Passeios pelo Bosque da Informação
estudos sobre representação e organização da informação e do conhecimento
Jaime Robredo e
Marisa Bräscher
(Organizadores)
ibict
Brasília DF - 2010

Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento

Jaime Robredo e Marisa Bräscher
(Organizadores)

Edição comemorativa dos 10 anos do Grupo de Pesquisa
Estudos sobre a Representação e Organização
da Informação e do Conhecimento – EROIC

ibict
Instituto Brasileiro de Informação
em Ciência e Tecnologia

© 2010
Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e
Universidade de Brasília (UnB). Programa de Pós-Graduação em Ciência da Informação da Faculdade de Ciência da Informação (FCI/PPGCInf).
Grupo de Pesquisa Estudos sobre a Representação e Organização da Informação e do Conhecimento (EROIC).

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT)
Diretor : Emir José Suaiden

COMITÊ EDITORIAL
Armando Malheiro da Silva – Universidade de Minho-Porto, Portugal.
Ely Francina Tannuri de Oliveira – Departamento de Ciência da Informação, Universidade Estadual Paulista (UNESP), Marília SP.
José Augusto Chaves Guimarães – Departamento de Ciência da Informação, Universidade Estadual Paulista (UNESP), Marília SP.
Lena Vania Ribeiro Pinheiro – Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Rio de Janeiro RJ.

Capa e foto: Juliano Serra
Diagramação, composição e links hipertextuais: Jaime Robredo

Passeios pelo bosque da informação : estudos sobre representação e organização da informação e do conhecimento / Jaime Robredo, Marisa Bräscher (Orgs.). – Brasília DF: IBICT, 2010. v + 335 p. ISBN: 978-85-7013-072-3. Edição eletrônica <http://www.ibict.br/publicacoes/eroic.pdf> .

1. Representação da informação. 2. Representação do conhecimento. 3. Organização da informação. 4. Organização do conhecimento. 5. Grupo de Pesquisa. I. Robredo, Jaime. II. Bräscher, Marisa. III. Título.
CDU 020 : 06

Como citar a presente obra:
ROBREDO, Jaime; BRÄSCHER, Marisa (Orgs.). **Passeios pelo bosque da informação:** estudos sobre a representação e organização da informação e do conhecimento – eroic . Brasília DF: IBICT, 2010. v + 335 p. ISBN: 978-85-7013-072-3. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf.

APOIO

Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq

Instituto Brasileiro de Informação em Ciência e Tecnologia

Universidade de Brasília – UnB

Coodenação de Aperfeiçoamento do Pessoal de Ensino Superior - CAPES

NOTA IMPORTANTE

A presente obra é uma edição comemorativa dos 10 anos de atividade do Grupo de Pesquisa 'EROIC – Estudos sobre Representação e Organização de Informação e do Conhecimento', cujos primeiros passos encontram-se registrados em Atas datadas de 21 de agosto, 11 de setembro e 16 de outubro de 1998. Sendo incluído, em 1999, no repertório de grupos de pesquisa do CNPq, após certificação pela Universidade de Brasília. O Grupo surgiu do espírito de colaboração de alguns professores do Departamento de Ciência da Informação e Documentação (CID), da Faculdade de Estudos Sociais Aplicados (FA), denominada posteriormente Faculdade de Economia, Administração, Contabilidade e Ciência da Informação (FACE), e à qual permaneceu vinculado até o dia 30 de abril do presente ano, quando o CID deixou de ser departamento para se tornar a mais nova faculdade da Universidade de Brasília, com o nome de Faculdade de Ciência da Informação (FCI). No Prefácio da obra, de autoria dos atuais líderes do Grupo, encontra-se um breve resumo de sua história e de algumas de suas realizações. O que se pretende destacar aqui, entretanto, é o aspecto inovador da presente obra, resultante de algumas de suas características, quais sejam, por exemplo: *i)* sua natureza de obra eletrônica, em forma de um e-livro que reúne em quatorze capítulos os resultados de pesquisas concluídas recentemente, ou ainda em andamento, de dezenove de seus vinte e três membros atuais, em solo ou em co-autoria; *ii)* a intenção de oferecer aos estudantes de graduação ou de pós-graduação um material didático atualizado e de fácil uso, que cobre diversos tópicos de interesse de várias disciplinas das grades curriculares, não somente dos cursos oferecidos pela jovem Faculdade, mas também de outras áreas que, cada dia mais numerosas, se interessam pelos processos informacionais, como atestam as variadas procedências dos candidatos a uma vaga na pós-graduação em ciência da informação; *iii)* o espírito de integração e intercâmbio de idéias entre os autores, e a necessidade de adequar os textos às normas da Associação Brasileira de Normas Técnicas, na busca das melhores práticas de apresentação dos capítulos, levou, juntamente com as modernas facilidades da computação gráfica, a resultados significativos quanto a qualidade atingida, praticamente sem custo; e *iv)* a preocupação de aproveitar ao máximo as facilidades oferecidas pelos recursos da Web, dentre eles a prática cada dia mais comum de 'copiar e colar', na intenção de extrair das bibliografias, para posterior uso, referências com endereços eletrônicos interessantes, em que porventura foi substituído o nome de um ou mais autores, citados na referência precedente, pelo traço '_____' pode levar a dramáticas perdas de preciosas informações. Assim, após reflexão, foi adotada a manutenção, nas bibliografias que acompanham os capítulos desta obra, dos nomes de todos os autores de trabalhos em coautoria. Esperamos que nos desculpem pela 'transgressão'.

Quanto aos nomes dos autores dos capítulos, no intuito de harmonizar a grafia, foi adotada a forma como aparecem nos respectivos currículos, no Sistema de Curriculos Lattes do CNPq. De igual forma, sugere-se que, ao citar os capítulos, se não houver razões expressas em contrário, os nomes dos autores sejam grafados, também, na forma registrada nesse sistema, sob o epígrafe `Nome em citações bibliográficas', na Seção `Dados pessoais' de cada currículo.

SUMÁRIO GERAL

PREFÁCIO – Jaime ROBREDO e Marisa BRÄSCHER, *p. 6-10*

I. ESTUDOS HISTÓRICOS E EPISTEMOLÓGICOS DA INFORMAÇÃO, *p. 11*
Capítulo 1. – Ciência da informação e Web semântica: Linhas convergentes ou linhas paralelas? – Jaime ROBREDO, *p. 12-47*
Capítulo 2. – A perfeição do nada: Da videologia à desideologia da percepção – Juliano Serra BARRETO, *p. 48-59*

II. REPRESENTAÇÃO E ORGANIZAÇÃO DA INFORMAÇÃO, *p. 60*
Capítulo 3. – O escopo da análise da informação – Dulce Maria BAPTISTA, Rogerio Henrique de ARAÚJO JR. e Eliana CARLAN, *p. 61-80*
Capítulo 4. – FRBROO – Conceitos básicos – João Alberto de Oliveira LIMA, *p. 81-92*
Capitulo 5. – Atributos dos Requisitos Funcionais para Registros Bibliográficos (FRBR) – Fernanda Passini MORENO, *p. 93-114*
Capítulo 6. – Organização da informação: Conceitos básicos e breve fundamentação teórica – Lígia Maria Arruda CAFÉ e Rodrigo de SALES, *p. 115-129*
Capítulo 7. – Organização da informação: Proposta de elementos de arquitetura da informação para repositórios digitais institucionais baseados na descrição física e temática – Fernanda de Souza MONTEIRO, *p. 130-145*

III. SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO, *p.146*
Capítulo 8. – Sistemas de organização do conhecimento: Antigas e novas linguagens – Marisa BRÄSCHER e Eliana CARLAN**,** *p. 147-176.*
Capítulo 9. – Categorização lingüística como esteio da organização do conhecimento – Fabio José Dantas de MELO, *p. 177-182*

IV. ESTUDOS MÉTRICOS DA INFORMAÇÃO, ***p. 183***
Capítulo 10. – Metrias da informação: História e tendências – Jaime ROBREDO e Jayme Leiro VILAN FILHO, *p. 184-258*
Capítulo 11. – Rede metodológica entre epistemologia, organização do conhecimento, bibliometria e tesauros: Concepção e construção do Tesauro Brasileiro de Ciência da Informação – Lena Vania Ribeiro PINHEIRO e Helena Dood FERREZ, *p. 259-276*

V. APLICAÇÕES E DESDOBRAMENTOS DA REPRESENTAÇÃO E ORGANIZAÇÃO DA INFORMAÇÃO E DO CONHECIMENTO, *p. 277*
Capítulo 12. – Descoberta do conhecimento em textos: Conceitos e convergências – Marcelo SCHIESSL, Hélia de Sousa CHAVES e Marisa BRÄSCHER, *p. 273-288*
Capítulo 13. – A contemporaneidade do tema governo eletrônico e perspectivas de pesquisas – Rinalda RIECKEN, *p. 294-319*

VI. OUTROS TEMAS E ATIVIDADES, *p. 320*
Capítulo 14. – Competência informacional em tempos de Web – Fábio Augusto Guimarães TEIXEIRA e Greyciane Souza LINS, *p. 321-335*

PREFÁCIO

Apresentamos um breve retrospecto do andamento e projeção das atividades do Grupo de Pesquisa "Estudos sobre a Representação e Organização da Informação e do Conhecimento - EROIC".[1] A título de curiosidade, reproduzimos, na *Figura 1*, a Ata da primeira reunião do grupo fundador do 'Grupo de Pesquisa' do Departamento de Ciência da Informação e Documentação (CID).

Universidade de Brasília (UnB)

Departamento de Ciência da Informação e Documentação (CID)

Reunião para proposta de criação de Grupo de Estudo e Pesquisa
21 de agosto de 1998

Pontos destacados

- Importância da recuperação da informação;
- Pós-graduação é diferente de pesquisa, pois esta também envolve a graduação;
- Proposta de criação da Faculdade de Ciência da Informação OU Instituto de Ciência da Informação. O segundo nome foi melhor aceito; pelos professores do Departamento, em razão de um instituto estar mais voltado para a ciência do que uma faculdade;
- Linhas de pesquisa a serem desenvolvidas pelo Grupo Cognitio:
  - Análise temática,
  - Indexação,
  - Processamento em linguagem natural,
  - Indexação automática,
  - Recuperação de informação,
  - Terminologia (sistema de conceitos),
  - Classificação,
  - Lingüística aplicada à ciência da informação,
  - Estudo de interfaces (interdisciplinaridade),
  - Lógica (sistema de conceitos),
  - Ciências cognitivas

Comentários feitos pela professora Haruka:

- Sobre as divisões da Lingüística. São elas: Semântica, Lexicologia, Morfologia, Semiótica, Lexicografia, Pragmática, Sintaxe e Terminologia;
- Proposta de estudo de interface e interdisciplinaridade entre Lingüística documentária, Lingüística aplicada e Ciência Cognitiva;
- Análise do discurso ≠ análise da informação;
  - Necessidade de [escolher um] nome para o Grupo.

Figura 1

Os trabalhos do Grupo, aos quais se refere a *Figura 1*, foram retomados na semana seguinte, quando se aprovou um calendário de reuniões mensais.

[1] O presente texto retoma e atualiza a apresentação feita por um dos organizadores do livro, por ocasião do II Workshop Internacional em Ciência da Informação (ROBREDO, J. "Breve retrospecto, andamento e projeção das atividades do Grupo de Pesquisa Representação e Organização da Informação e do Conhecimento". Brasília DF, 6 e 7 de dezembro de 2007. Disponível: http://dgp.cnpq.br/buscaoperacional/detalhegrupo.jsp?grupo=0240607COPBXUU.

Na *Figura 2*, se recolhem alguns dos comentários dos participantes das três reuniões que seguiram à primeira, o que permite ver como foram se delineando as linhas de pensamento que, posteriormente, passariam a ser chamadas 'linhas de pesquisa'

Universidade de Brasília (UnB)

Departamento de Ciência da Informação e Documentação (CID)

Extratos das Atas das reuniões do Grupo Cognitio, celebradas sucessivamente em 21 de agosto, 11 de setembro e 16 de outubro de 1998

Pontos destacados pelo Grupo [em ordem cronológica]

21 de agosto de 1998

Professor [Jaime] **Robredo:** Sugere que [...]: *i)* no início dos trabalhos [,,,] os participantes apresentem [...] sobre o que estão trabalhando; *ii)* sobre indexação automática, argumenta que esta proporciona maior objetividade que a indexação manual; *iii)* informa sobre artigo de sua autoria [a ser publicado] em *Ciência da Informação* [sobre o tema]; *iv)* [sugere a] divulgação periódica dos resultados das reuniões do Grupo.

Professor Odilon [Pereira da Silva]: Propõe que os integrantes do Grupo troquem [idéias sobre] matérias de interesse comum.

Outros comentários [autores não identificados]: *i)* [comparando] classificação e indexação, [surgem questionamentos sobre] a respectiva adequação de ambas para representar conceitos; *ii)* [o que] a classificação [aporta] à indexação.

11 de setembro de 1998

Professor [Jaime] **Robredo:** [Apresenta sua] linha de trabalho "medida da probabilidade de co-ocorrência dos descritores num mesmo documento". [Permite monitorar os temas de] interesse da comunidade científica [...] em um determinado período de tempo.

Rita Roberta [**Maria Barbosa Ferreira Porto**]**:** Concluiu pesquisa na área de terminologia, no PIBIC [Programa Institucional de Bolsas de Iniciação Científica].

Rita de Cássia [**do Vale Caribé**]**:** Elabora revisão da literatura em indexação, como trabalho da disciplina "Monografia em ciência da informação".

Rodolfo [sobrenome não informado]**:** Projeto [que procura] ir além da técnica informática [...] na busca de informações.

Professora Haruka [**Nakayama**]**:** *i)* enfatiza a interação entre biblioteconomia e lingüística; *ii)* biblioteconomia sem [clara] fundamentação teórica; *iii)* a terminologia oferece uma contribuição nos aspectos epistemológico e metodológico; *iv)* atraso do Brasil no aspecto normalização da língua.

Professora Cristiane [Galvão]: *i)* Pesquisa a parte conceitual da ciência da informação (como se constitui e como se relaciona com outras ciências; *ii)* como as pessoas apreendem os conceitos e como buscam informações; *iii)* estudo das linguagens documentárias mais voltadas aos textos científicos; *iv)* metodologias para transferência da informação tecnológica e mapeamento das estruturas textuais utilizadas nas unidades produtivas visando a compreensão do conteúdo; v) tradução de informações científicas pelo bibliotecário, para os não especialistas, visando tornar os textos inteligíveis para estes últimos.

16 de outubro de 1998

Professor Odilon [Pereira da Silva]: *i)* comenta sobre as disciplinas que leciona ou lecionou: Catalogação, Classificações (CDD e CDU), Indexação, e Bibliografia brasileira, e considera essa última disciplina inútil a partir do momento em que surgiram os sistemas automatizados de recuperação da informação; *ii)* lembra que o currículo de Biblioteconomia sofreu algumas mudanças [...], destacando o desparecimento da teoria da classificação.

Professora Cristiane [Galvão]: *i)* salienta a diferença entre classificação e linguagens documentárias modernas (classificação → lógica; linguagens documentárias → lingüística); ii) enfoca os aspectos da linguagem [do ponto de vista da] professora da disciplina "Análise da informação": como o leitor percebe o texto; como o usuário compreende um resumo; objeto da análise da informação; papel da análise da informação no ciclo documentário; terminolgia.

Professora Haruka [Nakayama] e **Rita de Cássia** [**do Vale Caribé**]**:** Propõem estudo sobre normas técnicas nacionais e internacionais referentes à análise da informação**Propostas** [proponentes não identificados]**:** *i)* [...] a cada duas semanas um membro apresenta um tema sobre o qual trabalham os membros do Grupo: estes podem dar um retorno ao autor na próxima reunião; ii) produção de documentos escritos.

Figura 2

SUMÁRIO GERAL

O Grupo, idealizado pela saudosa professora Haruka Nakayama, começou suas atividades, como vimos, com o nome de 'Grupo Cognitio', pretendendo dar destaque aos aspectos cognitivos no âmbito da ciência da informação.

Na falta de uma documentação fidedigna acessível (atas de outras reuniões, planos, etc.), é difícil traçar o percurso do Grupo, em seus primeiros passos[2], até sua oficialização, em 1999, no repertório de grupos de pesquisa do CNPq, certificado pela UnB, com o nome 'Representação e Organização do Conhecimento'. Sua missão foi assim estabelecida: "Promover o estudo e a pesquisa em temas relacionados com a análise da informação, indexação (manual ou auxiliada pelo computador), linguagem natural e linguagens documentárias, com ênfase nas dimensões lingüística, terminológica e comunicativa". Salvo engano, o Grupo era integrado inicialmente por: Haruka Nakayama e Jaime Robredo (líderes); Ailton Feitosa, Maria Cristiane Galvão, Renato Tarciso de Souza e Sebastião de Souza (membros participantes); Edgar Costa Oliveira, Paulo Borges e Rodolfo Vaz (pós-graduandos).

O Grupo passou por algumas mudanças resultantes da transferência da professora Haruka para o Departamento de Línguas Estrangeiras e Tradução, do Instituto de Letras da Universidade de Brasília, em 2002, e da transferência da professora Maria Cristiane para a Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, da USP, em 2005.

Por outro lado, a incorporação ao CID do professor Mamede Lima-Marques, em 2002, para lecionar no Programa de Pós-Graduação em Ciência da Informação (PGCINF), abriu uma nova frente de pesquisa em arquitetura da informação, incorporada em 2004, ao Grupo de Pesquisa, que muda, então, de nome para "Organização do Conhecimento e Arquitetura da Informação".

Enfim, em 2006, o PPGCINF aprova a criação de uma nova linha de pesquisa denominada "Arquitetura da Informação", e o Grupo volta a suas origens, passando a se denominar "Representação e Organização da Informação e do Conhecimento", mantendo a profícua colaboração com a linha "Arquitetura da Informação", bem como as duas já existentes anteriormente: "Comunicação da Informação" e "Gestão da Informação e do Conhecimento".[3]

[2] É claro que isso poderia ser feito a partir das publicações, comunicações, orientações, etc. de seus membros, que figuram nos currículos da Plataforma Lattes do CNPq, o que extrapola a intenção deste Prefácio. Aproveitamos a oportunidade para observar, a esse respeito, o perigo que representa para a preservação da memória da pesquisa brasileira o fato que, nas atualizações das informações sobre os grupos de pesquisa, as versões precedentes são apagadas. Fica lançada a idéia de se pensar em uma política de preservação da memória científica nacional.

[3] Por razões históricas que não cabe analisar aqui, o nome das Linhas de Pesquisa dos cursos de pós-graduação, registradas junto à CAPES, não coincide sempre com os nomes dos Grupos de Pesquisa que figuram no Diretório de Grupos de Pesquisa do CNPq. Assim, as Linhas de Pesquisa que figuram no PPGCInf são "Arquitetura da Informação", "Comunicação da Informação", "Gestão da Informação e do Conhecimento"', e "Representação e Organização da Informação e do Conhecimento", enquanto no Diretório de Grupos de Pesquisa do CNPq, figuram os Grupos: "Arquivos fotográficos", "Arquitetura da Informação", "Biblioteca e Sociedade", "Comunicação Científica", "Inteligência Organizacional e Competitiva", "Representação e Organização da Informação e do Conhecimento" e "Publicações Eletrônicas". O CNPq, por sua parte, admite a existência de linhas de pesquisa dentro dos Grupos de Pesquisa. Em nosso Grupo, são elas: "Estudos históricos e epistemológicos da informação", "Representação e organização da informação", "Sistemas de organização do conhecimento", "Estudos métricos da informação (infometria)", "Aplicações e desdobramentos da representação e organização da informação e do conhecimento" e "Outros temas e atividades", das quais as cinco primeiras figuram como as linhas registradas pelo CNPq como linhas pesquisa do Grupo (ordenadas alfabeticamente).

SUMÁRIO GERAL

O volume de pesquisas realizadas, de orientações concluídas em nível de mestrado e doutorado, e o número de trabalhos e de comunicações, publicados em periódicos ou em anais de eventos nacionais e internacionais, nos anos de atividade do Grupo, de 2000 a 2009, pode-se resumir da forma seguinte: *1)* Teses orientadas: 9; *2)* Dissertações orientadas: 36; *3)* Teses e dissertações em andamento: 15; *4)* Artigos: 16 ; *5)* Comunicações: 39; *6)* Livros e capítulos de livros: 6; *7)* Projetos com auxílio do CNPq : 5; *7.1)* Em andamento: 2 ("Construção de ontologias a partir da probabilidade de co-ocorrência de termos significativos, visando a construção de um arcabouço teórico generalizável da organização e recuperação de informações em ambientes Web, e seus desdobramentos práticos" – vigência: 2010 – 2014; e "Estudos comparativos: tesauros. taxonomias e ontologias " – vigência: 2010 – 2015).

Os valores acima, cotejados às linhas de pesquisa do grupo, demonstram uma forte tendência à consolidação destas últimas, que se reforça pelo fato marcante que representam os prêmios conquistados recentemente por alguns de seus membros: *i)* Rogério Henrique de Araújo Júnior – Melhor tese de doutorado (Prêmio ANCIB 2006); *ii)* Fernanda Passini Moreno – Melhor dissertação de mestrado (Prêmio ANCIB 2007); Auto Tavares da Câmara Júnior – Melhor trabalho do GT2 (VIII ENANCIB, 2007); *iv)* Lígia Maria Arruda Café e Marisa Bräscher Basílio Medeiros – Melhor trabalho do GT2 (IX ENANCIB, 2008).

A composição atual do grupo é: **Pesquisadores** (em ordem alfabética): Ailton Feitosa**,** Dulce Maria Baptista; Fábio José Dantas de Melo; Jaime Robredo (líder); João Alberto de Oliveira Lima; Juliano Serra Barreto; Marisa Bräscher Basílio Medeiros (líder); Rinalda Francesca Riecken; Rogerio Henrique de Araújo Júnior. **Pós-graduandos** (em ordem alfabética): Eliana Carlan; Fábio Teixeira, Fernanda de Souza Monteiro; Fernanda Passini Moreno; Greyciane Souza Lins; Hélia de Sousa Chaves Ramoa, Jayme Leiro Vilan Filho; José Marcelo Schiessl; Marcio de Carvalho Victorino; Mariana Baptista Brandt; Symball Rufino de Oliveira. Acrescentem-se a esses nomes: Armando Malheiro – Universidade de Minho-Porto, Lena Vania Ribeiro Pinheiro – IBICT/RJ e Lígia Maria Arruda Café – UFSC (membros externos) e, ainda: Leandro Rodor de Oliveira e Roberto Cantanhede (colaboradores externos).

A referida consolidação será reforçada se, como tudo parece indicar, for confirmada a incorporação ao Grupo de novos estudantes/pesquisadores interessados no estudo das atuais abordagens da representação descritiva de objetos – tanto convencionais quanto eletrônicos – para convertê-los em recursos informacionais.

Finalmente, pode-se informar que, para divulgar nossas realizações com mais detalhe do que permite o espaço alocado pelo CNPq, na Plataforma Lattes/Diretório dos Grupos de Pesquisa, para esse fim, encontra-se em fase avançada o estudo de um portal do Grupo, cuja inauguração esperamos anunciar em breve.

A partir daí, abrir novos espaços para listas de discussão, notícias, blogs, etc. poderá ser realidade já no presente ano. Assim, num percurso contínuo, com seus momentos altos e baixos, estamos comemorando o fruto de nossos esforços. Enfim, a maior prova de consolidação e de maturidade do Grupo, talvez seja a publicação do presente livro eletrônico, produto de múltiplas autorias e colaborações.

Nosso E- Livro, "*Passeios pelo Bosque da Informação: estudos sobre representação e organização da informação e do conhecimento – eroic*", que ora apresentamos, reúne quqtorze capítulos redigidos com a participação de dezessete de seus membros – pesquisadores, docentes, doutores e mestres, bem como mestrandos e doutorandos que buscam um aprimoramento profissional, os quais – em que pesem

as diferenças de suas experiências e níveis de conhecimento –, se igualam no seu interesse pela descoberta e aprofundamento dos conhecimentos que ainda se escondem nesse bosque da informação.

A idéia do livro foi reunir, num volume, tópicos de atualidade para estudo e aproveitamento dos alunos de graduação e pós graduação. Pretende-se que esses textos sirvam de ponto de partida para reflexão e discussão em sala de aula.

Para acessar o E-Livro, uma vez conectados à internet, basta abrir no navegador (por exemplo, Mozilla Firefox ou Internet Explorer) o endereço <http://www.ibict.br/publicacoes/eroic.pdf> que nos apresentará a capa. Rolando a imagem da capa até a página do SUMÁRIO GERAL, pode-se escolher um capítulo e iniciar a navegação.

A todos(as) os(as) colegas e amigos(as), e estudantes – hoje profissionais brilhantes –, sem esquecer os(as) esforçados(as) bolsistas de iniciação científica ou de otras modalidades de apoio, que muito contribuíram para o desenvolvimento do Grupo, nesses dez anos de labor e ousadia, nosso MUITO OBRIGADO!

Brasília DF, outubro de 2010

Jaime Robredo e Marisa Bräscher
Líderes do Grupo

I. ESTUDOS HISTÓRICOS E EPISTEMOLÓGICOS DA INFORMAÇÃO

Capítulo 1. – Ciência da informação e Web semântica: Linhas convergentes ou linhas paralelas? – Jaime ROBREDO, *p. 12-47*

Capítulo 2. – A perfeição do nada: Da videologia à desideologia da percepção – Juliano Serra BARRETO, *p. 48-59*

Capítulo 1

Ciência da informação e Web semântica: Linhas convergentes ou linhas paralelas?

Jaime Robredo[1]

SUMÁRIO DO CAPÍTULO 1

Resumo do Capítulo 1, *p. 13*
Como citar o Capítulo 1, *p. 13*

1. **INTRODUÇÃO DO CAPÍTULO 1**, *p. 14*
2. **WEB 1 WEB 2 WEB 3**, *p. 14*
3. **WEB SEMÂNTICA**, *p. 25*
4. **APLICAÇÕES REAIS DA WEB SEMÂNTICA NA CIÊNCIA DA INFORMAÇÃO EM SENTIDO AMPLO**, *p. 29*
   - **4.1 Situação atual em áreas predominantemente não documentárias**, *p. 30*
   - **4.2 Situação atual em áreas predominantemente documentárias**, *p. 32*
5. **REFLEXÕES CONCLUSIVAS**, *p. 41*
6. **BIBLIOGRAFIA DO CAPÍTULO 1**, *p. 42*

[1] Doutor em Ciências, Pesquisador Associado Sênior e Líder do Grupo de Pesquisa "Estudos sobre Representação e Organização da Informação e do Conhecimento – EROIC". Programa de Pós-Graduação em Ciência da Informação (PPGCI) – Faculdade de Ciência da Informação da Universidade de Brasília (UnB/FCI). CV Lattes: http://lattes.cnpq.br/9669_125022187444. E-mail: ssrrinfo@uol.com.br.

Resumo do Capítulo 1

Após uma detalhada discussão, fundamentada em numerosas opiniões, sobre o que significariam realmente as expressões 'Web 1.0', 'Web 2.0', 'Web. 3.0' e 'Web semântica', poder-se-ia concluir, que as fronteiras conceituais entre seus respectivos significados são bastante tênues, para não dizer imaginárias, pois a Web cresce e se desenvolve de forma vertiginosa, por diferentes caminhos que nos brindam com fantásticas inovações. Assim, passamos das rígidas redes de computadores à interconexão descentralizada, sem limites, entre usuários com autonomia crescente. Presta-se especial atenção à Web semântica em seus desenvolvimentos recentes, que apontam para novos paradigmas na representação e organização do conhecimento e seus desdobramentos práticos, na agregação de lógica à Web. Essas reflexões apontam para o importante papel que os novos profissionais da informação devem desempenhar na integração de ciência da informação e Web semântica e, para tanto, tornam-se indispensáveis e urgentes algumas mudanças nos currículos, no que tange ao ensino da analise, representação e organização da informação.
Palavras-chave : Web 1.0; Web 2.0; Web 3.0; Web semântica; Representação do conhecimento; Organização do conhecimento.

Abstract

Information science and semantic Web: Converging or parallel lines?
Following a detailed discussion, founded in numerous opinions about the real meaning of expressions such as 'Web 1.0', 'Web 2.0', 'Web 3.0', and 'Semantic Web', should it be possible to conclude that the conceptual borders of their respective meaning are significantly weak, not to say imaginary. The fast growing and development of the Web in last decades followed different roads, offering to us fantastic innovations. In fact, we jumped from the rigid computer networks to the user's unlimited decentralized connection, with a growing autonomy. Special attention is paid to the recent evolution and developments of the semantic Web, pointing to new representational knowledge organization paradigms and their practical developments in aggregating logic to the Web. These reflections indicate the important role to be played by the new information professionals in the integration of information science and semantic Web, and to make this achievement possible it is urgent and indispensable to reinforce some aspects of the curricula in that concerning the skills related to the knowledge analysis, representation and organization.
Keywords : Web 1.0; Web 2.0; Web 3.0; Semantic Web; Knowledge representation; Knowledge organization.

Como citar o Capítulo 1

ROBREDO, Jaime. Ciência da informação e Web semântica: Linhas convergentes ou linhas paralelas? *In*: Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento*. Brasília DF: IBICT, 2010. 335 p. ISBN: 978-85-7013-072-3. Capítulo 1, p. 12-47. Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

1. INTRODUÇÃO DO CAPÍTULO 1

A velocidade vertiginosa com que tem se desenvolvido a Internet, como um meio de divulgação e localização de informações de todo tipo e de qualidade e confiabilidade variáveis, parece que mereceria uma pausa para parar a pensar sobre o que a Web semântica (ou, como outros a denominam também, a Web 3.0) teria a contribuir nos futuros rumos do desenvolvimento da ciência da informação, entendida num sentido amplo, e de seus desdobramentos e aplicações. Mas, primeiramente e sem querer entrar em maiores considerações históricas, vamos lembrar o que parece que seriam hoje os significados mais comumente aceitos das denominações Web 1.0, Web 2.0, Web 3.0 e Web semântica, tratando, ao mesmo tempo, de situá-las nos seus respectivos contextos espaço-temporais. Atenção especial será prestada à Web semântica, às suas realizações até o momento e às suas aplicações reais na ciência da informação em sentido amplo, terminando com um sobre-vôo de suas tendências maia prováveis no Brasil e no mundo.

2. WEB 1.0, WEB 2.0, WEB 3.0

Começou-se a falar de Web (corte de *World Wide Web* – teia [de aranha]/rede de abrangência mundial –, ou simplesmente WWW) quando a Internet (rede internacional, rede interligada, etc.) se consolidou como uma realidade que ultrapassava os limites conceituais do seu idealizador Tim Berners-Lee (1989), quando ainda trabalhava no CERN (*Centre Européen de Recherches Nucléaires*) em Genebra, na Suíça, tornando realidade os sonhos ou premonições de Paul Otlet, com o seu *Mundaneum* (inaugurado em 1919, que veio a ser chamado por Rayward (2002) como "Internet de papel") e de Vanevar Bush (1945), com suas idéias esboçadas em seu copiosamente citado artigo "*As we may think*", publicado em *Atlantic Monthly*, em 1945, onde descreve sua engenhoca denominada *Memex (MEMory EXpanded)*; de fato, o primeiro sistema hipertextual multimídia. [2]

A Web, até os últimos anos do século passado, caracterizava-se por suas páginas estáticas, comportando textos, imagens e *links*, que somente podiam ser alteradas pelo *webmaster*. No início do novo século ocorre uma mudança progressiva e irreversível que torna as páginas dinâmicas e abre aos usuários a possibilidade de alterar e acrescentar dados. Nasce a Web dos usuários, e blogs e redes comunitárias se multiplicam e se expandem.[3] Sobre as origens da expressão Web 2.0, convém se reportar ao cenário traçado por Graham (2005), que diz:

> [...] 'Web 2.0' significa alguma coisa? Até há pouco tempo eu pensava que não, mas a realidade se tornou mais complicada. [...] Agora parece ter adquirido um significado. E ainda, aqueles que não gostam do termo estão provavelmente certos, pois se isso quer dizer o que eu penso, não temos necessidade dela. A primeira vez que ouvi falar de 'Web 2.0' foi na denominação da conferencia sobre a Web 2.0 em 2004. [4]

E o autor prossegue dizendo que, de acordo com Tim O'Reilly:

> [...] a expressão 'Web 2.0' surgiu "durante uma sessão de *'brainstorming'* entre O'Reilly e *Medialive International*." O que é *Medialive International*? "Produtores de conferências e shows comerciais sobre tecnologia" segundo o site da entidade. Assim, pode-se pensar que essa sessão de *'brainstorming'* era o seguinte: O'Reilly desejava organizar uma conferência sobre a Web e eles buscavam um nome para ela. [...] Havia algo como um desgaste semântico: sabiam que estavam acontecendo coisas novas e escolheram '2.0' para se referir a todo o que poderia vir pela frente." (A tradução é nossa) [5]

---

[2] Uma história dos primórdios da Internet encontra-se em Robredo (2005, p.235-252).

[3] Esta é, resumida, a visão de Raoul Mengis, bem sucedido executivo da empresa 1-Computer, na Suíça, especializada em consultoria e desenvolvimento de aplicações baseadas em Web 2.0 e Web 3.0 (MENGIS, 2008)

[4] O autor se refere à *Web 2.0 Conference* realizada de 5 a 7 de outubro de 2004 , no Hotel Nikko, em San Francisco, Califórnia.

[5] As traduções de todos os textos citados *ipsis litteris,* extraídos de originais em língua estrangeira, são do autor.

Segundo Tim O'Reilly (2006), "A Web 2.0 é a revolução dos negócios e da indústria dos computadores causada pela conversão da Internet em uma plataforma, e pela tentativa de entender as regras do sucesso da nova plataforma." O'Reilly (2005) reforça, assim, a visão da Web 2.0 como uma nova plataforma, expressa anteriormente: "Como muitos conceitos importantes, a Web 2.0, considerada como uma plataforma, não possui limites sólidos, senão mais do que isso, um núcleo gravitacional. Você pode visualizar a Web 2.0 como um conjunto de princípios e práticas que se agrupam como um verdadeiro sistema solar de sites que apresentam algum desses princípios, a distancias variáveis do núcleo."

Tim Berners-Lee, tem questionado, numa entrevista concedida a Scott Laningham da 'developerWorks' – um braço de IBM – em 28 de julho de 2006, se cabe utilizar a expressão 'Web 2.0' de alguma forma significativa, dado que muitos de seus componentes tecnológicos, já existiam desde os primeiros dias da Web. Parece oportuno reproduzir alguns trechos da resposta de Berners-Lee, que, no fim das contas, foi quem trouxe ao mundo a Web e é atualmente o diretor of the *World Wide Web Consortium*, à pergunta de Laningham sobre a Web 2.0: [6]

> LANINGHAM: Com a Web 2.0, você sabe, há uma explicação comum de que a Web 1.0 tratava da conexão entre computadores, tornando a informação acessível; e a Web 2.0 trata de conectar pessoas e de facilitar novos tipos de colaboração. Você vê assim a Web 2.0? [7]
>
> BERNERS-LEE: Absolutamente não. A Web 1.0 tinha tudo a ver com a conexão de pessoas. Era um espaço interativo e eu penso que a Web 2.0 é uma amostra evidente de jargão; ninguém sabe ainda o que significa. Se Web 2.0 é para você tratar de blogs e wikis, então é pessoas para pessoas. Mas isso é o que se pensou o tempo todo que era a Web.
> E de fato, você sabe, essa Web 2.0, entre aspas, significa usar as normas produzidas por todas as pessoas que trabalharam na Web 1.0. Isso significa usar o documento como modelo de objeto; isso vale para HTML e SVG, etc., isto é, usando http; assim, construindo coisas usando as normas da Web, mais, naturalmente, Java script.
> Assim, a Web 2.0 significa para alguns mudar em parte o pensamento do cliente tornando-o mais imediato, mas a idéia da Web como interação entre pessoas é realmente o que é a Web. Isso era o que se projetou para que a Web fosse um espaço colaborativo onde as pessoas poderiam interagir.
> Agora, eu realmente gosto da idéia das pessoas construindo coisas em hipertexto, algo como um espaço hipertextual comum, para explicar o que é o entendimento compartilhado, captando assim todas as idéias que levam a uma certa posição. Eu penso que isso é realmente importante. E penso que blogs e wikis são duas coisas divertidas, e penso que decolaram porque elas fazem grande parte da gestão da navegação para você, permitindo-lhe acrescentar conteúdo.
> Mas eu penso que haverá muitas mais coisas assim que acontecerão, diferentes caminhos em que as pessoas poderão trabalhar juntas.
> Os wikis semânticos são muito interessantes. São wikis em que as pessoas podem acrescentar dados, e esses dados podem ser trazidos à superfície e fatiados e cortados e picados usando todos os tipos possíveis de ferramentas semânticas, e é por isso que é tão emocionante o caminho por onde as pessoas e as coisas estão indo, embora eu pense que há uma grande quantidade de coisas novas que acontecem nesse sentido que ainda temos que inventar. [...]
>
> LANINGHAM: Posso formular uma questão bem abrangente, perguntando somente quais são as suas expectativas a longo prazo para a Internet e para a Web, e para as assombrosas invenções que todos nós estamos utilizando já? Que lugar você vê que elas ocuparão em nossas vidas, nas próximas décadas?

---

[6] O texto completo da entrevista encontra-se em: <http://www.ibm.com/developerworks/- podcast/dwj/cm-int082206txt.html>.

[7] O texto da entrevista publicado é uma transcrição direta de uma gravação ao vivo e, por isso, a linguagem é freqüentemente coloquial. Nossa tradução tratou de respeitar ao máximo a espontaneidade das expressões usadas pelo entrevistador e pelo entrevistado.

> BERNERS-LEE: Bem, a Internet foi realmente uma bela e bem arrumada plataforma, com base na qual me foi possível desenvolver a Web, e espero que a Web continue sendo uma bela e bem arrumada plataforma onde as pessoas continuem desenvolvendo outras coisas. Eu espero que a Web Semântica decole de forma que os dados, basicamente todos os dados que estão ai e aos quais você tem acesso, e as páginas Web estejam acessíveis como os dados, de forma que você poderá tratá-las como dados. Haverá grandes quantidades de aplicações construídas sobre essas idéias. Estamos começando a enxergar isso agora, mas é realmente assim, sabe, já estamos seriamente envolvidos no crescimento exponencial da Web Semântica, e isso é muito emocionante.
> Eu penso, também, que isso será outra vez uma nova plataforma onde surgirão muitas outras coisas a serem feitas. Isso será utilizado para interoperabilidade dos dados e para a integração dos dados através das empresas e entre empresas e entre aplicações. [...] Em geral, eu espero que nós, enquanto humanidade, podemos aprender a usar o espaço informacional para nos entendermos os uns aos outros, que nós mesmos podemos formar grupos de interesse, de forma que, entre nós, essa espécie de redes emaranhadas de grupos humanos abarquem o mundo de forma que.... não seja necessário clicar muitas vezes através da Web social para estabelecer o contato entre as pessoas, induzindo-as a se reunir para vencer os enormes desafios do mundo atual, no lugar de lutar umas contra outras. [...]
>
> LANINGHAM: Uma visão admirável, uma visão que vale a pena. Tim, foi um prazer e um honra. Obrigado pelo tempo que nos dedicou.

A entrevista de Berners-Lee mereceu atenção significativa, suscitando numerosos comentários geralmente favoráveis à sua posição oposta a visão reducionista de O'Reilly. Para não alongar excessivamente o debate, basta citar algumas passagens extraídas do texto de Nate Anderson (2006) publicado no blog da *Ars Technica*:[8]

> Uma rápida olhada em uma lista de sites declarados Web 2.0 é suficiente para entender do que [Berneers-Lee] está falando. Em que sentido esses sites estão fazendo algo qualitativamente diferente do que faziam outros que os precederam? Em que sentido esses sites fazem algo suficientemente parecido para que possam ser reunidos numa categoria única? Mesmo os que a defendem concordam que os maiores problemas da Web 2.0 são o excesso de 'Hiper' e a falta de uma verdadeira definição. Mas talvez a maior crítica ao fenômeno Web 2.0 seja seu fracasso em gerar produtos verdadeiramente importantes em quantidade suficiente. [...] O problema e que demasiados empreendedores iniciantes tratam de ser '2.0 de tal forma que o público fica absolutamente desconfiado. [...] A verdadeira questão não é se um site é "suficientemente 2.0" para ser 'Hiper', mas se o *site* é suficientemente útil e interessante, independentemente de sua abordagem.

Para concluir, pode-se dizer que a Web 2.0 é vista por alguns como uma segunda geração do desenho e da evolução da Web, que facilita a comunicação e o compartilhamento da informação, a interoperabilidade e a colaboração, com a subseqüente proliferação de redes comunitárias e sociais, hospedagem de serviços e aplicações, compartilhamento de vídeos, wikis, blogs e folksonomias. Sem uma verdadeira mudança de paradigma, valia a pena começar a numerar os passos sucessivos da web em seu progresso e sua evolução? Ainda, para quem desejar aprofundar mais a questão, poder-se-ia sugerir o artigo da Wikipedia intitulado Web 2.0.[9]

No que diz respeito à Web 3.0, as coisas são ainda mais complexas e...surpreendentes, como se verá ao considerar uma amostra de opiniões que parecem representativas de um universo bastante amplo.

---

[8] Situada em Chicago, Illinois, a *Ars Technica* é uma empresa de divulgação, análise de tendências tecnológicas e assessoria. Eles traduzem o nome 'em latim' como 'A Arte da Tecnologia' e inverteram, no seu marketing, a velha expressão latina *sine scientia ars nihil est* (sem conhecimento, a arte não é nada), para *sine ars, scientia nihil est*.

[9] Disponível em <http://en.wikipedia.org/wiki/Web_2.0>.

Pode-se começar pela opinião de Raoul Mengis (2008), já citado antes, para quem a W3 seria: "A Web pela Web, independente dos níveis de hardware e software; a Web em todo lugar e sobre quaisquer plataformas. [...] Uso crescente nos setores financeiro, produtivo, administrativo, comercial, propaganda e marketing. Código aberto, maior transparência."

Num "*twine*" [10] de alguns meses atrás, Aleksandar Malecic[11], jovem engenheiro eletrônico (Crepaja, Servia) – entusiasta de Web 3.0, gestão, teoria de sistemas, psicologia e filosofia – apresentou uma síntese bastante clara da "biografia" da Web 3.0 até praticamente os dias atuais, tomando como base alguns elementos da página em inglês da Wikipédia intitulada "Web 3.0" e retirada do acesso público em 24 de fevereiro de 2009[12].

> Web 3.0 é um dos termos usados para descrever o estado de evolução da Web que segue a Web 2.0. Dado que as possibilidades técnicas e sociais identificadas neste último termo devem ainda ser ainda realizadas completamente, a definição da natureza da Web 3.0 se torna extremamente especulativa. Em geral, isso se refere a os aspectos da Internet que, embora potencialmente possíveis, não são nem técnica nem praticamente realizáveis no momento atual. [...] O termo 'Web 3.0' foi introduzido para especular sobre a futura 'onda' de inovações da Internet. A visão da próxima etapa da evolução da *World Wide Web* varia enormemente, do conceito de tecnologias emergentes como a Web Semântica, transformado a forma como a Web é usada (e conduzindo a novas possibilidades da inteligência artificial), até a observação de que os aumentos da velocidade de conexão da Internet, as aplicações modulares da Internet
> e os avanços na computação gráfica haverão de desempenhar um papel fundamental na evolução da *World Wide Web*. [...] A suposta terceira geração da Web, que coletiviza os serviços baseados na Internet, pode ser chamada de 'Web inteligente' – com aplicações que fazem uso da Web semântica, de micro-formatos, da busca em linguagem natural, da mineração ou garimpagem de dados (*data mining*), do ensino com auxílio de máquinas, dos agentes de recomendação (*recommendation agents*[13]), e tecnologias de inteligência artificial, com ênfase na compreensão da informação com auxílio de máquinas , buscando elevar os níveis de experiência produtiva e intuitiva do usuário." (*sic*)

A expressão Web 3.0 apareceu pela primeira vez num polêmico e virulento artigo de Jeffrey Zeldman (2006), publicado na revista eletrônica '*A List Apart Magazine – for people who make websites*' [14], onde critica a Web 2.0 e outras tecnologias relacionadas:

> A Biblioteca Pública de Nova York organizou recentemente um debate sobre o andamento do projeto que Google vinha realizando com a cooperação de outras prestigiosas bibliotecas, para digitalização de livros. tornando-os acessíveis. O projeto entusiasma os futuristas, mas enfurece alguns editores. É obvio que a digitalização cria cópias virtuais. Os editores reclamam que essa duplicação viola os direitos autorais, mesmo se o conteúdo dos livros se oculta do público. À medida que eu assistia ao debate, meu desconforto com o ambiente "hiper" que cercava um tipo emergente de desenvolvimento da web se transformou numa explosão de ódio explícito. [...] ... ficou claro que a "Web 2.0 era não somente maior que o Apocalipse, mas também mais proveitosa. Proveitosa, claro, para os investidores como o palestrante. Ainda, a nova corrida do ouro não deve ser confundida com a bolha do ponto-com' dos anos 90 [do século passado]. Web 1.0 não foi destrutiva. Entende? Web 2.0 é totalmente destrutiva. [...] O que significa Web 2.0, se é que significa alguma coisa?

---

[10] Uma visão do *twine* pode-se encontrar em Richard MacManus (2007).

[11] O *twine* de Aleksandar Malecic (2009 ?), encontra-se acessível em: <http://www.twine.com/-item/120wyxj28-x6/web-3-0-wikipedia-the-free-encyclopedia>.

[12] URL original: <http://en.wikipedia.org/wiki/Web_3.0>.

[13] Os agentes eletrônicos de recomendação (*recommendation agents*) permitem elevar a qualidade dos serviços oferecidos em linha aos consumidores, ajudando-os a decidir na escolha de produtos. Ver, por exemplo: Ascoy *et al*. (2006).

[14] *A List Apart Magazine* (ISSN 1534 0295) explora o design, desenvolvimento e significado do conteúdo da web, com foco especial nas normas e melhores práticas da web. Ver, por exemplo: <http://www.alistapart.com/about>.

> [...] A você que está trabalhando sobre um produto de *software* social turbinado por *Ajax*, boa sorte, que Deus o abençoe e divirta-se. Lembre que outras 20 pessoas estão trabalhando sobre a mesma idéia. Assim, não o complique, despache-o antes que outros o façam mais desagradável de se ver que um multimilionário presunçoso.
> [...] A você que percebe as falhas porque passou o último ano aprimorando suas habilidades na web e atendendo clientes, ou fazendo um negócio ou, talvez publicando conteúdo, você que é especial e simpático, mantenha-se de cabeça erguida e não deixe que ninguém veja suas lágrimas.
> [...] No que me diz respeito, estou rompendo com o homem medíocre e pulando para a Web 3.0. Porque esperar?

No número 282 de *A List Apart Magazine,* de 21 de abril de 2009, apresenta-se uma relação completa, com uma breve sinopse, dos artigos publicados por Zeldman entre 1999 e 2009. Na sinopse do artigo acima referido, pode-se ler: "Web 2.0 é uma nova estrelinha num emaranhado de nós da rede, na cauda longa da experiência disruptiva de amanhã. Web 3.0 pensa que você é tão 2005." [15]

Victoria Shannon (2006), num artigo publicado no *International Herald Tribune*, comentando a abertura da 15ª Conferência Internacional sobre a *World Wide Web*, realizada em Edimburgo, destaca a frase de Tim Berners-Lee: "A Web está, simplesmente, caminhando para se tornar mais revolucionária". E comenta, parafraseando Berners-Lee:

> Embora Berners-Lee se insurja contra o uso do termo "Web 2.0" – o zumbido de uma abelha do *Silicon Valley* para descrever a Internet desde a depressão do ponto.com na virada do século – ele enfatiza que vê uma rede revigorada. Para muitos, do ponto de vista tecnológico, Web 2.0 significa uma Internet mais interativa, customizada, social, com amplo uso das mídias – sem mencionar a busca do lucro – do que era há uma década [...] É uma mudança aparente [...] Mas os especialistas em software, os executivos e os empreendedores do setor tecnológico, que participam da conferência de Edimburgo, estão procurando mais longe que isso, apontando para outra frase feita, talvez com menos apelo para o usuário: a Web semântica, outra criação de Berners-Lee.

Shannon prosegue: "Nesta versão da Web, sites, links, mídias e bases de dados são 'mais inteligentes" e capazes de prover automaticamente mais conhecimento que as atuais." Em seu artigo, ele transcreve os comentários de alguns participantes que, por sua oportunidade e atualidade, são resumidos a seguir:

> NIGEL SHADBOLT (Universidade de Southampton, Inglaterra): Agora vem o esforço para empurrar os desenvolvedores da Web a adotar os componentes e colocá-los no *software*, nos serviços e nos *sites*. Há um espaço evidente para a Web semântica nas ciências da vida, na medicina, na pesquisa industrial. Aí é donde se situa hoje o foco, em grande parte. Nós estamos buscando comunidades de usuários da informação para mostrar as vantagens. É um processo evolutivo. A grande questão é saber se isso evoluirá para o negócio ou para os consumidores. A conseqüência de uma Internet aberta e difusa é que podem surgir resultados inesperados de lugares não previstos. Por exemplo, algumas experiências anteriores, ao mostrar novas relações entre os dados da Web, deram lugar o surgimento do *Flickr*, um *site* de compartilhamento de fotos que os próprios membros categorizam, e o *FOAF – friend of a friend –*, um projeto e pesquisa para decrever os diversos *links* entre pessoas. Ambos acrescentam 'significado' onde esse contexto não existia antes, simplesmente mudando a programação subjacente para refletir os *links* entre bases de dados. Penso que em 10 ou 12 anos veremos o crescimento acelerado desse tipo de integração da Web semântica.

[15] Na falta de uma tradução fidedigna da palavra '*interwingled*', traduzimos a expressão metafórica '*on the interwingled longtail'* como 'num emaranhado de nós da rede, na cauda longa'. Agradecemos antecipadamente qualquer tradução melhor.

> PATRICK SHEEHAN (*3i Investments*, Londres): Os investimentos estão começando a acompanhar o período 'tudo azul' dos grandes sonhos da Web semântica. [...] A tecnologia ainda procede maioritariamente das universidades. Mas as empresas [de investimentos] de investimentos são uma realidade e resolvem problemas reais – só não fazem pesquisa. Acredito que a Europa, em lugares como a Universidade de Southampton, é o líder mundial da pesquisa em Web semântica, embora ainda resta por ver se a região poderá comercializá-la com sucesso.
>
> TIM BERNERS-LEE (Head, *World Wide Web Consortium* – W3C): As pessoas ficam me perguntando o que é a Web 3.0. Eu penso que pode ser quando você atingiu um elevado nível em vetores gráficos escaláveis – qualquer coisa que adquira uma forma ondulada, ou se dobre e pareça indefinida, nebulosa – na Web 2.0, e tenha acesso à Web semântica integrada através de um enorme espaço de dados, você terá um incrível recurso a dados.
>
> PATRICK SHEEHAN : Eu acredito que a Web semântica será profunda. Em tempo, será tão obvia como hoje nos parece a Web.

John Markoff (2006) escreve sobre o mesmo tema:

> Dos bilhões de documentos que formam a World Wide Web e dos links que os mantêm juntos, os cientistas do computador e número crescente de empreende-dores estão encontrando novos caminhos para garimpar a inteligência humana. A finalidade é acrescentar uma camada de significado no topo da Web atual que faça dela menos um catálogo e mais um guia – e ainda forneça os fundamentos para que os sistemas possam 'raciocinar' de forma semelhante aos humanos. Esse nível de inteligência [...] tem iludido os pesquisadores durante mais de meio século. Denominado Web 3.0, tal esforço está na infância, e idéia faz dizer aos céticos que se trata de uma visão inalcançável. Entretanto, as tecnologias subjacentes estão sendo vistas cada vez com maior interesse pelas grandes empresas [...] que investem em projetos concretos de aplicação prática. [...] Todos os projetos que pretendem criar a Web 3.0 beneficiam-se da potência crescente dos computadores, que podem rápida e completamente esquadrinhar a Web. Nova Spivack, fundador de *Radar*, empresa cuja tecnologia detecta relações entre pacotes de informações garimpando na Web, chama isso a Base de Dados da World Wide Web.
> A Web 2.0, que descreve funcionalidades que simulam a conexão de aplicações (mapeamento geográfico) e serviços (como o compartilhamento de fotos), na Internet, tornaram-se recentemente o foco dos 'ponto.com' tipo 'hiper' no *Silicon Valley*. Entretanto, o interesse comercial na Web 3.0 — ou "Web semântica" — pela idéia de agregar significado, só está começando.

Vale a pena acrescentar algumas opiniões formuladas nos anos seguintes por especialistas e profissionais da Web, que confirmam, se ainda for necessário, o conflito sobre o que seria a Web 2.0, a Web 3.0 e a Web semântica (estas duas últimas consideradas por alguns como sendo a mesma coisa), para tecer algumas considerações e tentar, se for possível, estabelecer uma posição mais clara. Da página da Wikipedia sobre a Web 3.0 retirada em fevereiro do presente ano, que pode ser encontrada em alguns blogs que ainda a conservam[16], foram escolhidos alguns exemplos ilustrativos.

No *Technet Innovation Sumit*, Jerry Yang (2006, *apud* Dignan *et al.*, 2006), fundador do Yahoo!, falou:

> A Web 2.0 está bem documentada e explicada. O potencial da Net alcançou uma massa crítica, com capacidades que podem se realizar em rede. Também estamos vendo dispositivos mas ricos, nos últimos quatro ou cinco anos, não só em hardware como consoles de jogos e dispositivos móveis, mas também na camada de software. Você no precisa ser um cientista da computação para fazer um programa. [...]
> A distinção entre profissionais, semi-profissionais e usuários será apagada, surgindo redes de negócios e aplicações.

[16] Ver por exemplo: 'Web 3.0' (2007) em Information online, 'What is Web 3.0?' (sem data) em Design 3.0, e 'Web 3.0' (2009) em Greenlivingpedia.

Reed Hastings (2006, *apud* Dignan *et al.*, 2006), fundador de *Netflix*[17], no mesmo encontro, definiu de forma simples as diferentes fases da Web: " A Web 1.0 era discar, largura de banda de 50K em média, a Web 2.0 é uma largura média de banda de 1 megabit e a Web 3.0 será o tempo todo 10 megabits, que será a Web tudo vídeo..."

Eric Schmidt (2007), executivo do Google, apresentou um vídeo no Fórum Digital de Seul, realizado em maio de 2007, no qual respondeu da seguinte maneira a uma pergunta sobre como definiria a Web 2.0 e a Web 3.0:

> A Web 2.0 é um termo de marketing, e penso que a Web 3.0 acaba de ser inventada. Mas se eu tivesse que imaginar o que é a Web 3.0, diria que é uma forma diferente construir aplicações... Minha previsão seria, em último termo, que a Web 3.0 viria a ser considerada como aplicações que se juntam. Existem numerosas características: as aplicações são relativamente pequenas, os dados estão em uma nuvem, as aplicações podem ser executadas em quaisquer máquinas, PC ou telefone celular; as aplicações são muito rápidas e customizáveis. Além disso, as aplicações estão distribuídas de forma viral: literalmente mediante redes sociais, e-mail. Você não gostaria de comprá-las e armazená-las... É um modelo de aplicação muito diferente do que vimos até agora no mundo da computação

Sob o título apelativo "Jimmy Wales, pai da Wikipedia, diz que a Web 3.0 não tem futuro", José Moura (2008) publica um breve texto, acompanhado de um vídeo de Wales – com subtítulos em português – cujo objetivo seria arrecadar fundos para manter viva a Wikipedia.[18]

Poder-se-ia convir que, sem falar de certa tendência observada de associar uma coisa com a outra e/ou de discutir/comparar varias coisas ao mesmo tempo (Web 2.0 e Web 3.0, Web 3.0 e Web semântica), não ficou muito claro quais seriam os significados mais geralmente aceitos. Assim convém explorar o conteúdo da página '*Web 3.0',* da Wikipedia retirada em 23 de fevereiro de 2009, como atestam a página de igual nome da *Greenlivingpedia* (2009) e a página de autoria de Dan Grigorovici (2009)[19], onde o autor clama sua indignação nos seguintes termos:

> Dia de lamentação: ontem, 23 de fevereiro de 2009, a 1 hora e 22 minutos, a entrada "web 3.0" na Wikipedia foi deletada! John Markoff e outros proponentes do rótulo, tristes; os oponentes se alegram. Já falei muitas vezes aqui e em outros lugares [...] que o termo "web 3.0", embora não seja uma denotação tão brilhante, poderia ficar como um denominador comum a ser usado para significar a adoção de uma grande corrente, mais ampla, do conjunto de tecnologias que a circundam. [...] De outro ponto de vista, se você mata a "web 3.0", então por definição, você deve matar a "web 2.0" e a "web 1.0" ou quaisquer outros termos/rótulos de tipo marketing.[20]

Note-se que a página 'deletada' chegou a atingir, segundo informação publicada na página intitulada também 'Web 3.0' por *Evolving Trends* (2008), mais de dois milhões de visitas durante os seis primeiros dias de sua publicação.

---

[17] Netflix (NASDAQ: NFLX) é um serviço de locação em linha de DVDs e discos Blu-ray. Com sede em Los Gatos, Califórnia, já reuniu mais de 100.000 títulos e cerca de 10 milhões de assinantes. (Fonte: Wikipedia http://en.wikipedia.org/wiki/Netflix).

[18] O vídeo, apresentado sem data nem referência, pode ser exibido no Blog do Moura. Ver Moura, José (2006).

[19] Ver, também a página 'Web 3.0' da Wikipedia que publica, para referência o *'log'* das 11 ações sucessivas de retirada das diversas partes da página entre 17 abr 2006 e 26 fev 2009. Ver Wikipedia (2006).

[20] O autor do presente capítulo aproveita a oportunidade, para expressar sua opinião, puramente pessoal, sobre a conveniência de banir o uso de 1.0, 2.0, 3.0 para qualquer tipo de Web. Ou haveria, também, 'webs' intermediárias, com decimais – 3.1, 3.5, 3.8, 4 versão beta, etc. – que não são, de fato, senão etapas sucessivas de um processo contínuo de desenvolvimento e progresso? Assim, somente quando ocorrer uma mudança de paradigma, bastaria adicionar um adjetivo adequado ao termo genérico Web, especificando o surgimento de um novo ramo na árvore do progresso. Parece que esse seria o caso da Web semântica, talvez da Web 3D e de outras que virão. Não se deve esquecer que um nome simples é sempre mais genérico que um termo adjetivado.

Na página da Greenlivingpedia[21], acima citada encontram-se algumas pistas do que seria uma definição expandida da Web 3.0, segundo a *Nova Spivack* (2006), que define a Web 3.0 como a terceira década da Web (2010–2020), durante a qual cabe pensar que algumas tendências tecnológicas complementares atingirão simultaneamente novos níveis de maturidade, dentre as quais:

- *transformação* da Web de uma rede de aplicações armazenadas separadamente e repositórios de conteúdo, em tudo mais compacto e interoperável;
- *conectividade ubíqua,* adoção da banda larga, acesso à Internet móvel e a dispositivos móveis;
- *computação em rede,* interoperabilidade dos serviços na Web, processamento distribuído, processamento em grade e processamento em nuvem;
- *tecnologias abertas*, APIs[22] e protocolos abertos, formatos de dados abertos, plataformas de *software* de código aberto e dados abertos;
- *identidade aberta*, 'OpenID', reputação aberta, identidade de '*roaming*' e dados pessoais portáveis;
- *a web inteligente*, tecnologias da Web Semântica tais como RDF, OWL, SWRL, SPARQL, GRDDL, plataformas de aplicações semânticas e armazenamento de dados com base em declarações;
- *bases de dados distribuídos*, a "*World Wide Database*" (tornada disponível pelas tecnologias da Web Semântica); e
- *aplicações inteligentes*, processamento da linguagem natural. Aprendizado com auxílio de máquinas, agentes autônomos.

Às quais se acrescentariam outras pesquisas sobre:

- *espaços 3D*, pesquisas lideradas pelo Consórcio 3C (*Second Life* [23], espaços 3D compartilhados); e
- *pesquisa sócio-tecnológica*, a inclusão do conceito de "Web 0.0" como o mundo real pré-existente (também proposto "sensual web").[24]

Realmente, não fica muito fácil buscar, ou tentar estabelecer com algum fundamento, uma delimitação clara entre as 'Webs' 1.0, 2.0 e 3.0, e ainda a coisa se complica devido à confusão entre Web 3.0 e Web semântica, e à luta travada entre o Google e Wikipedia, durante os últimos anos, pela conquista de uma posição hegemônica. Sobre a Web semântica são apresentadas algumas reflexões na próxima Seção, que se espera ajudem a vislumbrar aonde se dirige um dos caminhos mais promissores da evolução da Web nos próximos anos. Para os leitores interessados(as) em conhecer alguns detalhes da luta entre os dois gigantes mencionados, se reproduzem os títulos de algumas páginas da Web, indicando as datas de publicação e os respectivos endereços:

- *Wikipedia 3.0: The End of Google?* <http://evolvingtrends/-wordpress.com/2006/-06/26/wikipedia-30-the-end-of-google> 26 jun 2006:
  A Web Semântica (ou Web 3.0) promete 'organizar a informação do mundo' de uma forma dramaticamente mais lógica do que o Google jamais poderá conseguir com o desenho de sua máquina atual. Isso é especialmente verdade do ponto de vista da compreensão da máquina, como o oposto da compreensão humana. A Web Semântica requer o uso de uma linguagem ontológica declarativa como OWL para produzir ontologias de domínios específicos suscetíveis de ser usadas pelas máquinas para raciocinar encima da informação e formular novas conclusões, não se limitando a comparar palavras chave".

---

[21] Também publicada em *'User:Greyskinnedboy/Web 3.0'* (Wikipedia, 2009).
[22] Interface de programação da aplicação (*application programming interface* - API): conjunto de rotinas, estruturas de dados, classes de objetos e/ou prtocolos/or fornecidos por bibliotecas e/ou serviços de sistemas operacionais para dar suporte à construção das aplicações.
[23] Ver, por exemplo, Andrew Wallenstein (2007).
[24] Ver, por exemplo, Tristan Zand (2006).

[...] Entretanto, a Web Semântica, que ainda está em fase de desenvolvimento, na qual os pesquisadores estão tentando definir os modelos de desenho melhores e mais fáceis de usar, precisarão da participação de milhares de especialistas ao longo do tempo para produzir essas ontologias por domínios necessárias para que funcionem. As máquinas (ou o raciocínio baseado em máquinas, também chamado 'AI software' ou 'info-agentes' seriam então capazes de utilizar essas ontologias construídas com enorme trabalho – mas não manualmente – para elaborar uma visão (ou modelo formal) de como os termos individuais dentro da informação se relacionam os uns com os outros. Tais relações podem-se considerar como os axiomas (verdades assumidas inicialmente) que, junto com as regras que regem os processos inferenciais tanto permitem como impõem a interpretação (e bem-formado uso) dos termos pelos agentes para deduzir novas conclusões a partir dos dados existentes [...] O surgimento de uma Wikipedia 3.0 (como no caso da Web 3.0 ou Web Semântica) construída com base no modelo da Web Semântica proclamará o fim do Google como a 'Máquina Extrema que Responde'. Ela seria substituída pela 'WikiMente' que não seria uma mera máquina de busca como o Google, mas um verdadeiro 'Cérebro Global' [...]

- *Wikipedia vs. Google*. <www.globalbydesign.com.blog/2006/12/26/wikipedia-vs-google/> 26 dez 2006. John Yunker (2006), da Microsoft, escreve:
  Em nossa '*2006 Web Globalization Report Card*' , os dois *´top sites´* da Web global, acima de todos, eram Google e Wikipedia. As duas organizações suportam mais de 100 línguas e um *template´* de desenho global flexível e que se carrega rapidamente. É por isso que eu estava interessado em ler o que Jimmy Wales, o fundador da Wikipedia, está planejando para desenvolver uma máquina de busca com base Wiki. Evidentemente, isso é muito mais fácil de se dizer do que se fazer. Wikipedia precisou de uma enorme rede de voluntários dedicados, para chegar a alguma coisa de valor. Será que um grupo semelhante conseguirá manter-se unido em torno da máquina de busca?Será muito interessante de acompanhar. Jimmy escreve que as máquinas de busca falham 'pelas mesmas razões que provocam a falha dos softwares proprietários: a falta de liberdade, a falta de trabalho em comunidade´.

- *'Google Warming Up the Wikipedia 3.0 vision?'* <http://evolvingtrends/wordpress.com/2007/12/14/google-tries-again-to-co-opt-the-wikipedia-30-vision> 14 dez 2007:

  Google parece que está reinventando Wikipedia com seu novo produto, que eles chamam ***'knol'*** (ainda não acessível ao público). Numa tentativa de captar o conhecimento humano, Google aceitará artigos de usuários que se identificarão como seus autores. Se assim o desejarem, podem autorizar complementos... Outros usuários podem ser convidados a avaliar, editar ou comentar os artigos.

- '*Google versus Wikipedia.*' <http://11pixels.ciberarte.com.br/motores-de-busca/506/google-versus-wikipedia.html> 8 jan 2008:

  Nos últimos tempos a supremacia do Google vem sendo ameaçada pela Wikipedia, que depois do próprio Google, é um dos sites mais visitados do mundo. A convivência parecia pacífica até bem pouco tempo atrás, até mesmo porque os dois recursos serviam para fins bem distintos. Enquanto o Google é mais conhecido como o principal motor de busca da atualidade, a Wikipedia é uma enciclopédia livre construída pelos próprios visitantes do site. No entanto, em dezembro de 2006 o fundador da Wikipedia, Jimmy Wales, anunciou um novo projeto para se somar aos vários da *Wikimedia Foundation*, que hoje conta com a própria Wikipedia, com mais de 9 milhões de artigos em 253 línguas; o Wiktionary, um dicionário aberto; o Wikinews, para notícias; o Wikibooks, com livros e manuais grátis; o Wikiquote, uma coleção de citações e outros projetos que podem ser conferidos no site da fundação.

- '*WIKIA Search'.* <http://11pixels.ciberarte.com.br/motores-de-busca/506/google-versus-wikipedia.html> 8 jan 2008.

A *Search Wikia* é a nova cartada de Wales. Lançada oficialmente ontem (versão alfa), o sistema é definido pelos seus criadores como uma ferramenta de busca *open-source*, que rivaliza com as tradicionais e ultrapassadas ferramentas de busca proprietárias como o Google. O novo modelo proposto por Wale, assim como todos os projetos da Wikipedia, baseia-se em resultados cadastrados e administrados pelos próprios usuários. [...] O troco do Google chama-se Knol, uma ferramenta planejada pelo Google para abrigar artigos de assuntos e áreas diversas gerados pelos usuários. A diferença marcante em relação à

Wikipedia é que os autores dos artigos serão especialistas em assuntos diversos, convidados pelo Google, portanto, o conteúdo do Knol não poderá ser editado por qualquer um. Alguns vêem esta dramática diferença de forma positiva, já que o conteúdo da Wikipedia levanta várias críticas relacionadas à qualidade e à manipulação das informações.

- '*Yahoo! Answer vs. Google+Wikipedia vs. Powerset.*' <http://asc-parc.blogspot.-com/2008/05/yahoo-answer-vs-googlewikipedia-vs.html> 13 mai 2008. Ed. H. Chi (2008) escreve:
  Uma das coisas boas da Web é que todo esse conhecimento construido social e colaborativamente pode ser rastreado com facilidade. O algoritmo *PageRank* (vagamente baseado num voto coletivo e num mecanismo de promediação sobre os links) é, provavelmente, responsável pelo enorme ganho de produtividade em todo o mundo, e ainda esclarece muitas coisas.[25] Não é, pois, surpreendente que a Web 2.0 tente construir, sustentando-se nesse sucesso, para melhorar o compartilhamento do conhecimento e a descoberta da informação, como o ´*Yahoo! Answers*´. [...] nestes dias a Wikipedia parece ser a técnica mais recente para socializar a construção de coleções e enciclopédias (sem esquecer sistemas mais antigos como o *Open Directory Project*). A grande novidade são as máquinas de busca com base semântica, como o *Powerset*. [...] com sua abordagem baseada em significados. [...] Não há garantia de que sua pergunta seja satisfeita pelo conteúdo da Wikipedia, mas técnicas de busca tradicionais têm a vantagem de deixar que você conheça se a informação existe realmente na base de conhecimento (supondo que você sabe como formular a pergunta). Eu utilizei o Google para pesquisar dentro da Wikipedia, e consegui respostas bastante boas.

- *Google Co-Op: The End of Wikipedia?* <http://evolvingtrends/wordpress.com/2006/09/24/google-co-op> 24 set 2008:
  Google teve a idéia de se servir de ´*experts*´ para estruturar [e dar significado à] informação com a finalidade de melhorar a probabilidade de encontrar, a partir do artigo "Wikipedia 3.0?" Ou é resultado puro e natural da evolução de seu pensamento? Ou ambos? Segue um extrato do experimento 'Google Co-Op´anunciado recentemente, que mostra o caminho para melhorar a busca nos tópicos que você conhece melhor. Se você é medico, com conhecimento em uma doença específica,você pode contribuir usando as etiquetas no tópico saúde para anotar todas as páginas web que sabe que fornecem informação útil e confiável sobre essa doença. Seus pacientes e outros usuários do Google poderiam contribuir para você e se beneficiar de sua experiência, Você pode participar em vários tópicos que já foram trabalhados na área de saúde [...] Ou, se quiser algo diferente, vá enfrente e inicie um tópico seu. Mostraremos um exemplo de como proceder para marcar as páginas web sobre um tópico existente ou como criar seu próprio tópico.

- *Semantic MediaWiki: Towards Wikipedia 3.0* <http://korrekt.org/talks/Wikimania2007/SMW.html> e, também, <http://pt.wikipedia.org/wiki/Semantic_MediaWiki> 21 abr 2009.[26]
  Semantic MediaWiki (SMW) é uma extensão do MediaWiki, que usa tecnologias da web semântica (OWL, RDF, etc.) para que o ambiente wiki, além da navegação por meio de hiperligações, passe a ter uma estrutura suscetível de ser entendida pelas máquinas. Permite que os artigos (e hiperligações) tenham relações, atributos etc. Assim, o SMW pode ajudar na busca, organização, navegação, melhoria e compartilhamento do conteúdo do wiki. Por ser uma extensão, nenhuma parte do MediaWiki é sobrescrita, sendo assim, pode ser incorporada a wikis já em pleno funcionamento sem muito custo de migração.

- *'Google, Wikipedia & ProQuest'.* Comparison video. 2009. < <http://www.proquestk12.com/demo/Google_ProQuest/Google_ProQuest.shtml>
  Como pode ProQuest realizar pesquisas em profundidade em recursos empilhados, como Google e Wikipedia? Há um lugar e um tempo para os três, e este vídeo esclarecerá tudo.

[25] O autor apresenta o exemplo "Is 'watermelon' a melon?", perfeitamente ilustrativo em inglês, mas que perderia todo sentido se traduzido literalmente. Em português, um exemplo equivalente poderia ser: " 'árvore de conhecimento' é uma 'árvore'?".

[26] Tradução da Wikipedia do documento inglês: <http://en.wikipedia.org/wiki/Semantic_MediaWiki> 21 abr2009.

Neste ponto, são deixadas de lado as citações de opiniões, pois poder-se-ia pensar que dá para perceber que, mesmo aumentando a amostra mais e mais, não será possível estabelecer fronteiras e limites entre as mais ou menos ilusórias 'webs', que permitiriam dizer se tal ou qual aplicação ou serviço se enquadra com certeza em uma ou em outra web. Pois, olhando bem, cada uma delas se beneficia do que aportaram as anteriores, assim como qualquer uma, em cada momento de sua vida, estará contribuindo de algum modo para o surgimento de novas realizações, de novos sucessos ou de novos fracassos. Assim, parece bom desistir de qualquer explicação *Tim-Tim por Tim-Tim*[27] do que seriam essas webs. Para o(a) leitor(a) que desejar acompanhar esse debate ou aprofundar os detalhes da briga entre Google e Wikipedia, apresentam-se, a seguir, alguns números a título de curiosidade:

- Uma pesquisa no Google, em 7 de maio corrente, com a expressão "the end of google" rendeu 28.100.000 de respostas;
- Outra pesquisa, na mesma data, com a expressão "the end of wikipedia" rendeu 129.000 respostas;
- Outra pesquisa, na mesma data, com a expressão "google versus wikipedia" rendeu 1.480.000 respostas; e
- Outra pesquisa, na mesma data, com os mesmos termos anteriores invertidos, ou seja com expressão "wikipedia versus google" rendeu 285.000 respostas.

Quanto tempo gasto! E ainda não terminou. Será que vale a pena? Não seria melhor, em benefício de todos, juntar esforços, compartilhar de verdade?

Sobre a retirada mal explicada da página "Web 3.0" da versão inglesa da Wikipedia, apresentam-se, a seguir, aos(às) leitores(as) alguns dados para reflexão: Das mais de cem línguas em que publica a Wikipedia foram escolhidas algumas para ver o que dizem sobre a "Web 3.0". Os resultados da pesquisa realizada em maio foram os seguintes:

- Alemão: Remete para Semantisches Web,
- Checo: Destaca a discussão sobre o significado da Web 3.0; remete para Web semântica, microformatos e GoogleDocs.
- Dinamarquês: Remete para Web semântica.
- Espanhol: Breve página; destaca a discussão sobre o significado da Web 3.0; faz referência a bases de dados, inteligência artificial, Web semântica e Web 3D.
- Francês: Breve página apresentada como rascunho; sugere buscar, no Google, WikiWix ou Yahoo!, artigos sobre tecnologia semântica e indexação na Internet.
- Holandês: Não tem página.
- Italiano: Breve página.
- Polonês Destaca a discussão sobre o significado da Web 3.0; remete para Web semântica; remete para Semantic Web, Second Life.
- Português: Breve página; sem referências.
- Sueco: Breve página; atribui a origem da expressão a John Markoff; remete para Web semântica, inteligência artificial, Web 3D.

A julgar pelos resultados, parece que os 'Wikipedianos' pelo mundo afora não gostam muito da Web 3.0.

[27] Justificamos as letras maiúsculas como lembrança de Tim Berners-Lee e Tim O'Reilly.

Antes de retomar o fio da meada, poder-se-ia contar uma breve história que permitiria estabelecer uma relação com a preocupação (ou mania) de dar nomes a tudo e a qualquer coisa, para chamar a atenção, mesmo quando as coisas já tinham nome, e que, aparentemente, nada tem a ver com 'dar nomes aos bois'.[28]

Quem já andou por fazendinhas e fazendas, onde se criam bois, deve ter certamente observado que, com a convivência, acaba se estabelecendo uma relação de amizade entre os animais e os peões. Estes designam e chamam aqueles com nomes carinhosos, como 'Camarada', 'Valentão', 'Ruivo', e assim por diante. E quando o peão se aproxima e brada "Eeeeh, Valentão.., tooo chegaaandu!", o animal olha para ele como se tudo entendesse. Mas não podemos esquecer que, no melhor dos casos, a única resposta que cabe esperar é um sonoro "Mmmmmuuuuuuuuh!!!".

3. WEB SEMÂNTICA

Tim Berners-Lee *et al.* (2001), num famoso artigo publicado em *Scientific American*, apresentou suas idéias sobre o que seria a Web Semântica. Comparando com artigos mais recentes (ver, por exemplo, Wikipedia, 2009), fica visível que as idéias não têm mudado muito. Quanto às implementações...., isso será considerado mais na frente. Segundo Berners-Lee,

> [...] a Web Semântica não é uma Web separada, mas uma extensão da atual, na qual é dada à informação um significado bem definido, o que permite às pessoas e aos computadores trabalharem em cooperação. Os primeiros passos para entrelaçar a Web Semântica na estrutura da Web existente, já estão sendo dados. Num futuro próximo, esses desenvolvimentos deverão conduzir a significativas novas funcionalidades à medida que as máquinas irão se tornando bem mais capazes de processar e "entender" os dados que, por enquanto, se limitam somente a apresentar. A propriedade essencial da *World Wide Web* é sua universalidade. A força do hipertexto é que 'alguma coisa pode ligar-se a outra coisa'. A tecnologia da Web, portanto, não deverá estabelecer diferenças entre rascunhos de mensagens e apresentações cuidadas, entre informação comercial e acadêmica, ou entre culturas, línguas, suportes, etc. A informação varia de acordo com a grande variedade de eixos. Um deles é a diferença entre informação produzida basicamente para o consumo humano e aquela produzida principalmente por máquinas. [...] Até hoje a Web se desenvolveu mais rapidamente como um meio de documentos para pessoas, mais do que para dados e informações que possam ser processadas automaticamente. A Web Semântica quer conseguir isso.
> Como a Internet, a Web Semântica será tão descentralizada quanto possível. [...] A descentralização implica compromissos: a Web deve atingir uma consistência total em suas interconexões [...]

Berners-Lee estabelece algumas condições para que a Web semântica se torne realidade:

- *Representação do conhecimento*. – Para que a Web semântica possa funcionar, os computadores devem ter acesso a coleções de informações estruturadas e conjuntos de regras de inferência que podem "raciocinar" automaticamente. A "representação do conhecimento" – como essa tecnologia é denominada freqüentemente – contém a semente de aplicações importantes, mas ainda, não mudou o mundo. Para mostrar todo seu potencial, precisa se associar a um sistema global simples. Quando o sistema é suficientemente complicado para ser usado, sempre inclui questões que não podem ser respondidas. Por exemplo, como resolver o paradoxo de uma frase como "Esta frase é falsa"? Para resolver problemas como esse, todos os sistemas tradicionais de representação do conhecimento possuem suas próprias regras para fazer inferências a

[28] Lembremos, simplesmente, a tendência a introduzir neologismos parra substituir termos e expressões bem estabelecidos na ciência da informação, tais como 'recursos', no lugar de artigos ou imagens, '*creator*' no lugar de autor, 'arquitetura' no lugar de configuração, estrutura, modelagem, ou... 'ontologia' no lugar de grupo ou conjunto de termos com afinidade semântica.

partir dos dados. Assim, um sistema sobre genealogias, que cria árvores de família, deve incluir a regra "a esposa de um tio é uma tia". Embora os dados possam ser transferidos de um sistema para outro, as regras, geralmente, não podem. O desafio da Web semântica é fornecer uma linguagem que represente tanto os dados como as regras de raciocínio sobre os dados e que permita que as regras de um determinado sistema de representação do conhecimento possam ser exportadas para a Web. Ou seja, agregar lógica à Web.

- *XML e RDF.* – Duas importantes técnicas para o desenvolvimento da Web Semântica já estavam disponíveis quando Berners-Lee escreveu seu artigo: A Linguagem de Marcas eXtensível (*eXtensible Markup Language* – XML) e Marco para Descrição de Recursos (*Resource Description Framework* – RDF). XML permite a qualquer um criar suas próprias etiquetas (*tags*) – etiquetas escondidas como as usadas para 'anotar' as páginas Web ou partes de texto na página. Programas ou '*scripts*´ podem ser usados de formas sofisticadas, mas quem escreve o *script* deve conhecer para que usa as *tags* o autor da página. XML permite que os usuários acrescentem livremente estruturas aos seus documentos, mas não indicam nada sobre seu significado.
  O significado se expressa mediante RDF, que o codifica em grupos de ´tríades´ (*triples*) – frases elementares formadas por sujeito, verbo e objeto. As tríades podem ser escritas usando *tags* XML. Em RDF, as tríades do documento fazem *assertivas* sobre coisas particulares (pessoas, páginas Web ou quaisquer coisas) que têm *propriedades* ("é irmã de", "é autor de") com certos *valores* (outra pessoa, outra página Web). Essa estrutura se torna a forma natural de descrever a grande maioria dos dados processados pelas máquinas. Cada sujeito e objeto é identificado por um Identificador Uniforme do Recurso (*Uniform Resource Identifier* – URI), usado exatamente como o *link* ou Localizador Uniforme do Recurso (*Uniform Resource Locator* – URL) da página Web. Os URLs são o tipo mais comum de URI. Os verbos também são identificados por URIs, o que permite a qualquer um definir um novo conceito, um novo verbo, simplesmente definindo sua URI em algum lugar da Web. A linguagem humana se desenvolve usando o mesmo termo com significados diferentes, mas a automatização não. As tríades de RTF formam redes de informação sobre coisas relacionadas. Como RDF usa URIs pra codificar essa informação em um documento, as URIs garantem que os conceitos não sejam somente palavras no documento, mas que sejam amarradas a uma única definição que todo o mundo pode encontrar na Web. Por exemplo, imaginemos que temos acesso a várias bases de dados que contêm informação sobre pessoas, inclusive seus endereços Para localizar pessoas que moram numa determinada localidade, identificável pelo código de endereçamento postal (CEP), precisamos conhecer que campos identificam, em cada base, os CEPs e os nomes. RDF pode fazer isso mediante a tríade: "[o campo 9999 na base ABCD] [é um campo do tipo] [CEP]", e usando URIs para cada termo da tríade.

- *Ontologias.* – Como duas bases podem utilizar diferentes identificadores para o mesmo conceito, é preciso se servir de um programa que possa combinar ou comparar informações das duas bases de dados, "sabendo" que os dois termos significam a mesma coisa. A solução, para que o programa possa "descobrir" o significado comum, seriam, teoricamente, as coleções de informações denominadas "ontologias", termo cooptado da Filosofia pelos pesquisadores da Web semântica com um sentido que nada tem a ver com o estudo do ser. No jargão da Web semântica, ontologia é um documento ou arquivo que define formalmente relações entre termos. O tipo clássico é uma *taxonomia* e um conjunto de *regras inferenciais*. A taxonomia define classes de objetos e relações entre eles. Por exemplo, um *endereço* pode ser definido como um tipo de *local*, e os *CEPs* podem ser definidos como sendo aplicáveis exclusivamente a locais, etc. *Classes*, *subclasses* e *relações* entre *entidades* são poderosas ferramentas para o uso da Web. Pode-se expressar um grande número de relações entre entidades, atribuindo propriedades às classes e permitindo que as subclasses herdem essas propriedades.

- *Agentes.* – Outra questão importante é o uso de assinaturas digitais, que são blocos criptografados que os computadores e *agentes* podem utilizar para verificar que a informação provém de uma fonte confiável. Os agentes desconfiarão das assertivas lidas na Web semânticas, até checar as fontes de informação. Muitos serviços na Web existem sem semântica, e os agentes não têm como localizar entre eles algum que seja capaz de realizar uma função específica. Esse processo, denominado descoberta de serviços (*service discovery*), só pode acontecer quando existe uma linguagem comum para descrever o serviço, permitindo a outros agentes "entender" a função oferecida e como se servir dela. Agentes produtores e consumidores podem chegar a um entendimento compartilhado, intercambiando ontologias que fornecem o vocabulário necessário para uma negociação. Os agentes podem dar o impulso inicial a novas capacidades de "raciocínio" para a descoberta de novas ontologias. A semântica pode, também, facilitar o aproveitamento de um serviço que somente satisfaz a demanda parcialmente. Um processo típico implica a criação de uma "cadeia de valor", na qual subconjuntos de informação passam de um agente para outro, cada um "agregando valor", para construir o produto requerido pelo usuário final.
  Os primeiros passos já foram dados, e normas e recomendações estão sendo definidas para descrever a capacidade funcional dos instrumentos. Mais na frente, as linguagens tornar-se-ão mais versáteis para lidar com ontologias e lógica. A Web semântica não é uma simples ferramenta para realizar tarefas individuais. Ela foi concebida especificamente para, se desenhada de forma adequada, sustentar o desenvolvimento do conhecimento humano.

As considerações acima foram retiradas, resumidas e traduzidas a partir do referido artigo de Berners-Lee *et al*. (2001). A seguir são extraídos alguns elementos do artigo da Wikipedia (2009), também referido acima, que apresenta a Web semântica como uma extensão decorrente da evolução da *World Wide Web*, na qual a semântica da informação e os serviços da Web encontram-se definidos, tornando possível que a Web entenda e satisfaça as solicitações de pessoas e máquinas para se servir dos conteúdos da rede. Deriva da visão de Tim Breners-Lee, diretor do *World Wide Web Consortium* sobre a Web como um meio universal para intercâmbio de dados, informação e conhecimento. A Web semântica compreende um conjunto de princípios de '*design* ' [29], grupos de trabalho colaborativos e tecnologias de implementação. Alguns elementos da Web semântica são ainda possibilidades prospectivas, a serem implementadas. Outros elementos já têm especificações formais. Tais como o Marco para Descrição de Recursos (*Resource Description Framework* – RDF), vários Formatos de Intercâmbio (*RDF/XML, N3, Turtle, N-Triples*) e notações como o Esquema RDF (*RDF Schema* – RDFs) e a Linguagem de Ontologias Web (*Web Ontology Language* – OWL), todos eles visando a fornecer uma descrição formal de conceitos, termos e relações, dentro de um domínio de conhecimento. Uma relação mais completa das normas e recomendações estabelecidas até o momento de redigir o presente Capítulo encontra-se em *World Wide Web Consortium* (W3C, 2009)

O acima referido artigo da Wikipedia (2009) recolhe uma citação que lembra a visão original da Web semântica expressa pelo próprio Berners-Lee (BERNERS-LEE; FISCHETTI, 1999):

> Tenho um sonho para a Web [em que os computadores] se tornam capazes de analisar todos os dados da Web – conteúdo, *links* e transações entre pessoas e computadores. A 'Web Semântica', que fará com que isso seja possível, ainda não surgiu, mas quando o faça, o dia-a-dia do comércio, da burocracia e de nossa vida diária serão tratados por máquinas que falam com máquinas. Os 'agentes inteligentes' que o pediram durante séculos, finalmente se materializam (BERNER-Lee, 1999).

[29] Ver: World Wide Web Consortium (W3C, 2007). '*Design Issues: Architectural and philosophical points*'.

Os componentes da Web semântica compreendem as normas e ferramentas para XML, Esquema XML, Esquema RDF e OWL que estão organizados numa arquitetura própria denominada 'Torta (em camadas) da Web Semântica' (*Semantic Web Layer Cake*, também denominada *Semantic Web Cake* ou *Semantic Web Stack*) repesentada, na sua forma já clássica, na *Figura 1*.

Figura 1.- A torta da Web semântica (Wikipedia, 2009)

As funções e relações entre os componentes encontram-se em *World Wide Web Consortium* (W3C, 2004). Segue um resumo desses conceitos:

- *XML* fornece uma sintaxe elementar para estruturar o contexto nos documentos, e ainda associa não-semântica com o significado do conteúdo;
- *XML Schema* é a linguagem que fornece e delimita a estrutura e o conteúdo dos elementos contidos nos documentos XML;
- *RDF* é uma linguagem simples para expressar *modelos de dados* que se referem a objetos (*recursos*) e a suas relações. Um modelo baseado em RDF pode ser representado em uma sintaxe XML;
- *RDF Schema* é um vocabulário para descrever propriedades e classes de recursos baseados em RDF, com semânticas para hierarquias generalizadas dessas propriedades e classes;
- *OWL* acrescenta mais vocabulário para descrever propriedades e classes, tais como: Relações entre classes (p.ex.: disjunção), cardinalidade (p.ex.: "exatamente um"), igualdade, tipificação mais detalhada de propriedades, características de propriedades (p.ex.: simetria) e classes enumeradas;
- *SPARQL* é um protocolo e linguagem de busca para fontes de dados da web semântica.

As especificações em curso de elaboração são:

- *Regra do Formato de Intercâmbio* (*Rule Interchange Format* – RIF). como na "Camada Regras" da "Torta da Web semântica" (*Fig.1*). A intenção é reforçar a usabilidade e a utilidade da Web e de seus recursos interconectados, mediante:
  - *Servidores* que mostram os sistemas de dados existentes usando as normas RDF e SPARQL. Existem numerosas aplicações para realizar a conversão em RDF.
  - *Documentos "marcados"* com informação semântica (uma extensão das <meta> *tags* da HTML, usadas nas páginas da Web atual para fornecer informações às máquinas de busca que usam "*web crawlers*"[30]. Isso poderia ser informação, que a máquina pode entender, sobre o conteúdo do documento, que as pessoas podem entender (como autor (*creator*), título, descrição, etc, do documento) ou, simplesmente, metadados que representam fatos, ou qualquer coisa, como, por exemplo, os recursos e serviços oferecidos por um *site*. (Note-se que qualquer coisa' pode ser descrita com uma URI (*Uniform Resource Identifier*, de forma que a web semântica pode 'raciocinar' sobre animais, pessoas, localidades, idéias, etc.) A 'marcação' semântica, pode, freqüentemente, ser gerada de forma automática;
  - *Vocabulários de metadados* correntes (*ontologias*) e *mapas* entre vocabulários que permitem aos autores do documento conhecer como "marcar" seus documentos, de forma que os *agentes* possam utilizar a informação fornecida pelos metadados (assim. o *Autor* no sentido de Autor da página não poderia ser confundido com o *Autor* de um livro objeto de uma resenha);
  - *Agentes automáticos* que realizam ações para os usuários usando os dados da Web semântica;
  - *Serviços baseados na Web* (freqüentemente com agentes próprios), para fornecer informação específica aos agentes, (p.ex.: um serviço de confiabilidade onde o agente possa 'indagar' se um repositório em linha tem qualidade e é seguro).

4. APLICAÇÕES REAIS DA WEB SEMÂNTICA NA CIÊNCIA DA INFORMAÇÃO EM SENTIDO AMPLO

As idéias até aqui expostas fornecem – espera-se – uma visão suficiente do que é a Web semântica e para onde apontam seus futuros e ambiciososrumos. As opiniões entre os especialistas continuam ainda divididas. Alguns acham que a Web semântica já está aí, que basta aplicar o que já

[30] São programas de computador que percorrem a Web automática e metodicamente. Também chamados *ants* (formigas), *Web spider* (aranhas da Web), *Web robots* (robôs da Web), *etc.*

temos ao nosso alcance para pular a um paradigma superior que tornará corriqueiras as tarefas que conduzem a esse mundo de informação globalizada e a todos acessível. Há, também, os eternos pessimistas obsessivos que dizem, torcendo o nariz: "*Muito bonito, mas isso não é para amanhã, se é que algum dia chega...!*" A opinião do autor é um prudente meio termo – os latinos diziam '*In medio virtus*' –, com certa inclinação para o lado dos otimistas, dos sonhadores, dos visionários. Não temos visto já tantas coisas acontecerem que pareciam sonhos impossíveis? Trabalhando de forma colaborativa, aberta, e acreditando, é possível chegar.

Fica para os(as) leitores(as) o cuidado, se assim o desejarem, de aprofundar o assunto. Pode-se esperar que as numerosas referências, e os *links* que as acompanham, facilitem seu trabalho.

Nesta Seção, será mostrada com exemplos concretos, reais, qual é a situação da Web semântica e onde está sendo aplicada já com proveito. Para tanto, será dirigida a atenção para áreas onde a informação é o foco – e/ou o diferencial –, que se dividirá em *'predominantemente não documentárias'* (as que mais têm se desenvolvido com base na Web semântica) e '*predominantemente documentárias*' (nas quais caberia depositar as maiores esperanças). Deve-se esclarecer, enfim, que o termo *documento* será usado em sentido amplo, como '*objeto informacional*" materializado em quaisquer suportes, que também se poderiam chamar, aderindo à tendência, de '*recurso*', quando for o caso.

4.1 Situação atual em áreas predominantemente não documentárias

Diversas tecnologias próprias da Web semântica vêm sendo aplicadas com sucesso em algumas áreas que, em princípio, não são reconhecidas como propriamente integrantes do que no Brasil se entende geralmente – numa visão excessivamente restritiva – por 'Ciência da Informação'. Dentre essas áreas, foram escolhidos alguns exemplos já testados com profundidade diversa (do estudo de caso ao projeto em grande escala), que atestam de sua viabilidade. Do capítulo introdutório de uma recente obra, intitulada "*Semantic Web for Business*" (GARCIA, 2009), foram extraídas as seguintes linhas, que representam o espírito e o teor da volumosa obra:

> O objetivo da cooperação Indústria-Pesquisa é estabelecer uma relação de trabalho entre os pesquisadores da Web Semântica e os parceiros da indústria, visando acelerar a transferência dos resultados obtidos pelos primeiros em áreas em que podem ser aplicados como protótipos pelos industriais, no quadro dos casos escolhidos para estudo. A cooperação não busca somente realizar uma história de sucesso, em termos de uma pesquisa específica e de um negócio determinado, mas também de estabelecer o valor das tecnologias da Web Semântica nas aplicações industriais, no sentido mais amplo [...], resolvendo problemas de negócios e assegurando a aplicabilidade futura ao orientar os pesquisadores na direção das exigências reais do trabalho industrial.

Das nove áreas escolhidas nessa obra, os casos correspondentes aos provedores de tecnologia, à indústria de serviços, fármacos e saúde, mídia e comunicação, e o setor público e o governo representam 77 %, enquanto os restantes (transporte e logística, indústria automobilística, energia, e indústria de alimentos) representam 24 %. Um número importante dos casos estudados na Alemanha, Áustria, Bélgica, Chipre, Espanha, Estados-Unidos, Finlândia, França, Grã-Bretanha, Itália e Portugal, contou com o financiamento da Comunidade Européia, via os *Information Society Technologies – Framework Programs* (IST-FP5 [2001-2004], IST-FP6 [2004-2008] e IST-FP7 [2007-2013]). (LEGER *et al*. *Apud* Garcia, 2009).

Em outras regiões geográficas e na mesma época, dedica-se igualmente uma atenção crescente a pesquisas e projetos que envolvem parcerias público-privadas, com foco em aplicações práticas de tecnologias da Web semântica, freqüentemente, também com o apoio de órgãos de governo e/ou de fortes entidades financiadoras. Alguns exemplos:

- *Aplicação das tecnologias da Web semântica ao concerto de carros.* Projeto de pesquisa "MYCAREVENT", da Comissão Européia que reúne fabricantes de veículos, empresas de reparos, fabricantes de ferramentas para diagnóstico e especialistas em TI incluindo especialistas em Web Semântica (portal de serviço, terminologia e ontologia comuns). (BRIAN; MARTIN *apud* GARCIA, *op.cit*. 2009).

- *Abordagem integrativa de informação baseada na Web Semântica, para o mercado eletrônico com base em agentes.* – Se usam agentes de software para dar suporte a fornecedores e clientes na venda de produtos e serviços, num quadro de crescimento do e-comércio (integração da informação com base em ontologias, exploração do paradigma de mapeamento de ontologias, alinhando as necessidades dos consumidores e a capacidade do mercado, em modo semi-automático com base na tecnologia da web semântica; relações de confiança captadas pelas redes sociais). (VIAMONTE; SILVA *apud* GARCIA, 2009).

- *Web Semântica para Convergência de Mídias.* – Convergência de mídias, em 'casas' de mídias jornalísticas, que estão evoluindo para meios de comunicação altamente dinâmicos e multicanal, onde convergem os diferentes canais em um escritório de edição de notícias unificadas (construção de um marco ontológico baseado no jornalismo e nas normas multimídia existentes e traduzir os metadados atuais para a Web semântica). (PERDRIX *et al. Apud* Garcia, 2009).

- *Experiências na construção de serviços de e-negócios para telefonia móvel.* – As operadoras de telecomunicações podem melhorar suas redes compartilhadas e oferecer novos serviços resolvendo as limitações das arquiteturas orientadas para o usuário (*Service-Oriented Architectures* – SOA). Uso de tecnologia de serviços web para criar federações de provedores de serviços e ontologias para suportar de mecanismos avançados de compatiblização baseados em depósitos de metadados semânticos. Proposta de criação de uma Agência de Registro de Serviços que controla e reforça as políticas de anotação, garantindo a uniformidade das políticas de descrição dos serviços. (De FURIO *et al. Apud* Garcia, 2009).

- *Uso de ontologias em rede para tratar domínios de conhecimento intensivo: um estudo de caso no setor farmacêutico.* – Herrero Cárcel, citado por Garcia (2009) mostra como as ontologias em rede satisfazem as necessidades de um setor de conhecimento intensivo, como é o setor farmacêutico. O uso de semânticas ajuda a fazer a ponte entre as diversas representações de diferentes '*stakeholders*', resolvendo o problema que surge quando as ontologias usadas para modelar o domínio se tornam grandes demais e impossível de gerenciar. O cenário estudado é a nomenclatura de produtos no setor farmacêutico, visando o estabelecimento de uma ontologia de referência comum para os fármacos. Outro exemplo interessante nesse setor é o projeto *NextBio*: uma base que consolida dados experimentais de ponta, 'tagados' (*tagged*), das ciências da vida e os conecta via ontologias biomédicas, e os torna acessíveis mediante uma interface de busca. Agrupa grandes laboratórios em parceria com a Universidade de Stanford (DAVIES, 2009).
  Ainda, outro projeto interessante se refere ao desenvolvimento de bases de dados relacionais baseadas em tecnologias da Web semântica a partir de um conjunto de ferramentas, atualmente em uso na gestão de bases relativas à *Medicina Tradicional Chinesa*, heterogêneas e dispersas, que se integram mediante ontologias (HUAJUN CHEN *et al*., 2006)

- *Redes sociais.* – A popular aplicação de Web semântica "*Friend of a Friend*" (FoaF) descreve relações entre pessoas e outos agentes, em termos de RDF (WIKIPEDIA, 2009).
  O Projeto *SIOC* (*Semantically-Interlinked Online Communities*) fornece um vocabulário de termos e relações para modelagem dos espaços web (SIOC Project, 2009). Entre outros exemplos desses espaços de dados: foros de discussão, *blogs*, listas de endereços, galerias de imagens*, twitters*, etc.
- *Outros projetos multi-setoriais e multidisciplinares.* – Dentre os numerosos exemplos que poderiam ser citados – não serão contempladas aqui aplicações nas áreas jurídica, de e-governo, etc. que, por seus possíveis desdobramentos tornariam este trabalho quase interminável – foram

escolhidos dois: o Projeto *Linking Open Data* e Projeto *NEPOMUK.* O primeiro, *Linking Open Data Project* (2008) é um esforço comunitário que visa criar conjuntos de dados RDF interligados, de forma aberta e acessível, procedentes de uma ampla variedade de fontes. O objetivo é facilitar a publicação na Web dos dados de pequenas e médias aplicações. Beneficia-se do suporte do Grupo de Interesse *Semantic Web Education & Outreach* – SWEO[31], do *World Wide Web Consortium* – W3C (WIKIPEDIA, 2009). O Projeto "Ambiente em Rede para Gestão Unificada do Conhecimento baseado em Ontologias" **(***Networked Environment for Personalized, Ontology-based Management of Unified Knowledge – NEPOMUK*) reúne pesquisadores, desenvolvedores de *softwares* para a indústria e usuários industriais representativos, para desenvolver soluções abrangentes que permitam estender o '*desktop*' pessoal em um ambiente colaborativo que suporte tanto a gestão da informação pessoal, como o compartilhamento e intercâmbio através de relações organizacionais e sociais (NEPOMUK, 2007). O projeto foi encerrado em dezembro de 2008, após três anos de atividade, tendo sido cumpridos seus objetivos: construção da arquitetura, definição das ontologias e introdução do conceito de 'fruição' ao tempo que se facilita a integração da tecnologia com diversas aplicações de '*desktop*' existentes (ZAINO, 2008). Participaram dezesseis entidades acadêmicas e empresariais de sete países (Alemanha, França, Grécia, Irlanda, Itália, Suécia e Suíça).

- *Educação*. – Área em que a utilização de ferramentas da Web semântica se amplia rapidamente. Para uma rápida visão da evolução do uso das *tecnologias Web na educação*, ver Anderson (2008). Quatro anos antes, Anderson e Whitelock (2004) publicavam, no Reino Unido, um artigo, no *Journal of Interactive Media in Education*, no qual escreviam: "A 'rápida e incessante demanda por mudanças está forçando a educação formal e informal a se tornar mais eficaz e eficiente. Ademais, os custos de negligenciar a educação exacerbam a clivagem entre os que têm oportunidades para aprender e os que não têm. [...] A Web Semântica fornece uma visão a longo prazo de oportunidades para prover uma educação que independe das distâncias geográficas, temporais ou econômicas." Parece que um nova consciência das possibilidades oferecidas pelas tecnologias da Web semântica para melhorar a qualidade da educação a aumentar seu alcance, está abrindo caminho. Para não estender mais a explicação do que está se tornando evidente, só outros dois exemplos escolhidos em pontos bem distantes, encerrarão esta Seção: um, no Brasil, refere-se ao Primeiro e Segundo Workshops Brasileiros sobre a Web Semântica e Educação (2007, 2008), o outro é um artigo de Devedzic (2004), professor da Escola de Administração de Negócios, da Universidade de Belgrado. Observa-se uma visão bastante coincidente sobre as temáticas discutidas (linguagens para a Web semântica, arquitetura semântica, ontologias, avaliação das aplicações disponíveis, projetos, normas, sistemas de autoria, objetos educacionais, etc.). Pena que essa visão ainda seja incipiente no Brasil, onde a fratura educacional é vergonhosa e o nível de qualidade do ensino, em todos os níveis, extremamente desigual.

4.2 Situação atual em áreas predominantemente documentárias

Antes de abordar o exame da situação das atividades, serviços e produtos resultantes, nas áreas de Arquivologia, Biblioteconomia, Documentação, Museologia, Gestão da informação e do conhecimento, Inteligência empresarial e competitiva, Recuperação da informação, *Data* e *Text Mining*, Autoria coletiva, etc., em relação a Web semântica, parece oportuno passar em revista algumas práticas, alguns processos e – o que é mais importante

[31] O Grupo de Interesse Educação e Serviço na Web Semântica (*Semantic Web Education and Outreach* – SWEO), do W3C, encontra-se atualmente fechado. Seu `contrato oficial expirou em 31 de março de 2006. Note-se, entretanto, que algumas de suas atividades foram mantidas, mas agora funcionam independentemente. Para maiores detalhes, ver <http://www.w3.org/2001/sw/sweo/>.

– alguns conceitos e princípios, que vêm sendo usados desde priscas eras, que se aperfeiçoaram e se ampliaram no decorrer dos anos – ou séculos? – e que reaparecem agora, em ambiente Web, com outros nomes, apresentados com novas linguagens e representados, ordenados, processados, transmitidos e aplicados com novos códigos. Todas essas práticas se enquadram nos paradigmas de uma Ciência da Informação revisitada e expandida, como também se enquadrariam, pensando um pouco, as práticas e tecnologias dos ambientes Web.

Querem ver?

O *Quadro 1* – que não pretende ser completo – coloca em paralelo termos da Ciência da Informação, que poderia ser chamada de clássica, embora ainda não tenha saído da adolescência, e termos da Web Semântica, que também parece estar bastante longe da maturidade.

Quadro 1. Comparação de alguns termos e expressões da Ciência da Informação e da Web semântica

Web / Web semântica	**Ciência da Informação**
Agentes de recomendação	Políticas de aquisição; Lei de Bradford
Ambigüidade / desambiguação	Sinonímia; polissemia. (nos tesauros: *use*; *veja também*) [para tal termo *use* tal outro]
Compartilhamento	Catálogos coletivos
Formato de comunicação e intercâmbio de dados	Norma ISO 2709, velha de mais de 40 anos, é atualizada periódicamente
Identificadores / Localizadores (URLs, URIs...)	Número de chamada, etc.
Inferência	Curvas estocásticas sobre séries históricas.
Interoperabilidade	OCLC, etc.
Linguagens de marcas; hyperlinks	Remissivas
Metadados	Dublin Core (metadados e qualificadores)
Ontologias (pobres filósofos se remexendo nos seus túmulos seculares !)	Clusters temáticos; métricas da informação
Open systems/ Sistemas abertos	OPACs
Parsers	Indexação automática
Relações entre conceitos	Descritores compostos; adjacência; proximidade
Reuso	Pesquisa bibliográfica; este trabalho
Reverse file	Arquivo invertido
Tags	Etiquetas; *tags* ; campos de dados; Unesco/Unisist; CDS/ISIS; MARC
Taxonomias	Sistemas de classificação; tesauros
Template	Planilha/Folha de entrada

Enquanto isso, cabe indagar: Como estão as aplicações da ciência da informação, em sentido amplo, predominantemente documentárias?

Antes de responder, convêm fazer uma distinção entre bibliotecas gerais e bibliotecas especializadas. Entre as primeiras incluem-se as bibliotecas públicas de qualquer dimensão e, entre as segundas, cabe distinguir aquelas em que a especialização é temática (ciências da vida, ciências sociais, ciências exatas, humanidades, engenharias, legislação, etc.) daquelas que atendem especial e especificamente uma comunidade de usuários claramente diferenciadas, cujo exemplo mais claro seriam as bibliotecas escolares, cuja atenção se volta para um público infanto-juvenil.

As bibliotecas acadêmicas que atendem estudantes, docentes e pesquisadores são – ou deveriam ser –, ao mesmo tempo, um aglomerado de bibliotecas setoriais especializadas (por temas e por comunidades de usuários).

Essa 'classificação' – da qual é possível que muitos profissionais bibliotecários não gostem – visa a destacar algumas diferenças entre umas e outras bibliotecas, que influenciam cada dia mais a forma como devem ser administradas, para o maior benefício de seus usuários.

Ainda, aquelas bibliotecas que visam a oferecer serviços em áreas específicas, possuem características próprias quando se trata de processar e organizar de maneira otimizada as informações oferecidas aos seus usuários. Com efeito, embora quando, como todas as outras, devam cuidar de catalogar, classificar e ordenar seus volumes, seguindo normas bem estabelecidas, deveriam também prestar uma atenção especial à multiplicação e ao aprofundamento dos pontos de acesso que permitirão chegar aos volumes que contêm a informação desejada.

Imagine-se uma obra intitulada "Manual Prático de Economia Agrícola". Segundo seja escolhida como entrada principal 'Agricultura' ou 'Economia', será colocada na estante de 'Agricultura' ou de 'Economia', de tal forma que, embora o usuário tenha pretendido estudar o livro, as chances de encontrá-lo física e realmente, num sistema de estante aberto, são somente 50%. Se o título tem um subtítulo, por exemplo, "em clima tropical", como as palavras 'clima' e 'tropical', podem ser também consideradas entradas principais, o livro corre o risco de cair na estante de 'Meteorologia' ou 'Geografia', com o qual, segundo a escolha da 'cota', a probabilidade de ser encontrado cairia para 25%. A isso se deve acrescentar, a conveniência – que deveria se tornar exigência – de se fazer a descrição física e temática dos artigos que publicados nas coleções de periódicos e revistas, das comunicações nos volumes das atas dos congresso, dos capítulos, nas obras de autoria coletiva (o que se chama de catalogação analítica).

Senão, como ter acesso a informação contida nesses materiais? Todos os estudantes, professores e pesquisadores têm que folhear e manusear a enxurrada de novos periódicos, atas de congressos, livros de autoria coletiva que chegam todos os meses, aos centos, numa biblioteca central universitária? E então, quem tem tempo para estudar, para pesquisar? Por isso e outros detalhes que não serão discutidos aqui, as bibliotecas especializadas – e as gerais, que têm setores específicos (arte, literatura, etc.), se converteram desde os anos 50-60 do século passado em centros de documentação e informação especializados, praticando a catalogação analítica e a indexação em profundidade das unidades documentárias, se possível com bibliotecários também especializados, ou com profissionais da área, ou com indexadores profissionais.[32]

Com certeza, algum(a) leitor(a) dirá, encolhendo os ombros, ao ler os parágrafos anteriores: "E eu com isso, se já temos bases de dados de tudo, enciclopédias virtuais, motores de busca fantásticos?" Mas..., cuidado; as coisas não são bem assim. Quem coloca os registros bibliográficos nas bases de dados que podemos acessar na Internet? Quem realiza a indexação em profundidade? Quem paga os(as) indexadores(as) especializados(as)? É bom lembrar uma evidência que com freqüência se esquece: Se os registros colocados na Internet não são indexados em profundidade, com pertinência, se os metadados, qualificados ou não, que abrem o caminho para a interoperabilidade, não são corretos ou suficientes, se os resumos – de preferência em duas línguas – são vagos ou incompletos de forma que os analisadores automáticos de conteúdos, não podem descobrir novos termos que se somem

[32] É oportuno lembrar a existência de associações de bibliotecários especializados em diversos países, como por exemplo a *Special Library Association* – SLA, nos Estados Unidos, em muitos países as associações de indexadores, enquanto membros de uma profissão regulamentada. Não seria a hora de se pensar, no Brasil, sobre o assunto?

aos descritores ou palavras-chave indicados pelos autores e/ou pelos indexadores, a maioria dos registros que deveriam abrir o caminho aos documentos completos digitalizados, reunidos em repositórios acessíveis e acessáveis, ficarão perdidos no espaço virtual, invisíveis, flutuando como os corpos eternamente congelados, no vazio espacial, dos pobres astronautas que acidentalmente se afastaram mais da conta da nave, tentando algum concerto na parte externa, ou reparando algum dispositivo avariado por um meteorito.

A tendência hoje é que as universidades cuidem de seus repositórios destinados a reunir, conservar e tornar acessíveis na Web os registros e/ou documentos, na íntegra, que compõem a produção científica, técnica, artística e cultural da instituição. Pareceria que é nesse sentido que deveriam se orientar as pesquisas no sentido de aproximar a Ciência da Informação – e seus desdobramentos práticos – e a Web semântica. Mas..., convêm reiterar o afirmado acima: Sem uma indexação adequada, ou melhor, sem uma representação física e temática adequada, sem uma profunda análise dos conteúdos, poderão se aperfeiçoar os esforços colaborativos de bibliotecas e centros de documentação, mediante catálogos coletivos e bases de dados comuns ou interconectadas, mas a visibilidade, a interoperabilidade, o acesso aberto, e tudo o que a Web semântica promete para um futuro cada vez mais próximo, não acontecerão.

Outras aplicações da Web semântica, por exemplo, em projetos gerencias, já são possíveis, como foi mostrado na Seção anterior, e nada impediria sua aplicação quase que imediata, uma vez entendido que é preciso aprender a lidar com novas linguagens de representação, com aspectos semânticos aprofundados. Isso é mais fácil, também, porque o número de variáveis em jogo e menor, o universo com que se lida (atores, produtos) e seus limites mais claramente definidos e... mais fáceis de serem entendidos pelos responsáveis pela autorização de aplicar recursos em projetos gerencias que podem trazer, senão lucros, melhor relação custo/benefício. Mas, em se tratando de sistemas de informação documentária de grande porte, que exigem novas habilidades de um pessoal mais numeroso e mais qualificado, quando a influência e o resultado na pesquisa e no ensino não são imediatos, quando os produtos futuros são mais e melhores conhecimentos, a dificuldade de se fazer compreender pelos responsáveis das decisões orçamentárias é maior. Por isso, será necessário construir novos ambientes de colaboração entre as universidades do país e do exterior, com o governo e a indústria que deverão compreender, como demonstrado em outros países, que mais informação e conhecimento significa inovação e progresso, ou seja, mais riqueza.

É obvio que o uso dos sistemas abertos e dos novos protocolos (*Open Archives Initiative Protocol for Metadata Harvesting – OAI-PMH*)[33], aliados ao uso inteligente dos metadados tem dado também um amplo impulso às bibliotecas virtuais, às redes de bibliotecas e sistemas de informação, e aos repositórios documentais, acadêmicos ou não (exemplos: a rede OCLC[34], E-Lis, bem conhecida dos profissionais da informação, a *Ligue des Bibliothèques Européennes de Recherche*[35], a Iniciativa Alemã para Redes de Informação – *Deutsche Iniciative für Netzwerkinformation*[36], a *Bibliothèque Numérique Européenne*[37], e muitos mais[38]). Eis uma prova, embora limitada, de como as bibliotecas, os arquivos, os museus, bem como todas as profissões que derivam ou se relacionam com a ciência da informação, podem, sim, se integrar ao progresso e às mudanças que já são a marca do século XXI.

---

[33] Sobre o Protocolo para Colheita de Metadados da Iniciativa Arquivos Abertos (*Open Archives Initiative Protocol for Metadata Harvesting* – OAI-PMH), ver: <http://www.openarchives.org/pmh/>.

[34] Sobre o Centro de Bibliotecas Computadorizadas em Linha (*Online Conputer Library Center* – OCLC), ver: http://www.oclc.org/us/en/about/default.htm.

[35] Sobre a Liga de Bibliotecas Européias de Pesquisa (*Ligue des Bibliothèques Européennes de Recherche* – LIBER), ver: <www.libereurope.eu/>.

[36] Sobre a Iniciativa Alemã para Informação em Rede (*Deutsche Iniciative für Netzwerkinformation* – DINI), ver: < http://www.dini.de/ueber-dini/>.

[37] Sobre a Biblioteca Digital Européia (*Bibliothèque Numérique Européenne* - Europeana), ver: <http://www.europeana.eu/portal/>.

[38] Durante a redação deste Capítulo surge a notícia do lançamento, na terça-feira 21 de abril de 2009, da Biblioteca Digital Mundial – BNM, que reúne a Europeana e o *Google Book Search*. Para maiores detalhes, ver:< http://www.wdl.org/en>, ou .../en, .../pt (sete idiomas).

A título de informação apresenta-se um brevíssimo resumo de quatro exemplos, em escala mais reduzida que mostram algumas tendências de projetos e pesquisas relacionados com a ciência da informação e seus desdobramentos, e a Web semântica.

- Projeto ***S**emantic **I**nteroperability of **M**etadata and **I**nformation in un**L**ike **E**nvironments* – SIMILE (Interoperabilidade Semântica de Metadados e Informação em Ambientes Diferentes)[39]. Patrocinado pela Fundação Andrew W. Mellon[40], é executado pelas bibliotecas do *Massachussets Institute of Technology* – MIT[41] e pelo Laboratório de Ciência da Computação e Inteligência Artificial (*Computer Science and Artificial Intelligence Laboratory* – CSAIL)[42], também do MIT, busca reforçar a interoperabilidade entre recursos digitais, esquemas, vocabulários e ontologias, metadados e serviços.
- A Iniciativa Rhizomik, do *Grupo de Investigación en Interacción Hombre-Ordenador e Integración de Datos* – GRIHO[43], se inspira na metáfora de Deleuze e Guattari[44], aplicada ao estudo do conhecimento do ponto de vista cientifico, tecnológico e filosófico, em diversos domínios (Web semântica, interação homem-computador, sistemas complexos), como é o caso de uma ferramenta de gestão de conteúdos num contexto de várias mídias (imprensa, televisão, rádio, Internet), com base em tecnologias da Web semântica, visando a produção de conteúdos 'compreensíveis' para as máquinas, sem esquecer que a interação será iniciada e se realizará por e para pessoas (Perdrix *et al*, 2007).
- Uma pista interessante sobre *Museologia na Web* é fornecida por Denise Eler (2008), onde observa, num breve artigo que "Dentre as práticas emergentes utilizadas pelos museus online em países desenvolvidos destacam-se o podcasting, o tagging (classificação de conteúdo por usuários) e a estruturação semântica da informação (Web 3.0)". E conclui: "Se no início do surgimento da web, muitos temiam que o domínio virtual eclipsasse o museu físico, hoje se fala em um museu híbrido, reintegrado por uma variedade de técnicas e tecnologias. Mas o que dizer da situação dos museus brasileiros na internet?" Fornece importantes referências de Gutierrez (2007), Buckland *et al.* (2007) e Van Ossenbruggen *et al.* (2007).
- a área de *Bilioteconomia*, Blattman e Silva (2007) estudam a influência positiva de algumas ferramentas colaborativas da "Web 2.0" sobre certas práticas e serviços bibliotecários e documentários (escrita hipertextual colaborativa, ferramentas wiki, etc.).

---

[39] O foco de *SIMILE* é o desenvolvimento de ferramentas de código aberto (*open source*) que consolidem o acesso, a gestão, a visualização e o reuso dos recursos digitais, por parte dos usuários. Maiores informações em: <http://simile.mit.edu/wiki/SIMILE:About> e <http://simile.mit.edu/>.

[40] A *Andrew W. Mellon Foundation* patrocina atividades em: bolsas em educação superior; intercomunicação acadêmica; pesquisa em tecnologia da informação; museus e conservação de obras de arte ; representações artísticas. Ver: <http://www.mellon.org/>.

[41] As bibliotecas do MIT dão apoio aos programas de ensino e pesquisa do Instituto. Cinco grandes bibliotecas temáticas para planejamento e arquitetura, engenharia, humanidades, ciência, e gestão e ciências sociais, além de cinco bibliotecas especializadas e dos arquivos do Instituto, garantem o acesso a uma ampla gama de materiais, tanto impressos como eletrônicos (cerca de 3 milhões de volumes impressos, 20.000 assinaturas a publicações periódicas correntes, 500 bases de dados em linha, acesso na rede a mais de 30.000 periódicos eletrônicos, material multimídia, etc.). Para maiores detalhes, ver: <http://libraries.mit.edu/about/about.html>.

[42] É o maior laboratório interdepartamental do MIT, com mais de 800 membros, 90 pesquisadores principais de 8 departamentos e 500 estudantes. Sua missão e a inovação em tempo compartilhado, criptografia, redes, arquitetura de hardware, robótica, e fala e visão via computador. Ver: <http://www.csail.mit.edu/node/4>.

[43] Para maiores detalhes sobre outros projetos de pesquisa do grupo espanhol GRIHO, ver, por exemplo: <http://griho.udl.cat/es/recerca/linies.html>.

[44] A reterência original é: DELEUZE, Gilles; GUATTARI, Félix. *Capitalisme et Schizophrénie 1. 'Anti-Œdipe*, Paris: Minuit, 1972 e *Capitalisme et Schizophrénie 2. Mille Plateaux*, Paris: Minuit, 1980.

De 1992 a 1995 o Grupo de Estudo da IFLA sobre os Requisitos Funcionais para Registros Bibliográficos (*Functional Requirements for Bibliographic Records* – FRBRs) desenvolveu um modelo do tipo entidade-relacionamento[45], como uma visão geral do universo bibliográfico, pretendendo ser independente de qualquer código de catalogação ou implementação (Tillet, 2003). Num recente trabalho, Moreno (2007), com base no Relatório Final[46] do Grupo de Trabalho para os Requisitos Funcionais para Registros Bibliográficos (IFLA, 1997), destaca que:

> Os Requisitos Funcionais para Registros Bibliográficos, ao refletirem a estrutura conceitual do que um registro bibliográfico deve conter, fogem do pragmatismo que caracteriza as descrições e discussões sobre as formas de descrição. Isso significa que, ao desconstruir o objeto que será descrito: um livro, um documento eletrônico ou um registro sonoro, por exemplo, propondo um modelo conceitual com entidades, dotadas de atributos, indo além dos elementos de dados previstos na *International Standard Bibliographic Description* – ISBDs e nas *Anglo-American Cataloging Rules – AACRs* , os FRBR trazem as delimitações entre conteúdo e suporte, visando oferecer um maior número de opções ao usuário do catálogo eletrônico, através das relações entre estas entidades.

O modelo apresenta dez *entidades*, divididas em três grupos: O *Grupo 1* compreende *entidades* que são produto de *trabalho intelectual* ou *artístico* (*obra, expressão, manifestação* e *item*). O *Grupo 2* compreende as *entidades pessoa* e *entidade coletiva*, isto é, aqueles responsáveis pelo conteúdo intelectual ou artístico, pela produção física e disseminação, ou pela guarda das entidades do primeiro grupo. O *Grupo 3* compreende um *conjunto adicional de entidades* que servem como *assuntos de obras*, quais sejam: *conceito* (uma noção ou idéia abstrata), *objeto* (uma coisa material), *evento* (uma ação ou ocorrência) e *lugar* (um local). Os *relacionamentos* que ocorrem entre entidades dos FRBRs proporcionam informações adicionais que ajudam o usuário a fazer novas conexões entre uma entidade dada e outras entidades que com esta se relacionam. Esses relacionamentos podem ser entre obras, entre expressões da mesma obra, entre expressões de diferentes obras e entre manifestações, sendo *subdivididos em tipos*, servindo para que o usuário possa navegar pelas entidades que está buscando e por outras relacionadas. Os *atributos*, nos FRBRs, são similares aos elementos de dados do formato MARC[47]. Foram criados de acordo com o tipo de entidade e servem para mostrar as diferenças de conteúdo (intelectual ou artístico). Para responder às necessidades dos usuários, os atributos são definidos em relação às seguintes tarefas genéricas realizadas pelos usuários (*user tasks*) quando fazem buscas em bibliografias nacionais e catálogos de bibliotecas:

- uso dos dados para *encontrar* materiais que correspondam aos critérios estabelecidos para a busca do usuário;
- uso dos dados para *identificar* uma entidade;
- uso dos dados para *selecionar* uma entidade adequada às necessidades do usuário;
- uso dos dados para encomendar, adquirir, ou *obter* acesso à entidade descrita.

---

[45] A expressão entidade-relacionamento, que se encontra ao falar de FRBR, criada por Peter Chen em 1976, foi baseada na álgebra relacional. A proposta dos Diagramas E-R é representar o modelo de dados de uma determinada situação elegendo elementos chamados entidades e descrevendo a maneira como essas entidades interagem entre si. Define-se entidade como aquele objeto que existe no mundo real, com uma identificação distinta e com um significado próprio; qualquer coisa ou situação que possa ser claramente identificada e que possua características próprias e um significado. Em termos de relacionamento, todas as ações ou fatos que descrevem a iteração entre as entidades podem ser definidas como um relacionamento. Pode-se definir relacionamento como o fato, o acontecimento que liga dois objetos, duas 'coisas' existentes no mundo real. Amplamente usada na arquitetura de bancos de dados relacionais (por exemplo, no DSpace, para implementação de repositórios institucionais).

[46] Posteriormente revisto e ampliado. Ver IFLA (2008). Novamente revisto em fevereiro 2009. <http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf>. Consultado abr 2009.

[47] E, também, do formato dos campos dos registros dos sistemas baseados em Unesco/Unisist, como p.ex.: CDS/ISIS.

Pode-se também consultar com proveito o trabalho de Talavera Ibarra (2005) sobre o tema. A versão mais recente – corrigida e ampliada – sobre os FRBRs encontra-se no Relatório Final do Grupo de Trabalho sobre FRBRs, publicada em fevereiro de 2009 (IFLA, 2009), como indicado acima. Numerosos estudos de compatibilidade entre MARC e FRBRs foram realizados sob os auspícios da *Library of Congress* (2003).

A 'Descrição e Acesso ao Recurso' (*Resource Description and Access* – RDA), que pretende ser o novo padrão que virá a substituir a AACR2. Ainda em estudo, o novo padrão viria a representar uma evidente mudança de direção, apontando para o usuário e para a informação que este precisa (OLIVER, 2007). Embora as AACRs tenham sido traduzidas em várias línguas e adotadas em numerosos países, somente quatro países anglo-saxônicos (Austrália, Canadá, Estados Unidos e Reino Unido) integram o *Joint Steering Committee for Development of RDA* – JSC/RDA[48], que estuda o desenvolvimento desse padrão (GÓMEZ, 2007). Coyle e Hillmann (2007) descrevem as tensões que existem entre aqueles que defendem uma mudança incremental do processo de catalogação e os que desejam a inclusão decidida das bibliotecas na era digital e, uns meses mais tarde, a existência dessas tensões é corroborada numa breve reportagem intitulada "Controvérsias na catalogação", publicada.em *American Libraries* (2007). No Relatório do Grupo de Trabalho da *Library of Congress* sobre o 'Futuro do Controle Bibliográfico' (*On the Road*, 2008) se recomenda a suspensão dos trabalhos do JSC relativos ao RDA, assim como a integração com os trabalhos em desenvolvimento conjunto pela LC, DCMI, relativos aos FRBR, ao 'modelo de resumo' do DCMI e às tecnologias apropriadas da Web semântica. No mês de outubro de 2008, o JSC/RDA publicou a segunda minuta do documento "Descrição e Acesso ao Recurso", onde apresenta os objetivos e princípios para o desenho de RDA e Funcionalidade dos registros produzidos usando RDA (JSC/RDA, 2008). Em março de, 2009 a Biblioteca do Congresso propôs uma metodologia para realizar um teste conjunto das três Bibliotecas Nacionais dos Estados Unidos (*Library of Congress* – LC, *National Library of Medicine* – NLM e *National Agricultural Library* – NAL para testar a factibilidade e o custo do padrão RDA (Libray of Congress, 2009).

De todo o exposto pode-se deduzir que a catalogação encontra-se numa encruzilhada, sem achar o rumo para um melhor aproveitamento das tecnologias da Web e, mais particularmente da Web semântica, enquanto outros setores já se beneficiam de suas potencialidades e... dos lucros que as acompanham. Isso se traduz, inevitavelmente, numa inquietude e num desânimo crescente, acompanhado, com mais freqüência do que seria desejável, de uma diminuição progressiva dos resultados, para o qual, também, a enxurrada de documentos a processar em nada ajuda. Mey (2005) observa:

> [...] deparo-me agora, novamente, com proposições do tipo " a catalogação morreu"; "regras não fazem sentido", "nada disso é importante face às novas tecnologias", "metadados são a solução", entre outras variações. Ao mesmo tempo, diversos cursos de Biblioteconomia abriram concursos, nos últimos, dois anos, para a área de representação, nem sempre conseguindo docentes com mestrado, quem dirá doutorado. Parece que a área não desperta grandes interesses....

E continua questionando:

> [...] as centenas de documentos existentes na Internet sobre [as descrições bibliográficas] permanecem ocultos em suas línguas originais, inclusive os FRBR. Por isso, a pergunta: será que a catalogação realmente feneceu no Brasil? [...] Em síntese cabe questionar; por que ensinar ainda a catalogação? Que catalogação deve ser ensinada? Não se trata de perguntas retóricas: procuram-se respostas.

Semelhantes questionamentos se estendem até hoje. Num evento realizado em 22 de junho de 2007, que reuniu bibliotecários dos Estados Unidos e Canadá[49] durante a *'ALA Annual',* para discutir o futuro da catalogação, sob os auspícios da *Association for Library Collections & Technical Services /*

[48] Ver: *Joint Steering Committee for Development of RDA* - JSC/RDA. <http://www.collectionscanada/jsc/rda/html>. Acesso: abr 2008.

[49] Ver, por exemplo, a integração do MARC com XML em: http://www.loc.gov/standards/marcxml//, e os desdobramentos com o *Metadata Object Description Schema* (MODS) em: http://www.loc.gov/standards/-mods//, e com o *Metadata Authority Description Schema* (MADS) em: http://www.loc.gov/standards/mads/

Committee on Education, Training & Recruitment for Cataloging (ALCTS/CETRC), a *Library of Congress* e a *Catholic University of América*, foi fortemente destacado o foco das sessões: "[...] a desconexão entre o que se aprende e o que se ensina nos cursos de biblioteconomia e a realidade do trabalho que se exige no mundo real dos catalogadores." Vale a pena recolher algumas considerações apresentadas por Karen Calhoun, extraídas do seu relatório dirigido à *Library of Congress*, em 2006, intitulado *The Changing Nature of the Catalog and its Integration with Other Discovery Tools*:

> Os resultados das entrevistas estruturadas e da análise da literatura [...] indicam que os catálogos das bibliotecas atuais são pródigos em problemas e parcos em benefícios para os usuários. O custo-benefício da tradicional prática da catalogação está sob fogo cruzado. A roupagem mais típica do catálogo das bibliotecas de pesquisa é o apoio ao controle de inventário e como tecnologia do último momento para possibilitar a entrega dos materiais da biblioteca nas mãos dos usuários. A nova tecnologia para expandir o modelo de serviços do catálogo, de forma a cobrir uma parte maior do universo de informação acadêmica – metabusca – gerou muitas esperanças, mas não está satisfazendo as expectativas iniciais de aglutinar a fragmentada paisagem dos recursos informacionais acadêmicos (CALHOUN, 2006).

Ainda acrescenta:

> Os catálogos em linha foram um produto de sucesso. Como outros produtos, percorreram seu ciclo de vida. Nos últimos anos da década de 70 do passado século, eram uma raridade e sua adoção arrastou-se longamente. Dez anos mais tarde, os catálogos em linha já tinham decolado, e toda biblioteca que se respeitasse deveria ter um. Em 1995, os catálogos em linha já haviam amadurecido. Um novo ciclo de mudança começou, o qual já está praticamente completo. O produto da substituição tomou a forma de uma migração dos catálogos baseados em texto para sistemas baseados em tecnologias cliente-servidor e bases de dados relacionais. Hoje, o catálogo em linha está perdendo seu charme para os estudantes e para os professores e pesquisadores da academia.

E conclui:

> O uso do catálogo, em decadência frente a outras ferramentas de busca e recuperação de recursos informacionais, podem afundar em breve. [...] As bibliotecas especializadas em pesquisa tem investido e continuam a investir milhões para desenvolver e manter a sua capacidade de produzir catálogos locais – em 2004, as bibliotecas americanas de pesquisa (ARL libraries) gastaram aproximadamente 239 milhões de dólares, somente em serviços técnicos. À medida que a procura da informação se volta cada vez mais para os motores de busca, os dirigentes das bibliotecas de pesquisa precisam examinar os caminhos que podem trazer de volta sua capacidade para produzir catálogos em linha à altura de sua procura (CALHOUN, 2006).

Precisa dizer mais?[50]

A ciência da informação tem muito a contribuir para tirar a biblioteconomia e áreas afins – bem como a '*Library and Information Science (LIS)*' – do ´buraco negro' conceitual em que hoje se encontram aprisionadas, sem encontrar o caminho que levaria a trazer o imenso conhecimento escondido em catálogos e bibliotecas à luz da Web semântica, via metadados, taxonomias e ontologias, além das ferramentas pertinentes.

Os leitores, que acompanharam o até aqui exposto, são convidados a refletir sobre o que poder-se-ia chamar de 'Grande Fluxo Informacional' (*Figura 2*) interrompido pela demora de implementar já a junção dos FRBRs com os RDFs, criando o buraco negro que ameaça engolir as bibliotecas, que de virtuais passariam a ser invisíveis para a Web3 / Web semântica.

[50] Para quem deseje corroborar o desconforto existente com outras leituras, pode-se recomendar a leitura de Cyle e Hillmann (2007, *op. cit.*), Baptista (2007), Alvarenga (2003), Calhoun (2006, *op. cit.*), Beacon (2007), e Hall-Ellis (2007).

Figura 2 . – O 'Buraco Negro' do Grande Fluxo Informacional (ROBREDO, 2009).

O pior é que há bibliotecários 'tradicionais' que teimam em aplicar as novas tecnologias (hoje não tão novas, mas mutantes) para fazer o mesmo que já faziam, sem perceber que a tecnologia da informação permite fazer muito, mas muito mais; inclusive continuar a fazer o que já se fazia, muito melhor e para o mundo inteiro. Baptista (2007) descreve a ruptura paradigmática que caracteriza a representação física e temática do documento eletrônico – e/ou dos objetos informacionais digitalizados – e insiste na necessidade de integrar as práticas bibliotecárias e as novas tecnologias, para otimizar os resultados da busca e recuperação e situar a representação no contexto da ciência da informação, identificando interfaces com o fenômeno da cognição.

Caberia perguntar: Será que por uma brincadeira do destino os até há pouco tempo tão celebrados OPACs (*Online Public Access Catalogs*), e hoje perdendo fôlego, foram um prenúncio das bibliotecas tornadas *opacas* para a *Web*?[51] O problema parece ser duplo: os catálogos, apesar do uso de palavras-chave e metadados, carecem de maior especificidade e detalhe, especialmente quando se trata de buscas especializadas (SRIDHAR, 2004), e a visibilidade na Internet é limitada pelo parco aproveitamento das possibilidades que já oferece a Web semântica para passar da gestão da informação à gestão do conhecimento. Em outras palavras, a solução para trazer à luz a imensidão de conhecimentos acumulados nas bibliotecas, passa, como frisado dois parágrafos acima (GRADMANN, 2005), pela junção de FRBRs e RDF, com o mútuo enriquecimento das partes envolvidas (ciência da informação e ciência da computação) e o maior benefício para a humanidade.

Esperemos que um anjo do bem ajude os profissionais da informação a aproximar os FRBRs e o RDF, para um encontro amigável na *Web* semântica.

5. REFLEXÕES CONCLUSIVAS

Os novos profissionais da informação responsáveis pelas tarefas de representação deveriam se concentrar mais na indexação da melhor qualidade e na elaboração de resumos ou sinopses que realmente representem os conteúdos reais dos documentos. Resumos esses a publicar, assim como as palavras-chave, também em inglês, que é, de fato, a *lingua franca* da Web. E isso, porque justamente, a estruturação dos dados e informações, em campos de metadados específicos, direciona a varredura pelos motores de busca para os campos que contêm a informação que se procura. A qualidade da indexação por palavras-chave ou descritores e a realização de sinopses de qualidade, exigem uma leitura crítica, a desconstrução dos textos para descobrir conteúdos talvez não evidentes a primeira vista. Parece, pois, indispensável abordar, nas matérias que se relacionam com as descrições, o ensino de novas técnicas de leitura, de raciocínio. Sem entrar em detalhes, parece conveniente reaprender a pensar, a ler, a entender. Uma pitada de filosofia, de lingüística, de semântica, de teoria das classificações (taxonomias, hierarquias, relações dos conceitos) e, naturalmente, lógica, merece ser considerada.

E, finalmente, um melhor conhecimento do básico das tecnologias da informação e da comunicação para poder trocar idéias e buscar caminhos para trabalhar e pesquisar em conjunto com outros profissionais, que também se interessam pela informação e seu processamento e uso.

Que tal promover alguns encontros entre profissionais que acreditam no futuro? Que tal repensar o conteúdo das atuais disciplinas denominadas 'fontes de informação', 'bibliografia especializadas' ou similares, por outra(s) que se poderia(m) chamar 'fontes e sistemas de informação especializados'

[51] Os dados das transações na base de dados Aleph OPAC entre julho de 2004 e junho de 2008, informados pela FCLA (*Florida Center for Library Automation*), apontam para uma diminuição do número de buscas: 1 jul 2004 – 30 jun 2005 -> 15.877.763 buscas;
1 jul 2005 – 30 jun 2006 -> 35.094.511 buscas;
1 jul 2006 – 30 jun 2007 -> 48.780.182 buscas,
1 jul 7007 – 30 jun 2008 -> 21.205.281 buscas

e que seriam ministradas fazendo uso de fontes e bases de dados reais (embora sendo virtual a maioria dentre elas),acessíveis em convênio com a Biblioteca' Central e/ou as bibliotecas setoriais das universidades federais, e cujas assinaturas seriam pagas per estas ou por órgãos de apoio adequados do Ministério da Educação ou, ainda, por um Ministério das Universidades, como já existe em alguns países europeus? Que tal tornar realidade a 'alfabetização' ou 'competência' informacional (*information literacy*), da qual tanto se fala, desde o começo dos cursos de graduação?

6. BIBLIOGRAFIA DOCAPÍTULO 1

A LIST A PART (Portal). A list a part for people who make websites. Disponível: http://www.alistapart.com/about Acesso: mai 2010.

AKSOY, Lerzan; BLOOM, Paul N.; LURIE, Nicholas H.; COOIL, Bruce**.** 'Should Recommendation Agents Think Like People?' *Journal of Service Research*, v.8, n.4, 2006. p.297-315. Disponível: http://jsr.sagepub.com/cgi/content/abstract/8/4/29. Acesso: mai 2009.

ALVARENGA, Lídia. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, n.15, 1º semestre 2003. Disponível: http://www.encontros-bibli.ufsc.br/Edicao_15/-alvarenga_representacao.pdf>. Acesso: abr 2008.

ANDERSON, Nate. *'Tim Berners-Lee on Web 2.0: "nobody even knows what it means'.* arstechnica.com, January 9, 2006. Disponível: http://arstechnica.com/news.ars/post/20060901-7650.html. Acesso: abr 2009.

ANDERSON, Paul. 'Semantic Web Tecnologies – has their time come in education? *JISC – The TechWatch Blog*, Notes from te Future. July 7, 2008. Disponível: http://notesfromthefuture.jiscinvolve.org/2008/07/17/semantic-web-technologies-%E2%80%93-has-their-time-come-in-education/. Acesso: maio 2009.

ANDERSON. Terry; Whitelock, Denise. 'The Educational Semantic Web: Visioning and Practicing the Future of Education'. *Journal of Interactive Media in Education.* n.1, 2004. Special Issue on the Educational Semantic Web. ISSN 1365-893X. Disponível: http://www-jime.open.ac.uk/2004/1/. Acesso: abr 2008.

BAPTISTA, Dulce M. O impacto dos metadados na representação descritiva. *Revista ACB: Biblioteconomia em Santa Catarina*, Florianópolis, v.12, n.2, p.177-190, jul./dez., 2007. Disponível: http://www.acbsc.org.br/revista/ojs/viewarticle.php?id=241&layout=abstract. Acesso: mai 2008.

BEACON, Matthew. Training Issues Managers Face. *In*: ALCTS/CETRC Preconference. (*What They Don't Teach in Library School: Competencies, Education and Employer Expectations for a Career in Cataloging*). *Papers*. Washington, DC, June 22, 2007. Disponível: http://www.loc.gov/catdir/cpso/-careercat.html. Acesso: mai 2008.

BERNERS-LEE, Tim. '*Information Management: A Proposal*'. Geneva, CERN, March 1989, May 1990. (W3 Archive). Disponível: http://www.nic.funet.fi/index/FUNET/history/internet/w3c/proposal.html. Acesso: abr 2009.

BERNERS-LEE, Tim; FISCHETTI, Mark. '*Weaving the Web : The Original Design and Ultimate Destiny of the World Wide Web by It's Inventor'*. San Francisco CA: Harper, 1999. ISBN 0062515861. Disponível: http://www.abebooks.com/servlet/SearchResults?an=Berners-_Lee&tn=Weaving+Web&paratrk=16303 (Ref. 1st Edition, 1999). Acesso: mai 2009.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities". *Scientific American*. May 17, 2001. Disponível: http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>. Acesso: abr 2009.

BERNERS-LEE, Tim. "*Views on Web 2.0".* Interview for Scott Laningham – "developerWorks" IBM. 28 July 2006, Disponível: https://www.ibm.com/developmerworks/mydevelopmentworks/blogs/scott/entry/coming_up_tim_berners_lee?lang=en. Também: "developerWorks Interviews" Disponível: http:www.ibm.com/developerworks/potcast/dwi/cm-int082206txt.html. Acesso: May 2010.

BLATTMAN, Úrsula; SILVA, Fabiano Souto Corrêa da. Colaboração e Interação na Web 2.0 e Biblioteca 2.0. *Revista ACB: Biblioteconomia em Santa Catarina*, v.12, n.2, jul/dez 2007, p. 191-215.

BRAZILIAN WORKSHOP on Semantic Web and Education, First. São Paulo SP, 29-30 nov 2007. Disponível: http://www.grow.ic.ufal.br/events/first-brazilian-workshop-on-semantic-web-and-education-1?searchterm=First>. Acesso: abr 2009.

BRAZILIAN WORKSHOP on Semantic Web and Education, Second. Fortaleza CE, 12-14 nov 2008. Disponível: http://www.grow.ic.ufal.br/events/WSWEd2008.en?set_language=en-us>. Acesso: abr 2009.

BRIAN, Martin. '*Applying Semantic Web Technology to Car Repairs'. In*: Garcia, R. (ed.). *Op. cit.*, 2009. Section II, Chapter II.

BUCKLAND, Michael K.; GEY, Fredric C.; LARSON, Ray R. Access to Heritage Resources Using What, Where, When, and Who. *In: Museums and the Web 2007*, 2007, San Francisco, California. Museums and the Web 2007: *Proceedings*. Toronto: Archives & Museum Informatics. Toronto, Canadá : Archives & Museum Informatics, 2007. Disponível: http://www.archimuse.com/mw2007/papers/buckland/buckland.html. Acesso: mai 2009..

BUSH, Vanevar.'*As we may think*". *Atlantic Monthly*, v.176, Jul 1945, p.101-108. Disponível: http://www.theatlantic.com/doc/194507/bush. Acesso: abr 2009.

CALHOUN, Karen.*The Changing Natura of the Catalog and its Integration with Other Discovery Tools*. Report to the Libray of Congress. Washington , DC: Library of Congress. March 17, 2006. Disponível: http://www.loc.gov/catdir/Calhoun-report-final.pd. Acesso: mai 2008.

CALHOUN, Karen. On Competition for Catalogers*. In*: ALCTS/CETRC Preconference. (What They Don't Teach in Library School: Competencies, Education and Employer Expectations for a Career in Cataloging). *Papers*. Washington, DC: June 22, 2007. Disponível: http://www.loc.gov/catdir/cpso/careercat.html. Acesso: mai 2008.

CHI, Ed. H. '*Yahoo! Answer vs. Google+Wikipedia vs. Powerset.'* Posted May 13, 2008. Disponível: <http://asc-parc.blogspot.-com/2008/05/yahoo-answer-vs-googlewikipedia-vs.html>. Acesso: mai 2008.

CONTROVERSIES in Cataloging: The debate over AACR2's successor. *American Libraries*, oct 2007, p.66-67 (Special Report-Technical Services).

COYLE, Karen; HILLMANN, Diane. Resource Description and Access (RDA): Cataloging Rules for the 20th Century. *D-Lib Magazine*, v.13, n.1/2, Jan/Feb 2007. 11 p. ISSN 1082-9873. Disponível: http://dlib.org/dlib/january07/coyle/01coyle.html. Acesso: fev 2007.

DAVIES, Kevin. "The NextBio Thing in Bioinformatics". Bio-IT World.com (White paper), 2009. Disponível: http://www.bio-itworld.com/issues/2006/dec-jan/nextbio/. Acesso: mai 2009.

De FURIO, Ivano; FRATTINI, Giovanni; MAGALDI, Massimo; ROMANO, Luigi. '*Experience in Building Mobile E-Business Services:Service Provisioning and Mobility*' . *In*: Garcia, R. (ed.). Op. cit., 2009. Section II, Chapter XIII.

DEVEDZIC, Vladan**.** Education and the Semantic Web'International. *International Journal of Artificial Intelligence in Education*. v.14. 2004, pp. 39-65. Disponível: http://iospress.metapress.com/content/hr4v08qm6vy8y3t7/. Acesso: abr 2009.

DIGNAN, Larry; Diaz, Sam; STEINER-Threlkeld, Tom. '*TechNet Summit: The new era of innovation*'. Panel. Technical Innovation Summit, Nov 15, 2006. Stanford University, Standford, California*. ZDNet blog***,** Nov 15, 2006 (Posted by Dan Farber). Disponível: http://blogs.zdnet.com/BTL/?p=3959. Acesso: mai 2009.

EVRI (portal). Disponível: http://www.evri.com/welcome. Acesso: mai 2010

FAWZI, Marc. '*Wikipedia 3.0: The End of Google?' Evolving Trends*. Publicado em 26 jun 2006. Disponível: <http://evolvingtrends/wordpress.com/2006/-06/26/wikipedia-30-the-end-of-google. Acesso: mai 2009.

GARCIA, Roberto (ed.) *Semantic Web for Business: Cases and Applications.* Oviedo, Spain: University of Lleida, 2009. 444 pages. ISBN: 978-1-60566-066-0. Excerpts: Disponivel: http://www.igi-global.com/downloads/excerpts/8258.pdf. Acesso: mai 2009.

GRAHAM, Paul. *Web 2.0*. 2005. Disponível: http://www.paulgraham.com/web20.html. Acesso: abr 2009.

GREENLIVINGPEDIA. *Web 3* , 26 fev 2009. Disponível: http://Greenlivingpedia.org/Web_3. Acesso: abr 2009.

GRIGOROVICI, Dan. *The Death of Web 3.0 (on Wikipedia) – death by Wikipedia censorship*. Feb 24, 2009. Disponível: http://www.web3beat.com/2009/02/the-death-of-web-30-on-wikiped.html. Acesso: abr 2009.

'Google Co-Op: The End of Wikipedia?' Evolving Trends, Posted September 24, 2008. Disponível: http://evolvingtrends/wordpress.com/2006/09/24/google-co-op. Acesso: set 2008.

'Google versus Wikipedia'. Publicado por '*aozuas',* January 8, em Motores de busca, Redes sociais e Web 2.0 Disponível: http://11pixels.ciberarte.com.br/motores-de-busca/-506/google-versus-wikipedia.html. Acesso: mai 2009.

'Google, Wikipedia & ProQuest'. ProQuest K-12. Comparison video. 2009. Disponível: http://www.proquestk12.com/demo/Google_ProQuest/Google_ProQuest.shtml. Acesso: mai 2009.

'Google Warming Up the Wikipedia 3.0 vision? ' Slashdot - News for nerds, stuff that matters. December 14, 2007. http://evolvingtrends/wordpress.com/2007/12/14/google-tries-again-to-co-opt-the-wikipedia-30-vision/ Acesso: mai 2009.

GRADMANN, Stefan. *"rdfs:frbr – Towards an Implementation Model for Library Catalogs Using Semantic Web Technology"*. *Cataloging & Classification Quarterly*, v.39, n.3/4, 2005, p. 63-75.

GUTIERREZ, Homer; HEIMBERG, Jessica. Dallas Museum of Art Presents The ARts Network**.** *In*: Museums and the Web 2007, 2007, San Francisco, California. Museums and the Web 2007: Proceedings. Toronto: Archives & Museum Informatics. Toronto, Canadá : Archives & Museum Informatics, 2007. Disponível: http://www.archimuse.com/mw2007/papers/gutierrez/gutierrez.html. Acesso: mai 2009.

HALL-ELLIS, Sylvia D. "*Cataloging Education: A New Emphasis for the LIS Curriculum*". *In*: ALCTS/CETRC Preconference. (What They Don't Teach in Library School: Competencies, Education and Employer Expectations for a Career in Cataloging). *Papers* Washington, DC: June 22, 2007. Disponível: http://www.loc.gov/catdir/cpso/-careercat.html. Acesso: mai 2008.

HERRERO CÁRCEL, Germán. *Towards the Use of Networked Ontologies for Dealing with Knowledge-Intensive Domains: A Pharmaceutical Case Study'. In*: Garcia, R. (ed.). *Op. cit.*, 2009. Section II, Chapter XIV.

HILL, Janet Swan. *The Brick Wall: Recruiting People to a Career in Cataloging. In*: ALCTS/CETRC Preconference. (What They Don't Teach in Library School: Competencies, Education and Employer Expectations for a Career in Cataloging). *Papers*. Washington, DC: June 22, 2007. Disponível: <http://www.loc.gov/catdir/cpso/-careercat.html. Acesso: mai 2008.

HUAJUN Chen; YIMIN WANG; HENG Wang; YUXIN Mao; JINMIN Tang, CUNYIN Zhou; AININ Yin; ZHAOHUI Wu. *Towards a Semantic Web of Relational Databases: a Practical Semantic Toolkit and an In-Use Case from Traditional Chinese Medicine. In:* 4th International Semantic Web Conference (ISWC'06)**.** Athens, GA, USA, November 5-9, 2006, *Proceedings*. Springer-Verlag, Athens, USA: Springer-Verlag,2006, pp. 750-763, ISBN 3-540-49029-9. Disponível: http://www.aifb.uni-karlsruhe.de/Publikationen/showPublikation?publ_id=1266. Acesso: mar 2009.

IFLA Study Group on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records: Final report.* UBCIM Publications – New Series, v.19. München: K.G. Saur, 1998.

IFLA Study Group on the Functional Requirements for Bibliographic Records. "*Functional Requirements for Bibliographic Records - Final Report".* Approved by the Standing Committee of the IFLA Section on Cataloguing, September 1997. As amended and corrected through February 2008. Disponível: http://www.ifla.org/VII/s13/frbr/index.htm. Acesso: 9 mar 2008.

IFLA Study Group on the Functional Requirements for Bibliographic Records. "*Functional Requirements for Bibliographic Records - Final Report".* Approved by the Standing Committee of the IFLA Section on Cataloguing, September 1997. As amended and corrected through February 2009. Disponível: http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf. Acesso: abr 2009.

KRÖTZSCH, Markus; VRANDECIC, Denny; VOLKEL, Max. *Semantic Media Wiki.* WIKIPEDIA, the Free Encyclopedia. 2005. Disponível: http://pt.wikipedia.org/wiki/Semantic_MediaWiki Last modif. April 21, br 2009. Acesso: mai 2009.

KRÖTZSCH, Markus; VRANDECIC, Denny. *Semantic Media Wiki* Prsentation, 2007. Disponível: http://korrekt.org/talks/Wikimania2007/SMW.html**.** Acesso: mai 2009.

LIBRARY of Congress. *Testing Resource Description and Access (RDA) – Proposed Methodoloy for U.S National Libraries RDA TEST*. March 13, 2009. Disponível: http://www.loc.gov/bibliographic-future/rda/testing.html. Acesso: abr 2009.

LEGER, Alain; CHARLET,Jean; HEINECKE, Johannes; HOBSON, Paola; NIXON, London J.B; GOASDOUÉ, François; SHVALKO, Pavel. '*Semantic Web Take-off in a European Industry Perspective*'. *In*: Garcia, R. (ed.). Op. cit., 2009.

LIBRARY of Congress**.** *Functional Analysis of the MARC21 Bibliographic and Holding Funds: Display for Multiple Versions from MARC21 and and FRBR.* Disponível em: http://www,loc.gov/marc/marc-functional-analysis/multiple-versions.html. Acesso em 10 mai 2008.

LIBRARY of Congress Working Group on the Future ob Bibliographic Control. *"On the Road".* Report. January 9, 2008, 44p.

LINKING OPEN DATA - W3C SWEO Community Project. *Project Description*. April 30, 2009. Disponível: http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData. Acesso: mai 2009.

MacMANUS, Richard. *Twine: The First Mainstream Semantic Web App?* (October 18, 2007). Disponível: http://www.readwriteweb.com/archives/twine_first_mainstream_semantic_web_app.php. Acesso: mai 2009.

MARKOFF, John. *Entrepreneurs See a Web Guided by Common Sense. New York Times* November 12, 2006. Disponível: http://www.nytimes.com/2006/11/12/business/12web.html?_r=1&pagewanted=1. Acesso: abr 2009

MENGIS, Raoul. '*Web 1.0 Web 2.0 Web 3.0'*. Disponível: http://knol.google.com/k/raoulmengis/web30/1gr1q7rtu4hhn/4#Web_1(2E)0_Web_2(2E)0_Web_3(2E)0. Acesso: abr 2009.

MEY, Eliane S.A. *Algumas questões sobre o ensino da representação descritiva, ou a catalogação na berlinda*. Disponível: www2.fgv.br/bibliodata/geral/docs/TextoElianeSerraoAlvesMey.pdf/ (2005). Acesso: mar 2008.

MORENO, Fernanda P. Requisitos Funcionais para Registros Bibliográficos – FRBR: um estudo no catálogo da Rede Bibliodata' *In*: *VIII ENANCIB – Encontro Nacional de Pesquisa em Ciência da Informaçã*o, Salvador BA: 28-31 out 2007 (*Trabalhos.* GT2 – Organização e Representação do Conhecimento). Disponível: http://www.enancib.ppgci.ufba.br/artigos/GT2-230.pdf. Acesso: fev 2008.

MOURA, José. *Jimmy Wales, pai da Wikipdia, diz que a Web 3.0 não tem futuro*. BM – Blog do Moura (Informática, Tecnologia, utilidade pública, pessoal). Publicado em 21 nov 2008. Disponível: http://www.blogomoura.com/2008/11/jimmy-wales-pai-da-wikipedia-diz-que-a-web-30-nao-tem-futuro/. Acesso: abr 2009.

NEPOMUK – The Social Semantic Desktop. '*NEPOMUK*, Main1 (Project objectives, Project summary, Partners…)**,** Version 1.22. Last modified on April 13, 2007. Disponível: http://nepomuk.semanticdesktop.org/xwiki/bin/view/Main1/. Acesso: mai 2009.

OLIVER, Chris. Changing to RDA. *Feliciter,* n.5, 2007, p. 250-253 (Uma publicação bimestral da Canadian Library Association www.cla.ca).

O'REILLY, Tim. *What is Web 2.0: Desihn Patterns and Busines Models for the Next Generation of Software*. September 30, 2005. Disponível: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/30/What-is-Web-20.html. Acesso: abr 2009.

O'REILLY, Tim. *"Web 2.0 Compact Definition: Trying Again'.* July 28, 2006. Disponível: http://radar.oreilly.com/archives/2006/12/web_20_compact.html. Acesso: abr 2009.

PERDRIX, Ferran; GARCIA Roberto; GIL, Rosa; OLIVA, Marta; MACIAS, José A.; Moya, Diego. *Interfaces basados en la Web Semântica para la Gestión de Contenidos Multimedia*. 2007. (Proyecto S5T (TIN2005-06885) subvencionado por el Ministerio de Educación y Ciencia en el Programa Nacional de Tecnologias Informáticas. Disponível: http://rhizomik.net/~roberto/papers/Interaccion2007.pdf. Acesso: mai 2009.

PERDRIX, Ferran; GIMENO, Juan Manuel; GIL, Rosa; OLIVA, Marta; GARCIA Roberto. *Semantic Web for Media Convergence: A Newspaper Case*' . *In*: Garcia, R. (ed.). *Op. cit.,* 2009. Section II, Chapter IX.

RAYWARD, W. Boyd. *Anticipating the Digital World: Paul Otlet and his Paper Internet*. Jean Bartel Memorial Lecture, Leeds Metropolitan University, May 31st, 2002. Disponível: http://leep.lis.uiuc.edu/publish/wrayward/Otlet/Bartelslecture1.pdf. Acesso: abr 2009.

ROBREDO, Jaime. *Documentação de Hoje e de Amanhã***,** 4ª ed. rev. ampl. Brasília DF: Edição de autor. 2005. 409p. ISBN 85-905920-1-4.

ROBREDO, Jaime. Sobre Arquitetura da Informação. *Revista Ibero-Americana de Ciência da Informação*, n.2, 2009

SHANNON, Victoria. *A 'more revolucionary' Web*. International Herald Tribune. May 25, 2006. Disponível: http://wwww.iht.com.bin/print-ipub.php?file=/2006/05/23/business/-web.php. Acesso: mai 2009.

SCHMIDT, Eric. *What is Web 1.0, Web 2.0 and Web 3.0.* Vídeo apresentando uma fala no Seul Digital Forum, May 21-31, 2008. O vídeo pode ser baixado em: Trak.In' India Business Buzz: Disponível: http://machiko4455.blogspot.com/2008/04/web10-web20-web30.html e em: http://trak.in/tags/business/2007/08/13/what-is-web10-web20-and-web-30-simple-explanation-for-non-techies/. Acesso: abr 2009.

SJC/RDA. *RDA – Resource Description and Access: Objectives and Pprinciples*. Draft, Rev. 2, October 28, 2008. Disponível: http://www.collectionscanada.gc.ca/5rda-objectivesrev2.pdt. Acesso: mar 2009.

SIOC Project. C*SIOC Core Ontology Specification*. SiocProject.org, 2009. Disponível: http://sioc-project.org/ontology. Acesso: mai 2009.

SPIVACK, Nova. 'The Third-Generation Web is Coming'. *KurzweilAI.net*. December 17, 2006. Disponível: http://www.kurzweilai.net/meme/frame.html?main=/articles/art0689.html?m%3D3. Acesso: mai 2009.

SRIDHAR, M.S. "*Subjet searching in OPAC of a special library: problems and issues*". 2002, 13 p. Disponível: http://eprints.rclis.org/archive/00008849/01/subsearching.pdf Acesso: out 2008.

TALAVERA IBARRA, Ana M. FRBR: Requerimientos Funcionales de los Registros Bibliográficos. *In*: Nuevas tendencias en la normalización y sistematización de la información. Lima: Biblioteca Nacional del Perú, 2006. p.157-186 (I Encuentro Internacional de Catalogadores. Lima, 24-26 oct 2005. *Ponencias y conclusiones*).

TILLET, Barbara. *What is FRBR? - A Conceptual Model for the Bibliographic Universe.* Washington DC: Library of Congress – Cataaloging Distribution Service, 2004. ("*O que é FRBR? Um Modelo Conceitual para o Universo Bibliográfico"* (Trad. de Lídia Alvarenga e Renato R. Souza.) Disponível: http://www.loc.gov/cds/FRBR.html. Acesso: mar 2009.

VAN OSSENBRUGGEN, Jacco; AMIN, Alia; HARDMAN, Lynda; HILDEBRAND, Michiel; Van ASSEM, Mark; OMELAYENKO, Borys; SCHREIBER, Guus; TORDAI, Anna; de BOER, Victor; WIELINGA, Bob; WIELEMAKER, Jan; NIET, Marco de; TAEKEMA, Jos; van ORSOUW, Marie-France; TEESING, Annemiek. Searching and Annotating Virtual Heritage Collections with Semantic-Web Techniques**.** *In*: *Museums and the Web 2007*, 2007, San Francisco, California. Museums and the Web 2007**:** *Proceedings*. Toronto: Archives & Museum Informatics. Toronto, Canada: Archives & Museum Informatics, 2007. Disponível: http://www.archimuse.com/mw2007/papers/ossenbruggen/ossenbruggen.html. Acesso: mai 2009.

VIAMONTE, Maria João; SILVA, Nuno. '*A Semantic Web-Based Information Integration Approach for na Agent-Based Electronic Market'* . *In*: Garcia, R. (ed.). *Op. cit.*, 2009. Section II, Chapter VIII.

YUNKER, John. *Wikipedia vs. Google*.. *Global by Design,* December 26, 2006. Disponível: www.globalbydesign.com.blog/2006/12/26/wikipedia-vs-google/. Acesso: mai 2009.

WALLENSTEIN, Andrew. *'Hollywood hot for Second Life'* . The Hollywood Reporter. February 13, 2007. Disponível::. <http://www.hollywoodreporter.com/hr/content_display/television/features/e3i49998ef2b580e2b5461e3dfb1faedb43?imw=Y>. Acesso: mai 2009.

'WEB 3.0'. Evolving Trends. June 26, 2006. Disponível: http://evolvingtrends.wordpress.com/web-30/. Acesso: abril 2009.

'WEB 3.0'. Information online – The first choice for business news and press releases, October 21, 2007. Disponível: http://.information-online.com/book/export/html/127 Acesso: mai 2009.

'*WHAT is Web 3.0?'* Design 3.0 – The Internet Changing, Portfolio (sem data) Disponível: http://www.design3point0.com/articles/web3.html. Acesso abr 2009.

WIKIPEDIA, the Free Encyclopedia. *FOAF_(software)*', 2009. Disponível: http://en.wikipedia.org/wiki/FOAF_(software). Acesso: mai 2009.

WIKIPEDIA, the Free Encyclopedia. *The Semantic Web Stack*, March 30, 2009. Disponível: http://en.wikipedia.org/wiki/Semantic_Web_Stack>. Acesso: mai 2009

WIKIPEDIA, the Free Encyclopedia. *User: Greyskinnedboy/Web 3.0*, March 19, 2009. Disponível: <http://en.wikipedia.org/wiki/User:Greyskinnedboy/Web_3.0> Acesso: mai 2009.

WIKIPEDIA, the Free Encyclopedia. *Web 2.0* . November 4, 2008. Disponível: http://en.wikipedia.org/wiki/Web_2.0. Acesso: abr 2009.

WIKIPEDIA, the Free Encyclopedia. *Web 3.0* . 2008 (Ultima atualização 20 mar 2009). Disponível: http://en.wikipedia.org/wiki/Web_3.0 Acesso: abr 2009.

WIKIPEDIA, the Free Encyclopedia. *World Wide Web Consortium*. (Last modified May 14 2009). <http://en.wikipedia.org/wiki/W3C>. Acesso: mai 2009.

WORLD Wide Web Consortium - W3C. *OWL Web Ontology Language Overview.* W3C Recommendation 10 February 2004. Disponível: <http://www.w3.org/TR/owl-features/>. Acesso: abr 2009.

WORLD Wide Web Consortium - W3C. *Design Issues: Architectural and philosophical points*'. 2007. Disponível: http://www.w3.org/DesignIssues/. Consultado: abr 2009.

WORLD Wide Web Consortium - W3C. *W3C Semantic Web Activity*. Last update: Apr, 10, 2009. Disponível: http://www.w3.org/2001/sw/. Acesso: mai 2009.

ZAINO, Jennifer. `*The Social Semantic Desktop Project Wraps Up*'. *Semantic Web* – The Voice of Semantic Web Tecnology. December 3, 2008. Disponível: http://www.semanticweb.com/article.php/3788691. Acesso: mar 2009.

ZAND, Tristan. *Web 3.0 back to the real world / back to our senses.* June 2006. Disponível: http://www.zzz.ch/bootymachine/web3.0. Acesso: mai 2009.

ZELDMAN, Jeffrey. *Web 3.0 A List Apart – for people who make websites*. (Business, Industry, Politics and Monay, Scripting, Usability). January 16, 2006. Disponível: http://www.alistapart.com/articles/web3point0/. Acesso: mai 2009.

Capítulo 2

A perfeição do nada: da videologia à desideologia da percepção

Juliano Serra Barreto[1]

SUMÁRIO DO CAPÍTULO 2

Resumo do Capítulo 2, *p. 44*
Como citar o Capítulo 2, *p. 44*

1. **INTRODUÇÃO DO CAPÍTULO 2**, *P. 45*
2. **ADMIRÁVEL MUNDO AUDIO VISUAL**, *p. 47*
3. **A BABELIZAÇÃO DO CONHECIMENTO**, *p. 49*
4. **NADA É PERFEITO**, *p. 50*
5. **CONCLUSÃO**, *p. 52*
6. **REFERÊNCIAS DO CAPÍTULO 2**, *P. 53*

[1] Doutor em Ciência da Informação. Professor Adjunto, Departamento de Desenho Industrial – Instituto de Artes (IdA/DDI) da Universidade de Brasília (UnB). CV Lattes: http://lattes.cnpq.br/9650639440622019. E-mail: serra@unb.br.

Resumo do Capítulo 2

Diante do boom informacional e dos consequentes efeitos sobre o aparato perceptivo humano, evidencia-se a necessidade urgente de uma relação mais vigorosa entre a ciência da informação e outras ciências que estudam a cognição humana. Tal intercâmbio pode cobrir lacunas nestas áreas e ampliar o escopo holístico da ciência da informação.
***Palavras-cha*ve**: Comunicação, Cognição, Material audiovisual.

Abstract

The perfection of nothing: from videology to the desideology of perception
Amidst the informational boom and looking for the deep effects on the human perceptive apparatus, we see the urgent necessity of a vigorous relationship between the Information Science and other sciences of the human cognition. This can cover gaps in these areas and extend the Information Science's multidisciplinary scope.
Keywords: Communication, Cognition, Audiovisual.

***Como citar o capítulo 2*:**

BARRETO, J. S. A perfeição do nada: da videologia à desideologia da percepção*. In* : Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios pelo bosque da informação*: *estudos sobre representação e organização da informação e do conhecimento*. Brasília DF: IBICT, 2010. 335 p. ISBN: 978-85-7013-072-3. Capítulo 2, p. 48-59. Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

1. INTRODUÇÃO DO CAPÍTULO 2

A consciência, que por si mesma nos possibilitou, na visão da aparência, a apreensão da Idéia que nela se manifesta, poderia finalmente ser obrigada a exclamar, como Fausto: "Que espetáculo! Mas – ai de mim – só um espetáculo! Por onde te agarrarei, natureza infinita?"
(Richard Wagner).

Vivemos agora os efeitos de uma explosão informacional, onde a quantidade de informação disponível multiplica-se em progressão geométrica pela intensificação do uso extensivo do computador, da televisão e da rede mundial. A sensação é a de imersão plena em um oceano de conhecimentos no qual, porém, o signo já não tem mais significado, pois o conteúdo total equivale ao de uma Babel de mensagens sem sentido nem destinatário. Um desafio está colocado sobre toda a humanidade, na escolha das rotas e caminhos a serem navegados neste oceano. Observamos atônitos a crescente multiplicação da violência e, ao mesmo tempo, a atrofia de nossa susceptibilidade ao sofrimento humano. E não somente isso, mas também a nossa sensibilidade à arte, à verdade, à ética está se transformando rapidamente.

Com isso, surge um novo tipo de *homo sapiens* que, mais do que elaborar e trocar concepções simbólicas (nossa diferença primordial com outras espécies) vai se especializando em consumir indiscriminadamente as representações audiovisuais impostas pela mídia. É o alvorecer do *homo videns*, o homem que vê, o homem do pós-pensamento, como sugere o filósofo e cientista político italiano Giovanni Sartori (2001). Nesta troca de papéis estaríamos, segundo esse autor, perdendo inteiramente nossa capacidade reflexiva e crítica. Não existem mais critérios, e não se sabe a utilidade ou validade do que está sendo avidamente consumido. Para Sartori, o fenômeno tem duas faces: a desinformação, ou a deformação mesma da informação, intencional ou não, e a subinformação, a informação insuficiente, reduzida, descontextualizada.

É o que, percebido também por Umberto Eco (2004), nos distancia de nós mesmos na troca de uma memória coletiva por impressões instantâneas e individualistas: a montagem de referências individuais via rede mundial de computadores ameaçaria o caráter coletivo da cultura. Elimina seus filtros naturais, construídos por acumulação durante séculos, separando o que é relevante do que não é. Ele dá um exemplo: "Saber quando Júlio César nasceu é importante, enquanto a data da morte de sua mulher, não. Com a Internet essa valoração se anula. Os critérios passam a ser individuais e não mais se subordinam a valores culturais definidos consensualmente" (ECO, 2004). A Internet acaba criando esse mundo instantâneo e individual, no qual se relativizam a memória da cultura e o caráter coletivo dos povos. Eco diz que se assusta com a imagem do internauta solitário, em seu quarto, sobrevoando o que julga ser a totalidade do mundo por meio de cliques do seu mouse.

Figura 1. – A Web na cabeça – a imagem à esquerda mostra, em verde, as áreas do cérebro ativadas durante a leitura de um livro. A imagem à direita registra a atividade cerebral durante a navegação na internet. As áreas do cérebro envolvidas são semelhantes às da leitura, com um acréscimo importante, destacado em vermelho. Trata-se do córtex pré-frontal, que permite às pessoas tomarem decisões rapidamente, enquanto avaliam informações complexas (Fonte: Revista Veja, edição nº 2125, de 12 de agosto de 2009).

Assim como os filtros naturais são alterados, são também modificadas as condições cognitivas gerais, e consequentemente, a capacidade de ver, assim como de entender o que se vê, são minimizadas. A percepção sensorial, sobretudo visual, tem evoluído de forma surpreendente, nos adaptando a novos mundos sensoriais insuspeitos. À medida que desenvolvíamos, no caso do mundo visível, o uso sofisticado da linguagem visual e dos instrumentos de captação, e portanto da habilidade de produzir imagens cada vez mais complexas, o sistema visual humano sofria sucessivos refinamentos. Os primeiros filmes exibidos em comunidades isoladas, no princípio do século XX precisavam ser comentados por um "explicador" para que fossem compreendidos. Nesse caso a impressão de realidade era quase nula, mas quando a simbologia tridimensional do cinema foi reconhecida como tal, o narrador deixou de ser necessário. O aprendizado da percepção tem sido parte de nossa evolução, mas diante da sociedade da informação é possível que esteja ocorrendo uma descontinuidade nesse sentido.

Experimentos na área de neurociências indicam que estão se processando modificações importantes na maneira como o cérebro funciona a partir do contato próximo com máquinas e computadores. Em entrevista à Revista *Veja*, Gary Small[2] diz que "... a exposição à rede fortalece alguns circuitos neuronais (ver *Figura 1*). Com isso, fazemos mais com o cérebro, gastando menos energia." E explica que:

> Tecnicamente, a superexposição a estímulos constantes na Internet afeta a maioria dos circuitos corticais e a camada externa da área cinzenta do cérebro, o que inclui os lobos frontal, parietal e temporal. O resultado disso é que ocorre um reforço nos circuitos cerebrais que controlam as habilidades tecnológicas. Mas os circuitos relacionados a habilidades sociais são negligenciados.... Existe uma frase, citada pelos céticos: "Minha avó dizia que a TV iria apodrecer meu cérebro – o que de fato aconteceu". A verdade é que não sabemos o que acontecerá, mas precisamos reconhecer que a revolução digital traz efeitos bons e ruins ao cérebro. O importante é que ainda temos controle sobre aquilo a que escolhemos expor nossa mente (SMALL, 2009).

2. ADMIRÁVEL MUNDO AUDIOVISUAL

O poder hegemônico vem ao longo da História construindo barreiras e armadilhas para impedir o livre pensar e a livre expressão, e nestes casos se procura obter o controle do pensamento principalmente pela obstrução do acesso ao conhecimento e às idéias, proibindo-se publicações, queimando-se originais, calando-se as bocas. Acreditavam os ditadores (e alguns políticos ainda acreditam!), que controlando o meio físico, o suporte, seja o papel ou as ondas hertzianas, estariam impedindo a proliferação das idéias e portanto de ações indesejadas. A base necessária para homogeneização/higienização do pensamento. Pois se tal ameaça não se deu dessa forma no passado, pode estar em vias de ocorrer. Sim, agora, e de modo paradoxal, rápido e inesperado.

Na sociedade da informação parece haver um outro componente capaz de conformar o pensamento, de limitar a capacidade crítica, e de subverter a sensibilidade que costumamos imputar à condição humana. E não opera pela limitação da circulação de idéias, muito pelo contrário, exacerba justamente a publicação, a constância da oferta, a visibilidade total. Pois, a medida em que se avoluma esta torrente de informações vamos passando de observadores críticos, capazes de avaliar o que vemos e o que sentimos em relação ao que vemos, para meros visualizadores, espectadores atônitos de um fluxo ininterrupto de imagens e sons, e que induzem continuamente a ações automáticas.

> Entendemos, pois, que, na superinformação, a repetição torna visível o invisível, mas oculta a realidade pela banalização entorpecedora dos sentidos. Perdemos, então, a possibilidade de distinguir a nuance, pela rapidez com que temos de processar a procissão de signos desfilando em velocidade crescente (SARTORI, 2001).

[2] Neurocientista, autor de: "*iBrain: Surviving the technological alteration of the modern mind*", em colaboração com Gigi Vorgan, publicado por Harper Collins, New York, 2008.

Como os meios de comunicação destroem nosso referenciais de espaço e tempo, constituintes da percepção, e se instituem a si mesmos como espaço e tempo o espaço é o "aqui" sem distâncias, sem horizontes e sem fronteiras; o tempo é o "agora" sem passado e sem futuro (KHEL; BUCCI, 2004, p. 34-35), perdermos a capacidade de abstração e também a capacidade de distinguir entre a verdade e a mentira (SARTORI, 2001). Dessa forma, a percepção passa a ser moldada pelos impulsos da tecnologia que constrói sonhos e motiva desejos. Como em *Metrópolis*, de Fritz Lang, ela nos transforma em autômatos (RIBEIRO, 2001). Autômatos que estarão sempre prontos para aplaudir o nada, imputando a perfeição ao vazio. E no exercício desse poder mais do que devastador, a mídia realiza o que Noam Chomski chamou de "fabricação do consenso", a simples imposição totalitária do que as pessoas devem pensar e fazer de si mesmas e das outras.

Assim a mídia acaba por produzir os indivíduos que o mercado necessita pois estrutura antecipadamente a percepção da realidade numa sociedade do espetáculo e do consumo, em que as pessoas são despojadas de sua subjetividade. A exaltação do indivíduo e o culto à imagem do outro cria abismos subjetivos em que a personalidade comum já não é capaz de se reconhecer (KHEL; BUCCI, 2004, p. 67). Em vista desta alteração perceptiva pode-se questionar, na sociedade da informação, os seus efeitos éticos sobre a humanidade e, sobretudo, esta capacidade de moldar nossos corpos e mentes. A possibilidade de metamorfosear o corpo humano, oferecendo um novo campo perceptivo ao qual se adaptar, pode estar nos levando a um novo estágio humano, o *homo videns*, entendendo aí a instauração de um novo aparato sensorial, não exclusivamente visual, já que o cérebro é capaz de "ver" de várias outras maneiras, além de interpretar as formas construídas opticamente. O mais preocupante é que este poder transformador de nossa percepção é um instrumento que se revela ideal para o controle das massas, pois se antes sabíamos que "o que os olhos não vêem, o coração não sente" agora, com tanto mais o que ver, podemos dizer que, mesmo que os olhos vejam e revejam, o coração já não sente mais.

> Por ora, lamentavelmente, o que temos no controle é uma mídia servil à banalização da violência, coluna fundamental de todo um sistema de exploração, injustiça e perversas disparidades. É preciso inocular-nos a doença psicológica da passividade, é preciso tornar cada vez mais comum o absurdo, para que não nos escandalizemos, para que não nos revoltemos, para que achemos que as coisas são "normais" do jeito que são. Por isso, praticamente tudo o que passa em rede aberta de televisão é lixo, vulgaridade, erotização chula, escândalo, tragédia, fofoca, terror, violência. Até mesmo os programas humorísticos, que deveriam, por definição, trazer alegria e leveza às nossas vidas, estão estribados nos mais grosseiros preconceitos, no mais despudorado cinismo e, às vezes, na mais cruel selvageria ideológica voltada a escrachar os pobres e trabalhadores (vide o caso do global "Sai de Baixo" [3]). Os canais por assinatura são um pouco melhores mas, de maneira geral, não fogem à regra. A maioria dos videogames deveria fazer chorar qualquer adulto que tenha um mínimo de discernimento ético. Em termos de contexto familiar, nossas crianças e jovens convivem mais com eles do que conosco, é forçoso reconhecer (RIBEIRO, 2001).

[3] Comédia vulgar dominical da Rede Globo em 2004

A propósito, a psicanalista Tania Rivera lembra a previsão realizada em 1913 por um discípulo de Freud, Lou Andreas-Salome: "o futuro do filme poderá contribuir muito para a nossa constituição psíquica' (RIVERA, 2009, p. 63).

3. A BABELIZAÇÃO DO CONHECIMENTO

Uma aprendizagem informacional, ou educação para a informação, como tem sido também referida, e que suporte o tsunami da sociedade da informação, deve, antes de tudo, buscar a redescoberta da reflexão, da consciência da percepção e da sensibilidade com relação aos outros e a si mesma. Como sugere Castilho:

> A inclusão digital é a discussão do momento. Fala-se todo o tempo em integrar as pessoas à esse admirável novo mundo online. Entretanto, sem fornecer uma base educacional e cultural sólida às pessoas, não se obterá inclusão nenhuma. Uma pessoa que apenas saiba operar um mouse de computador continuará tão ignorante quanto a que não sabe. Não é a tecnologia que irá salvar as pessoas, assim como não é o oceano de informações online que irá deixar ninguém mais sábio. É preciso a base e essa base continuará sendo a educação, que se inicia com a alfabetização completa e correta, com os conceitos culturais de comunidade e identidade transmitido às pessoas (CASTILHO, 2005).

Em 2007 contou-se 281.000.000.000 Gigabites ativos no universo digital, o que significa cerca de 45GB por pessoa no planeta. Saber como encontrar um determinado conhecimento distante, se relacionar com máquinas e estar em rede é importante, mas não deveria comprometer a capacidade de ver apropriadamente, e de entender cada vez melhor o que está ao alcance de nossos olhos e mentes.

> A informação é em si ambivalente, tanto em quem a pronuncia, quanto em quem a recebe. Em todos os momentos passa pelo filtro da subjetividade, além de sua dimensão estar limitada pelo aparato perceptor e conceitualizador. Mas é esta ambivalência que resgata sempre a possibilidade de criar, inventar. Se tudo fosse apenas lógico, seria apenas repetitivo. O mundo da informação é agitado, conturbado, porque é, ao mesmo tempo, intrinsecamente manipulado e impossível de ser totalmente manipulado (DEMO, 2000). O problema portanto não é mais a ocultação e a manipulação da informação, e sim o "achatamento" da percepção, a evolução para um estado de desconhecimento da dimensão subjetiva, ou seja, a impossibilidade de perceber a ambivalência da informação, o que soterra por fim a criatividade e a inventividade, pela limitação paulatina do aparato perceptor e conceitualizador. O fato de existirem infinitas informações facilmente acessíveis não significa que sabemos para que servem.

O que está sendo ameaçado é a própria constituição da personalidade, em um processo franco de des-ideologização da existência individual, operando no aparato perceptivo e destruindo a possibilidade do auto-conhecimento. Entendemos nesse caso, as implicações do conceito de Ideologia indicado por Althusser, que diz, "*A ideologia é uma 'representação' da relação imaginária dos indivíduos com suas condições reais de existência*" (*in* "Aparelhos ideológicos de Estado"citado por Eugênio Bucci em (KHEL; BUCCI, 2004, p. 36)), e ainda que é isso o "[...] que garante ao mesmo tempo o reconhecimento mútuo entre sujeitos e o Sujeito, e entre os próprios sujeitos, e finalmente o reconhecimento do sujeito por si mesmo."

> As novas tecnologias eliminaram o tempo e o espaço real, trocando-o pelo ficcional. Permitiram a união de imagens, vozes e figurações de vida num tempo de simulação e invadiram o imaginário formatando-o. Se, anteriormente, a oralidade fornecia pistas e inspiração aos artistas, conferindo-lhes a função de revelar, com ilustrações, as histórias de uma cultura popular, a sociedade virtual, em suas redes, emaranha a imaginação através da formatação automatizada. A repetição acelerada de imagens estereotipadas gera o tautismo, cunhado por Lucien Sfez juntando a tautologia ao autismo, que, ao dar visibilidade, termina por tornar opaca a transparência da imaginação no homem globalizado (BALESTRERI, 2006).

Também Jaime Robredo, em sua discussão sobre as novas abordagens dos sistemas de informação, já alertava para uma tendência à superficialidade ocasionada pelo descompasso entre o tempo de recepção cultural e a velocidade crescente da transferência de informação, evidenciado na citação a Kornwachs:

> [...] assim podemos perguntar se é possível manter uma relação razoável entre a rápida disponibilidade da informação e a forma relativamente lenta do processo cognitivo humano [...] O maior problema com essa diferença entre as constantes de tempo técnica e humana, resulta do fato que os estados do processo cognitivo e o impulso de buscar novas informações importantes ou decisivas, apresenta-se de forma assincrônica (ROBREDO, 2003, p. 164).

E o mesmo fenômeno pode interferir inclusive na atividade científica, como observa Wilson Luiz Sanvito, Professor de Neurologia da Faculdade de Ciências Médicas da Santa Casa de São Paulo:

> O lema da sociedade e, principalmente, da comunidade científica é a obtenção do máximo de informação para se manter "*up to date*". Isso é um equívoco. Esse estado permanente de superinformação provoca ruído no sistema e não há tempo suficiente para uma reflexão crítica das mensagens recebidas para classificá-las sob a forma de conhecimento articulado. Estabelece-se uma cacofonia na cabeça do receptor das informações. Por outro lado, a pulverização das ciências contribui para a falta de integração de seus diversos ramos. Cada homem de ciência atua no seu próprio nicho e acaba perdendo a perspectiva de um saber integrado, o que determina uma perda do quadro de referências. A fragmentação do saber teve como conseqüência uma babelização do conhecimento. Nós vivemos uma época de especialização desabusada, em que o especialista é micro-cultíssimo e macro ignorante (SANVITO, 2000).

4. NADA É PERFEITO

Adoramos a perfeição, porque não a podemos ter; repugná-la-íamos,se a tivéssemos. O perfeito é o desumano, porque o humano é imperfeito"
(Fernando Pessoa)

Como sabemos... a perfeição, alijada do destino humano, não é mesmo necessária e, uma vez que nada é perfeito, podemos encontrar perfeição ainda, e apenas, na ausência de ação ou coisa, no silêncio. A rotina seria perfeita se não houvesse novidade. Assim a falta de mensagem pode viabilizar

a conexão ideal, sem estática ou modulações, contínua e plena. Aí a experiência real se encontra planificada e a inspiração não encontra ápices de onde possa alçar vôos criativos. Também não existe o silêncio, e então não há comunicação. A interação exige o silêncio, e a comunicação é um processo interativo.

Ciro Marcondes Filho (2007), sociólogo da USP, ao investigar a comunicação humana, percorre os diversos conceitos afeitos à comunicação na produção dos filósofos ocidentais, para concluir que: "As comunicações são antes extralingüísticas e promovidas pela interação humana" (p. 88); "Não nos comunicamos pela língua estruturada, porque ela mascara a comunicação" (p. 83).

Diz, citando Bergson: "as ideias que nos pertencem são as que menos podem ser expressas em palavras" (p. 85). "Não existe comunicação porque somos sistemas fechados" (p. 85). Pessoas e sistemas se voltam para si mesmos, segundo Luhman, e nisso se autoregulam, são autopoiéticos, sem trocas com o mundo exterior. "Na linguagem estruturada a comunicação torna-se ritualizada, não diz nada, por isso buscamos outras formas, menos ineficazes. Por exemplo, no silêncio, no toque físico, nos ambientes" (p. 93). A linguagem é como uma roupa (social, pretensamente comunicativa) com que cobrimos a transparência do nosso corpo nu. Segundo Nietzsche, pode existir pensamento mesmo onde não há palavras, pois o corpo pensa, mas não fala (p. 97). "Há labirintos na comunicação, pelos quais a realização da comunicação é o poder de driblar a proibição de se comunicar imposta pela 'sociedade da comunicação' " (p. 98). No comunicar, a percepção não se traduz em linguagem objetiva, esta é pobre, estéril, mero formalismo. Ela é antes sentida, vai além dos signos, como diz Merleau-Ponty, 'rumo ao silêncio deles' " (p. 101).

Como na vida social, que, para Henri Bergson, tem mais importância prática do que nossa existência interior e, baseada na linguagem estruturada, não permite mais o movimento, solidifica o entendimento fixando as sensações, e ocultando, ou mesmo destruindo, as impressões delicadas e fugidias da consciência individual (MARCONDES Filho, 2007, p. 84).

> Em suma, para um espírito que siga pura e simplesmente o fio da experiência não haveria vazio, não haveria o nada, mesmo que relativo ou parcial, nem negação possível. Um espírito como esse veria fatos se sucederem a fatos, estados a estados, coisas a coisas. O que observaria a todo momento são coisas que existem, estados que aparecem, fatos que se produzem. Ele viveria no atual e se fosse capaz de julgar, só afirmaria sempre a existência do presente (BERGSON, 1979, p. 256).

Como no cinema, ao projetar a sucessão de uma série de momentos fixos (fotografias) que não têm movimento, e provocar a ilusão daquele movimento que foi subtraído ao evento registrado, surge na velocidade midiática a mesma ilusão de completude onde na verdade se instaura o vazio.

Mas um vazio diferente, indolor, incapaz de sensibilizar ou ativar mudanças pois não implica em falta, mas em abundância. Teríamos perdido o medo, o horror ao vazio, e em consequência toda possibilidade de ação? Sim, se acreditarmos que toda ação procede de uma ausência:

> [...] toda ação humana tem seu ponto de partida em uma insatisfação e, por isso mesmo, num sentimento de carência. Não agiríamos se não nos propuséssemos um objetivo, e só procuramos uma coisa porque sentimos falta dela. Nossa ação procede pois do "nada" a" alguma coisa", e tem por própria essência bordar "alguma coisa" nos bastidores do "nada" (BERGSON, 1979, p. 258).

Para a professora Laura Coutinho, preocupada com o alcance educacional do cinema, "vivemos em um mundo em que a visibilidade chegou à exaustão. Não há mais o que não se veja" (COUTINHO, 2009, p.84). E além disso "Não sabemos mais ver sem as câmeras, e esse olhar é sempre ficção e realidade juntas" (p.86).

Ao desvelar a íntima relação entre o cinema de Antonioni e a filosofia de Heidegger, evidenciada nos filmes pela presença de temas como a "percepção do vazio, do nada (ou do ser) e a existência sufocada pela tecnologia, o ruído, o mecanicismo, a diluição e a publicidade" (CABRERA, 2006, p. 316), o filósofo Julio Cabrera reforça a idéia heideggeriana de que o tédio é um sentimento que permite o "acesso a um mundo mais fundamental do que o habitual, mais profundo e mais próprio" (p. 321). Mas o tédio é justamente o que não se espera mais encontrar no cinema. No fluxo audiovisual contínuo e pleno não "existe" o Nada.

Então, se o vácuo existencial foi dominado pela cornucópia midiática, e a escuridão da dúvida iluminada pelo hipnótico écran da informação instantânea... o que nos resta? "Somente o coração silente da clareira é o lugar do silêncio do qual pode irromper algo assim como a possibilidade do comum pertencer de ser e pensar, isto é, a possibilidade de acordo entre presença e apreensão" (HEIDEGGER, 1984, p. 79).

5. CONCLUSÃO DO CAPÍTULO 2

A Ciência da Informação, entre seus objetos de pesquisa, que se entendem como componentes de um sistema de conhecimento holístico, poderia e deveria abordar mais a fundo os aspectos acima discutidos, e incluir o estudo da cognição e da percepção como um de seus focos prioritários, diante da possibilidade que se apresenta, que é a de atingirmos uma estágio da sociedade em que não teremos mais leitores críticos, mas apenas processadores de informação. A ênfase dada na Ciência da Informação à tecnologia da informação e em suas conseqüências macro-sociais deixa escapar exatamente o momento da compreensão e da interiorização subjetiva da informação, e que tem importância fundamental na chamada Sociedade da Informação.

Em outras ciências estes aspectos são inevitavelmente tratados de forma parcial, pois não se considera a informação como um elemento principal, em geral estão mais preocupadas com aspectos fisiológicos e neurais, na visão da Ciência da Cognição, essencialmente biomédica, e com muitas aplicações nas Engenharias, ou com processos mentais abstratos, como no caso da Psicologia ou da Comunicação. Em ambos os campos não se concebe a informação em todas as suas dimensões e implicações, pois não é considerada em si mesma e, embora um vetor ativo em qualquer teoria, apresenta-se sem estatuto ou história.

Cabe à Ciência da Informação preencher tal lacuna, o que não tem acontecido, pois a vertente acadêmica preponderante é a que se dedica aos sistemas concebidos pelo homem, e poucos cientistas estudam os sistemas ditos naturais, conforme a delimitação didática apresentada por Robredo (2003), e que ainda no mesmo texto, sugere a identificação necessária da Ciência da Informação com as neurociências e as ciências cognitivas e psicossociais. Nosso singelo objetivo com essa pequena argumentação é de justamente reforçar estes vínculos interdisciplinares e de alertar para a urgência de tal investida, diante da velocidade inerente aos processos informacionais da atualidade.

7. REFERÊNCIAS DO CAPÍTULO 2

BALESTRERI, Ricardo. *Cidadania e direitos humanos: um sentido para a educação*. 2006. Internet. Disponível: http://www.dhnet.org.br/educar/balestreri/sentido/inabidicaveis.html. Acesso: 12 jun 2005,

BERGSON, Henri. *A evolução criadora*. Rio de Janeiro, RJ: Zahar Editores, 1979.

CABRERA, Julio. *O Cinema pensa*. São Paulo, SP: Rocco, 2006.

CASTILHO, Daniela. *Superinformação sem filtro*. 2005. Internet. Disponível: http://www.havesometea.net/MadTeaParty/archives/001618.html. Acesso: 12 jun 2005.

COUTINHO, Laura Maria. O olhar cinematográfico reflexões sobre uma educação da sensibilidade. *In*: *O Cinema e seus Outros.* Brasília: LGE Editora, 2009.

DEMO, Pedro. Ambivalências da sociedade da informação. *Ciência da Informação* Brasília DF, v. 29, 2000.

ECO, Umberto. O dia-a-dia da internet. *Revista Bovespa*. [S.l.: s.n.], 2004.

HEIDEGGER, Martin. *Conferências e escritos filosóficos*. São Paulo, SP: Abril Cultural, 1984. (Os pensadores).

KHEL, Maria Rita; BUCCI, Eugênio. *Videologias*. São Paulo, SP: Boitempo, 2004.

MARCONDES Filho, Ciro. *Até que ponto, de fato, nos comunicamos?* São Paulo, SP: Paulus, 2007.

PESSOA, Fernando. *Lvro do desassossego*. São Paulo SP: Companha das Letras, 2006.

RIBEIRO, José Luiz. Midiocracia: Torpor e torpezas. *Lumina* UFJF, v. 5, 2001.

RIVERA, Tania. A imagem e o escuro. *In*: CUNHA, Renato (Ed.). *O Cinema e seus outros*. Brasília: LGE Editora, 2009.

ROBREDO, Jaime. *Da Ciência da Informação Revisitada aos Sistemas Humanos de Informação*. Brasília, DF: Thesaurus; SSRR Informações, 2003.

SANVITO, Wilson Luiz. O fascínio das descobertas médicas. *Jornal da Tarde,* 2/12/2000, 2000.

SARTORI, Giovanni. *Homo Videns; Televisão e pós-pensamento*. Florianópolis, SC: EDUSC, 2001.

SMALL, Gary. A internet transforma o seu cérebro. *Veja*. Ed. Abril, 2009. Disponível em: http://veja.abril.com.br/120809/internet-transforma-cerebro-p-96.shtml. Acesso: 13 de mar 2010.

WAGNER, Richard. *Beethoven* (Orginal de 1870. Publicação no Brasil: São Paulo, SP: L&PM, 1987).

II. REPRESENTAÇÃO E ORGANIZAÇÃO DA INFORMAÇÃO

Capítulo 3. – O Escopo da Análise da Informação – Dulce Maria BAPTISTA, Rogerio Henrique de ARAÚJO JR. e Eliana CARLAN, *p. 61-80*

Capítulo 4. – $FRBR_{oo}$ – Coceitos básicos – João Alberto de Oliveira LIMA, *p. 81-92*

Capítulo 5. – Atributos dos Requisitos Funcionais para Registros Bibliográficos – Fernanda Passini MORENO, *p. 93-114*

Capítulo 6. – Organização da Informação: Conceitos básicos e breve fundamentação teórica – Lígia Maria Arruda CAFÉ e Rodrigo SALES, *p. 115-129*

Capítulo 7. – Organização da Informação: Proposta de elementos de arquitetura da informação, para repositórios digitais institucionais, baseados na descrição física e temática – Fernanda de Souza MONTEIRO, *p. 130-145*

Capítulo 3

O escopo da análise da informação

Dulce Maria Baptista[1], **Rogério Henrique de Araújo Júnior**[2] **e Eliana Carlan**[3]

SUMÁRIO DO CAPÍTULO 3

Resumo do Capítulo 3, *p.62*
Como citar o Capítulo 3, *p.62*

1. INTRODUÇÃO DO CAPÍTULO 3, *p. 63*
2. CONTEXTUALIZAÇÃO DA ANÁLISE DA INFORMAÇÃO NA CIÊNCIA DA INFORMAÇÃO, *p.64*
2.1 Dimensão conceitual, *p.65*
2.2 Dimensão estratégica, *p.68*
2.3 Dimensão operacional, *p.69*
3. O CONTEXTO DA ANÁLISE DA INFORMAÇÃO, *p.71*
4. CONSIDERAÇÕES FINAIS, *p.77*
5. REFERÊNCIAS DO CAPÍTULO 3, *p.79*

[1] Doutora em Ciência da Informação. Professor Adjunto. Programa de Pós-Graduação em Ciência da Informação (PPGInf). Faculdade de Ciência da Informação da Universidade de Brasília (UnB/FCI). CV Lattes: http://lattes.cnpq.br/5063841212997160 . E-mail: dmbp@unb.br.

[2] Doutor em Ciência da Informação. Professor Adjunto. Programa de Pós-Graduação em Ciência da Informação (PPGInf). Faculdade de Ciência da Informação da Universidade de Brasília (UnB/FCI). CV Lattes: http://lattes.cnpq.br/4752236489901895. E-mail: araujojr@unb.br .

[3] Mestre em Ciência da Informação. Programa de Pós-Graduação em Ciência da Informação (PPGInf). Consultora OPAS/OMS – área de disseminação da informação. E-mail: eliana.carlan@gmail.com.

Resumo do Capítulo 3

Neste capítulo busca-se contextualizar a análise da informação no âmbito da ciência da informação. A partir da identificação das dimensões conceitual, estratégica e operacional em que ocorre o processo de análise da informação, procede-se à descrição de cada uma dessas dimensões, tendo em vista a ênfase no significado da informação, e a conseqüente necessidade de representação condensada dos conteúdos temáticos. Nessa perspectiva, são considerados diferentes tipos de informação especializada, tal como representados em tabelas de classificação, como exemplos de representação condensada. A atribuição de palavras chave e a elaboração de resumos reflete, por sua vez, a necessidade de se compatibilizar a linguagem natural em que uma pergunta é formulada e a resposta que o sistema de informação tem a oferecer. Para ser representada, a informação precisa ser organizada e descrita. Considerando os diferentes suportes ou documentos em que é registrada, a descrição se vale de linguagens bibliográficas ou documentárias, e de interfaces com a terminologia e a lingüística. O contexto, em si, da análise da informação, engloba ações preliminares, ações de descrição, de representação, e outras ações complementares, que ampliam de forma significativa o escopo da análise da informação.
Palavras-chave: Significado da informação; representação do conhecimento; representação condensada; linguagem natural; representação da informação; organização da informação; suporte da informação; registro da informação.

Abstract

The scope of information analysis

In this chapter it is intended to focus information analysis within the context of information science. Starting with identification of conceptual, strategic and operational dimensions in which IA takes place, each one of those aspects is described, with emphasis on the meaning of information, and on the resulting need of summarized representation of contents. Having this in view, different types of specialized information are taken into account, such as shown in classification tables, as examples of condensed representation. Assignment of key words and abstracting reflect, on the other hand, the need of matching the natural language used in formulating a question and the response that is offered by an information system. Information needs to be organized and described in order to be represented. Considering different supports in which it may be recorded, description makes use of bibliographic or documentary languages, and of interfaces with terminology and linguistics. The context, itself, of information analysis, comprises a number of preliminary procedures, as well as description, representation and other complementary activities, which significantly enlarge the scope of infomation analysis.
Keywords: Information meaning; content representation; condensed representation; natural language; information representation; information organization; information support; information record.

Como citar o Capítulo 3

BAPTISTA, D. M.; ARAÚJO JR., Rogerio Henrique de; CARLAN, Eliana. Atributos dos Requisitos Funcionais para Registros Bibliográficos (FRBR). *In:* Jaime Robredo e Marisa Bräscher (Orgs.). *Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento.* Brasília DF: IBICT, 2010. 335 p. ISBN: 978-85-7013-072-3. Capítulo 3, p. 61-80. Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

1. INTRODUÇÃO DO CAPÍTULO 3

Considerando-se a amplitude semântica da expressão "Análise da informação" (AI), cujos termos já admitem por si sós uma ampla gama de significados, que tal expressão é aplicável a diferentes realidades, e que definições e conceitos costumam funcionar como ponto de partida para a compreensão de fenômenos e processos, procura-se nesse capítulo explorar inicialmente a questão da conceituação – já que não parece haver uma definição pronta e acabada para análise da informação, pelo menos no contexto da ciência da informação. A escassez da literatura no âmbito dessa ciência, e ao mesmo tempo a evolução nas práticas ligadas à análise documentária, representação, organização e recuperação da informação registrada, com ênfase, inclusive, nos conteúdos temáticos, indicam que todas essas práticas se referem a um processo mais amplo que se inclui, certamente, no escopo da análise da informação como um todo.

Numa abordagem preliminar, e entendendo a informação como elemento básico constitutivo de todo conhecimento, seja este simples ou complexo, e a utilização da própria informação para fins variados (pontuais, cotidianos, utilitários, científicos, tomada de decisão, etc.), pode-se observar que as pessoas se convertem, de um modo geral, e segundo seus interesses, em analistas de informação. E isso ocorre quando buscam a informação a partir de determinados critérios para processá-la cognitivamente e utilizá-la, também de acordo com critérios específicos e individuais. Nessa perspectiva abrangente, parece possível a identificação de três níveis em que se realiza a análise da informação: *1)* Nível intuitivo; *2)* Nível racional; *3)* Nível profissional. No nível intuitivo, o indivíduo entende, por exemplo, que o clima chuvoso recomenda que use um guarda-chuva para não se molhar. Ou que precisa caminhar com cuidado em determinadas calçadas para não cair no buraco. Essas e outras situações, típicas da vida cotidiana das pessoas, apelam à intuição e ao senso comum, não exigindo, em princípio, um maior nível de racionalização ao lidar com elas.

Já o nível racional exige, por exemplo, que alguém procure adequar as prestações do aluguel do imóvel onde reside à sua renda, caso deseje escapar da inadimplência. Pelo mesmo motivo, a utilização do cartão de crédito demanda um certo nível de racionalidade. Essas e inúmeras outras situações demandam níveis variáveis e crescentes de análise racional da informação, esteja esta contida nas circunstâncias da vida prática, na vida profissional, seja obtida por meio de comunicação informal, registrada em documentos e/ou publicações, disponibilizadas em meio eletrônico ou no espaço virtual, sempre em função de um uso previsto e estabelecido pelo próprio usuário, e de resultados e repercussões potenciais ou imediatos em sua vida.

Com relação ao nível profissional de análise da informação, parece pertinente considerá-lo como característico de algum tipo de mediação especializada que se torna necessária entre produtores e usuários da informação, sempre para fins de esclarecimento e de facilitação do acesso a fontes e conteúdos. Embora a distinção entre produtores e usuários se justifique, no caso, apenas do ponto de vista metodológico, na medida em que os indivíduos são

simultaneamente produtores e usuários de informação, e que, inclusive, os níveis racional e profissional de análise da informação não são mutuamente excludentes, mas ao contrário, bastante inter-relacionados.

A mediação se aplica a diferentes áreas e interesses, sendo que de acordo com tais áreas e interesses, assume características próprias, tornando-se portadora de uma singularidade que a distingue dos demais tipos de mediação. Dessa forma, e tendo em conta que na maioria das áreas a informação constitui atividade meio, o que as difere basicamente da ciência da informação – que tem por objeto a informação registrada, em si –, torna-se em princípio mais clara a visão do que consiste a análise da informação no contexto do jornalismo, por exemplo, do direito e da literatura de ficção, entre outras áreas. Torna-se claro que no âmbito da ciência da informação, análise da informação não é sinônima ou correlata do comentário jornalístico, seja este político, econômico, conjuntural, etc. Por exclusão, e por analogia, não se trata também da exegese do texto religioso ou filosófico, do diagnóstico médico, da hermenêutica jurídica ou da crítica literária. O que se pode inferir, de fato, é que existem diferentes tipos de análise da informação, o que talvez comprometa uma definição ou conceituação abrangente.

Por outro lado, e considerando-se o nível profissional dessa análise, observe-se o foco nos conteúdos temáticos da informação, o que servirá, por sua vez, a objetivos diferenciados. Numa visão aproximada, e portanto genérica, talvez seja possível entender análise da informação como sendo um conjunto de práticas que, viabilizadas pela articulação entre modelos conceituais e técnicas originárias dos campos de aplicação da ciência da informação, serve à compreensão, organização e recuperação dos conteúdos presentes em diferentes registros e suportes.

2. CONTEXTUALIZAÇÃO DA ANÁLISE DA INFORMAÇÃO NA CIÊNCIA DA INFORMAÇÃO

Se, de acordo com Le Coadic (2004, p. 4), "a informação comporta um elemento de sentido", e análise da informação, conforme dito acima, se preocupa com o conteúdo temático da informação, e, ainda, sendo a informação registrada o objeto da ciência da informação, torna-se pertinente estabelecer os elementos significativos que distinguem a natureza e os objetivos da análise da informação no contexto dessa ciência, em oposição à sua natureza e objetivos no âmbito das demais áreas do conhecimento. Se naquelas áreas, a análise da informação, em seu nível mais profissional, busca compreender e interpretar sentidos, sentidos esses presentes em algum tipo de registro, na ciência da informação a análise da informação serve à descrição, representação e organização de conteúdos – não necessária-mente à sua interpretação – independentemente do suporte em que se encontrem, para que se tornem disponíveis ao usuário (para que este, sim, os compreenda e interprete a partir do que Dervin (1998) entende como construção de sentido, na qual os condicionamentos individuais adquirem grande importância) a partir de uma linguagem documentária previamente construída, e que

funciona como mediação entre a linguagem natural daquele que busca informação e as fontes ou registros em que a informação buscada está efetivamente contida.

Não se quer sugerir aqui que esse tipo de mediação seja mecânico ou destituído do componente intelectual, mas que esse componente está fortemente presente na construção, desenvolvimento e uso das linguagens destinadas a representar conteúdos temáticos da forma mais isenta e objetiva possível. A esse respeito, inclusive, é bastante conhecida a situação enfrentada por indexadores no sentido de reduzir ao mínimo o coeficiente de subjetividade inerente às políticas e práticas de indexação. Conexões corretas entre conceitos e termos a partir de critérios eminentemente intelectuais resultarão na qualidade da representação de conteúdos e na recuperação eficaz da informação.

Em função dos diferentes níveis de complexidade, do conhecimento especializado, do aparato tecnológico envolvido, e da diversidade de aplicações e utilidades, parece possível identificar, no âmbito da ciência da informação, três dimensões caracterizadoras da análise da informação:

- Dimensão conceitual,
- Dimensão estratégica,
- Dimensão operacional.

2.1 Dimensão conceitual.

A dimensão conceitual da análise da informação é aquela que busca associá-la aos conteúdos temáticos da informação, identificando seu alcance e limites, estabelecendo interfaces disciplinares e objetivos que a caracterizam como uma das atribuições da ciência da informação. Nessa dimensão estariam incluídos conceitos, definições, hierarquia e tipologia da informação. No que se refere a conceitos e definições, esses parecem estar intimamente vinculados às múltiplas e variadas definições de informação presentes na literatura, na medida em que esta passa a ser o objeto de análise, e a partir da qual se extraem conceitos destinados à sua representação condensada, e por meio desta, à sua recuperação. Dentre as definições, e considerando o espaço e o interesse desse capítulo, nunca é demais citar Le Coadic, para quem a informação "[...] é um significado transmitido a um ser consciente por meio de uma mensagem inscrita em um suporte espacial-temporal: impresso, sinal elétrico, onda sonora, etc. Essa inscrição é feita graças a um sistema de signos (a linguagem), signo este que é um elemento da linguagem que associa um significante a um significado: signo alfabético, palavra, sinal de pontuação". (LE COADIC, 2004, p. 4)

Já o conceito, em si, como elemento de significado presente nos diferentes registros, constitui objeto de estudos que procuram justamente entendê-lo tanto do ponto de vista semântico como lógico e lingüístico, a fim de situá-lo como unidade de significado, a qual servirá como base para a construção de linguagens documentárias específicas. O artigo da autoria de Dahlberg (1978) tornou-se clássico nesse sentido.

Com relação à chamada hierarquia da informação, é interessante observar que, além de contemplar os níveis de evolução inerentes à construção do conhecimento, a expressão está atualmente vinculada a uma série de domínios e aplicações que incluem desde processos cognitivos a gestão de conteúdos, webdesign e engenharia de sistemas, conforme demonstra uma busca no Google, por *Information hierarchy,* realizada em maio de 2009 – com cerca de 46.900 resultados. Parece válida ainda a esquematização apresentada por Páez Urdaneta (1992), que descreve quatro componentes da pirâmide do conhecimento, do mais elementar ao mais complexo, os quais contribuem à compreensão do processo informacional no contexto das organizações. São esses: dado; informação; conhecimento; inteligência. O esquema procura, inclusive, fornecer elementos para atribuição de valor à informação, e funciona ainda hoje como referência em trabalhos de pesquisa, seminários e aulas.

Uma visão não hierárquica e questionadora da relação entre informação e conhecimento encontra-se, por outro lado, expressa no trabalho de Silva (2003), o que demonstra também que a relação entre uma coisa e outra continua sendo objeto de reflexão e pesquisa, não parecendo haver uma explicação única ou definitiva para a questão.

A tipologia da informação estaria implícita, por exemplo, nas tabelas de classificação utilizadas tradicionalmente na ordenação física de documentos a partir de classes de assuntos, tais como a Classificação Decimal de Dewey (CDD), Classificação Decimal Universal (CDU), a Classificação da Biblioteca do Congresso (*Library of Congress Classification* – LCC), a Classificação Internacional de Patentes (CIP), entre outras, que a partir de origens filosóficas da identificação e classificação das áreas do conhecimento (Araújo, 2006), funcionam também como pontos de acesso a documentos e informações dos mais variados tipos e em níveis precisos de particularização, a partir de códigos e notações. Nessa perspectiva, torna-se possível, inclusive, identificar diferentes e variados tipos de informação especializada, tais como: informação jurídica, informação econômica, arquitetônica, geográfica, e muitos outros, sejam de natureza científica, artística, gerencial, etc., a partir mesmo das possibilidades de descrição criadas por esses esquemas e tabelas de classificação de assuntos. Talvez não haja exagero em considerar esses códigos de classificação como as representações mais condensadas da informação.

A necessidade de compatibilizar a linguagem natural em que é expressa a necessidade informacional do usuário com aquilo que um sistema de informação pode oferecer como resposta determina também a criação e o desenvolvimento de outros tipos de representação condensada, seja na atribui-

ção de palavras chave ou na elaboração de resumos. A construção de tesauros, vocabulários controlados e índices se vale amplamente das interfaces com a lógica e a lingüística, que auxiliam no estabelecimento de relações hierárquicas e associativas entre termos e conceitos, e no controle terminológico para efeito de desambiguação, entre outros aspectos que repercutem decisivamente no processo da recuperação.

A esse propósito, é interessante observar que a obra de Lancaster (2004) que trata especificamente de indexação e resumos, e que consta em bibliografias da disciplina "Análise da informação", integrante de currículos de biblioteconomia em várias universidades brasileiras, não faz uma referência específica à expressão *análise da informação* como um tópico, em si, ou como processo típico da ciência da informação. A expressão vem sendo aplicada nitidamente em outras áreas mais e menos correlatas, como, por exemplo, na área do jornalismo radiofônico, em negócios, em análise climática, em tecnologia da informação, no ensino a distância, entre outras, tal como demonstra uma busca no Google, por *information analysis*, com retorno de nada menos que 1.060.000 resultados. Percebendo essa fluidez conceitual, a qual se reflete em abundância de referências dispersas ou inespecíficas, Robredo (2005, p.124) dedica todo um capítulo de seu livro *Documentação de Hoje e de Amanhã* à análise da informação, no qual focaliza "[...] a representação do conteúdo informacional dos documentos segundo diversos pontos de vista e possíveis aplicações".

Na literatura anglo-saxônica, análise da informação estaria implícita na análise de conteúdo (*content analysis*), sendo esta entendida, por sua vez, como uma metodologia de pesquisa aplicável a diferentes realidades, como descreve, por exemplo, o estudo de White e Marsh (2006). Observe-se que vários tipos de análise coexistem: análise de conteúdo e análise do discurso (ROCHA; DEUSDARÁ, 2005); análise documentária (KOBASHI, 1996; GUIMARÃES, 2003); análise de assunto (FUJITA, 2003). Tendo em conta as origens multidisciplinares desses e de numerosos outros estudos, parece lícito questionar se todas essas modalidades de análise não estariam incluídas, em maior ou menor grau, no escopo mais amplo da análise da informação como um todo.

Torna-se claro, em todo caso, que no âmbito da ciência da informação, a análise da informação compreende um conjunto não fechado de processos que se caracterizam por graus variáveis de complexidade, e nos quais a representação descritiva e temática – e as linguagens que a viabilizam – assumem fundamental importância, não só diante da massa de informação que circula em todos os sentidos e nos mais variados suportes, mas principalmente em função das necessidades do usuário.

A dimensão conceitual de análise da informação envolve ainda, e cada vez mais, a necessidade de um relacionamento interdisciplinar entre profissionais de informação e profissionais de outras áreas, na medida em que, dependendo do público a que se destina, a informação precisará ser orga-

nizada segundo critérios de exaustividade – mas principalmente de especificidade – que assegurem ao usuário a recuperação precisa e eficaz das fontes e conteúdos que efetivamente busca, em oposição ao mar de referências dispersas, obtidas ou não, a partir da navegação na internet. Nesse sentido, inclusive, se justifica a existência de sistemas especializados (centros de pesquisa, unidades de informação ou bibliotecas especializadas), em que, trabalhando com conceitos e terminologias típicas de diferentes áreas (como, por exemplo, direito, agricultura, saúde, energia atômica, etc.), e procurando interagir com os profissionais dessas áreas, o profissional da informação adquire habilidades que o convertem no especialista em informação jurídica, econômica, química, etc. Por outro lado, no contexto mais genérico e abrangente de uma biblioteca pública, ele se valerá de linguagens já amplamente utilizadas, tais como os cabeçalhos de assunto, podendo contribuir inclusive ao desenvolvimento e atualização desses instrumentos, na medida em que surjam novas áreas de interesse ou de assuntos suscetíveis de serem exploradas no âmbito dessas bibliotecas.

2.2 Dimensão estratégica

A decisão quanto ao projeto e implementação de formas diferentes de se organizar a informação com vistas à sua ágil recuperação requer o estabelecimento de políticas e estratégias por parte de gestores e profissionais. Em outras palavras, da mesma forma que, no ambiente organizacional a informação pode se converter em instrumento para a tomada de decisão e elemento de vantagem competitiva, o seu tratamento, independentemente da natureza da organização (empresarial, acadêmica, etc.) requer um tipo de análise da informação que possa responder a uma série de perguntas que, conforme o caso, se tornam indispensáveis.

Tais perguntas refletem um nível de problematização de questões relacionadas a diferentes opções e possibilidades em termos da adequação de objetivos e metas à missão institucional, e por conseguinte à capacidade de planejamento, execução e acompanhamento de projetos. Nesse caso estariam incluídas, por exemplo, as decisões referentes ao tipo de indexação que seria mais apropriado para determinado acervo. Exaustiva? Específica? Automática? Assistida por computador?

A aquisição ou assinatura de uma base de dados de legislação, por exemplo, constitui sem dúvida uma decisão estratégica em função de seus usuários efetivos e potenciais. Pode ser extremamente útil para determinado público (tribunais, escritórios de advocacia, varas cíveis, etc), porém sem relevância para outro. A construção de repositórios institucionais, em seu turno, demanda claramente o estabelecimento de políticas que determinem desde as publicações a serem nele incluídas, às comunidades participantes e aos metadados que serão utilizados na representação descritiva dos objetos informacionais, entre uma série de outros aspectos. Nas organizações, a transformação do conhecimento tácito em explícito requer uma integração de es-

forços de análise da informação, no sentido de se estabelecer critérios, objetivos, e recursos que deverão ser mobilizados. A conveniência de se criar e desenvolver uma ontologia deve ser objeto de análise da informação. Qual a finalidade? Para que clientela? As respostas a estas e outras indagações refletem sem dúvida a dimensão estratégica de análise da informação, seja no contexto das organizações ou no âmbito da pesquisa científica voltada ao tratamento e organização da informação. Não deixa de ser interessante observar também o duplo aspecto da dimensão estratégica de análise da informação: tanto a informação, em si, pode ser de natureza estratégica para uma organização, como o processo de escolha e decisão quanto à forma de tratá-la e organizá-la refletem as políticas e estratégias adotadas em determinado ambiente.

2.3 Dimensão operacional

Em função dos aspectos conceituais inerentes à informação como fenômeno, processo, e objeto de tratamento e mediação especializada, e tendo em vista o foco específico desse capítulo na organização – e não na recuperação –, a dimensão operacional da análise da informação engloba uma série de etapas e procedimentos que justamente transformam o documento (recurso/-objeto) em informação disponível.

Assim, a análise documentária, sinônimo de análise da informação, visa elaborar representações condensadas do conteúdo informacional de uma população de documentos, estando diretamente relacionada à síntese e ao tratamento da informação, com o objetivo da organização e recuperação de informações e de atender as necessidades dos usuários com maior precisão.

A síntese é a compreensão que se tem do objeto analisado e que será, de alguma forma, representado. Para organizar a informação é preciso descrevê-la. Descrever um objeto é enumerar suas características ou as relações desse objeto com outros que o identifiquem. Portanto, uma descrição necessita de uma linguagem, que para a biblioteconomia é chamada de linguagem bibliográfica.

A descrição física ou representação descritiva do objeto que está sendo analisado é um tipo de linguagem bibliográfica. O processo da descrição física envolve uma operação de identificar as características físicas do objeto tornando-o único, e inclui, por exemplo, elementos como autor, título, editor, ISBN, etc. Os elementos identificados são termos selecionados que compõem um vocabulário que está representado em códigos. Os códigos são compostos por campos, descritores ou metadados aplicados num conjunto de regras definidas em âmbito internacional como é o caso das *Anglo-American Cataloguing Rules* (AACR), o *Código de Catalogação Anglo-Americano,* que vem sendo revisado inúmeras vezes. Cabe aqui incluir, sob a visão de Taylor e Joudrey (2009 p. 9), o modelo conceitual para dados bibliográficos e de autoridade que está sendo desenvolvido pela *International Federation of Library Associations and Institutions* (IFLA), a Federação Internacional de Associações e Instituições Bibliotecárias: os *Functional Requirements for Bibliographic*

Records (FRBR) – Requisitos Funcionais para Registro Bibliográfico, e os *Functional Requirements for Authority Data* (FRAD) – Requisitos Funcionais para Dados de Autoridade. Ainda, segundo as autoras supracitadas, está agendado para 2009 a publicação do novo código de regras chamado *Resource Description and Access* (RDA), a ser traduzido para algo como Descrição de, e Acesso a Recursos.

Outro tipo de linguagem bibliográfica é a que descreve as informações contidas no documento, também denominadas de descrição de conteúdo ou representação temática. A descrição de conteúdo é um conjunto de operações que descrevem o assunto de um documento, também denominado na ciência da informação por atinência[4]. E que, por meio da abstração da essência temática e de termos e/ou símbolos escolhidos melhor representem o conteúdo informacional. Dos produtos gerados pelo processo operacional da análise da informação citam-se:

Classificação **–** linguagem documentária na qual os descritores permitem representar os conceitos e objetos de uma área do conhecimento de forma sistemática e hierárquica. Assim definida por Souza (1943 p. 21) como "o arranjo dos livros em classes de assuntos, ao mesmo tempo que lhes destinam lugares nas estantes, de acordo com esses assuntos". De acordo com sua estrutura hierárquica, as classificações atribuem, a cada descritor um indicador que pode ser formado por números, letras ou uma combinação de números e letras, identificando a qual grupo pertence.

Indexação **–** processo no qual se escolhe o termo ou os termos mais adequados para descrever o conteúdo de um documento. O produto dessa indexação são os índices e o nível da indexação varia de acordo com as necessidades dos usuários e das unidades de informação. Segundo Lancaster (2004 p. 9), o indexador deve formular algumas perguntas sobre o documento, a fim de identificar o objetivo do documento e responder às questões de interesse dos usuários. De que trata? Por que foi incorporado ao acervo? e Quais de seus aspectos serão de interesse para nossos usuários?

Sumarização ou resumo **–** "representação sucinta, porém exata, do conteúdo de um documento", assim definido por Lancaster (2004 p. 100). Um resumo deve ser breve, exato e claro. Consolidado, inclusive pela Associação Brasileira de Normas Técnicas, e segundo a norma (NBR 6028, 2003), o resumo pode ser: informativo, indicativo e crítico. O resumo informativo, informa ao leitor finalidades, metodologias, resultados e conclusões do documento, de tal forma que este possa, inclusive, dispensar a consulta ao original. O resumo indicativo porém, indica apenas os pontos principais do documento, não apresentando dados qualitativos, quantitativos, etc. De modo geral, não dispensa a consulta ao original. E, diferentemente dos outros dois, o resumo crítico

[4] O termo atinência está sendo usado para a tradução do termo inglês *aboutness,* traduzido por Antônio Agenor Briquet de Lemos no livro *Indexação e resumos: teoria e prática* de F.W. Lancaster (2004).

é redigido por especialistas com análise crítica de um documento. Também chamado de resenha. Quando analisa apenas uma determinada edição entre várias, denomina-se recensão. A análise da informação envolve, também, a comunicação entre autor-texto-leitor e um momento de tomada de decisão. O processo de leitura documentária, de acordo com Cintra (2002) possui alguns fatores que influenciam a análise como: a qualidade do texto, que implica na veracidade da informação e nível de estruturação do texto; o conhecimento prévio do leitor ou enciclopédia particular; as estratégias de leitura; objetivo da instituição e terminologias da área. A análise da informação, no contexto da organização e representação do conhecimento, busca aporte teórico em relações interdisciplinares com a terminologia e a lingüística, principalmente para a elaboração dos sistemas de representação do conhecimento, onde existe uma abordagem diferenciada no tratamento dos termos escolhidos para representar um domínio, visando maior precisão na busca e a eliminação de problemas, como a ambigüidade na interpretação, fato comum na linguagem natural.

Terminologia – campo do conhecimento que estuda as linguagens de uma determinada área. Essas linguagens são compostas por um conjunto de termos que representam um sistema de conceitos de uma área específica, chamadas de linguagens especializadas. Trabalha com o termo, o conceito e as relações entre conceitos. No contexto da análise da informação, estuda-se o controle terminológico observando-se a: homonímia – uma palavra que remete para vários significados, sem que haja relação semântica entre os termos; polissemia – palavras com a escrita igual e significados semelhantes, porém há uma relação semântica entre os termos; e sinonímia – várias palavras com um mesmo significado.

Lingüística **–** a matéria da lingüística é constituída por todas as manifestações da linguagem humana (SAUSSURE, 1977), tendo como objeto de estudo a descrição e análise da capacidade da linguagem levando em conta os aspectos fundamentais: estrutura, uso e forma. No nível descritivo dos sistemas lingüísticos damos destaque a: morfologia – que estuda o morfema, unidade mínima que forma uma palavra; sintaxe – estuda os princípios e processos que presidem a construção de frases em línguas particulares; e semântica – que estuda os significados das palavras num nível profundo de conceito.

3. O CONTEXTO DA ANÁLISE DA INFORMAÇÃO

A análise da informação é composta por ações de descrição, representação e organização de conteúdos visando à construção de linguagens documentárias voltadas para o armazenamento e a recuperação da informação. Dessa forma, a análise da informação deve ser recortada da compreensão comum de avaliação de conteúdos informacionais para dar simplesmente conhecimento sobre especialidades ou mesmo apoiar o processo decisório, para um tipo de análise que é foco da tradução sintético-analítica da informação.

Uma visão esquemática das ações básicas da análise da informação, correlacionadas com o ciclo documentário, pode ser cotejada na *Figura 1*.

De outro modo, pode-se delimitar o escopo e o contexto do trabalho de análise da informação afirmando que as ações de descrição, representação e organização de conteúdos possuem mecanismos associados às próprias características das etapas que compõem o ciclo documentário. Essa correlação inclui ações preliminares e complementares, que descrevem todo o processo de análise da informação:

Figura 1 – As ações básicas da análise da informação e o ciclo

I. Ações preliminares:

- Dada uma população de documentos ou unidades bibliográficas, seleciona-se uma amostra de documentos que se associam tematicamente a interesses específicos de usuários ou à própria especificidade do sistema de informação. Esse mecanismo coincide com a etapa de *seleção* no ciclo documentário; e
- Captura ou incorporação da amostra de documentos e unidades bibliográficas que serão matéria-prima ou alvo do processo de análise da informação. Essa característica associa-se às etapas de *aquisição* e *registro* no ciclo informacional.

II. Ações de descrição:

- A Identificação e registro das características que descrevem de modo representativo os documentos visa à apreensão de elementos textuais, a fim de formar juízo sobre o conteúdo dos documentos em análise. Esse mecanismo é o que se denomina, no âmbito do ciclo documentário, de *descrição bibliográfica*. Os metadados de uma unidade bibliográfica são, por excelência, resultado da identificação e registro da descrição bibliográfica; e
- Tradução sintético-analítica do conteúdo dos documentos e unidades bibliográficas. Tem por objetivo compreender a essência informativa de cada item a ser analisado, por meio de um resumo ou frases que contenham o tema central do conteúdo dos documentos. A tradução sintético-analítica está relacionada com a etapa de *análise* ou *condensação* do ciclo documentário.

III. Ações de representação:

- A mediação entre a linguagem natural e a linguagem documentária é o mecanismo que corresponde à tradução de um documento ou unidade bibliográfica em termos documentários. Esse mecanismo confunde-se com a própria definição de indexação que, segundo Araújo Junior (2007), envolve uma leitura analítica do documento, a fim de identificar e selecionar palavras-chave (indexadores) que possam representar de forma fidedigna o seu conteúdo. Esse mecanismo corresponde à etapa de *indexação* no ciclo documentário; e
- Escolha de palavras ou conjunto de palavras que de modo sumário representem o conteúdo dos documentos; corresponde ao próprio processo de indexação onde a familiarização, análise e a conversão de conceitos em descritores são, de acordo com Rowley (1988), os três estágios centrais do processo. As ações de representação impactam diretamente na etapa de *armazenamento da representação condensada dos documentos* no ciclo documentário. Dentre as ações de descrição, típicas da análise da informação, os elementos descritivos do conteúdo do documento são os dados que caracterizam, de forma condensada e unívoca, cada documento, permitindo que estes elementos possam ser registrados em diversos suportes para posterior processamento.

IV. Ações complementares:

- O acompanhamento e a avaliação do processamento da informação são mecanismos complementares à atividade da análise da informação, pois têm como finalidade ajustar incongruências entre a representação temática e o processamento efetivo da informação.
- O acompanhamento pode ser considerado como uma espécie de controle de qualidade da análise, pois é um mecanismo fundamental para a geração de produtos a serem utilizados no tratamento da informação para sua posterior recuperação pelos usuários.
- O mecanismo de avaliação permite a verificação da adequação conceitual das palavras, ou do conjunto delas, na representação do conteúdo dos documentos de modo unívoco e sumário, a fim de garantir a representatividade dos termos e assertividade na recuperação da informação.

Essas últimas ações correspondem, de forma equivalente, às etapas de *processamento da informação condensada* e *produtos do processamento* no ciclo documentário.

A *Figura 2* apresenta a posição do acompanhamento e da avaliação como ações complementares da análise da informação.

A descrição temática é um dos principais objetivos da análise da informação e está diretamente associada ao processo de sua recuperação, pois a busca pelos termos que representam de fato a informação analisada é um fator crítico de sucesso. Nesse sentido, a análise da informação abrange de

modo similar ao ciclo documentário as etapas de *interrogação e busca*, onde o usuário vai *recuperar a informação* que completa o sentido e a finalidade do ciclo documentário como também da análise da informação.

Figura 2 – Acompanhamento e avaliação na análise da informação

Figura 3 – Similaridades entre a análise da informação e a interrogação e busca

A *Figura 3* ilustra as similaridades das ações da análise da informação com as etapas finais do ciclo documentário de interrogação e busca voltadas para a recuperação da informação.

Considerando a finalidade do trabalho de análise da informação que é a recuperação da informação, descreve-se a seguir o contexto dessa atividade no campo de pesquisa da ciência da informação.

A ciência da informação é o campo de pesquisa que investiga os processos e os sistemas que envolvem coleta, processamento e disseminação da informação. Todavia, o conceito da ciência da informação não é uma verdade absoluta, pois as contribuições que buscaram delimitar essa ciência foram provenientes de inúmeros campos do saber e disciplinas distintas.

Em clássico artigo, Wersig e Neveling (1975) constataram que as diversas formações das pessoas que ingressaram em meados do século XX no campo de pesquisa da ciência da informação, que ainda não tinha nenhum sistema educacional estabelecido, foram provocadas por uma série de diferentes interesses envolvidos com o trabalho de informação. Essa situação permitiu que cada participante da discussão pudesse concordar sobre a existência de algo chamado ciência da informação, desde que estivesse baseada em sua formação específica (WERSIG; NEVELING, 1975).

Ainda hoje, o objeto de estudo da ciência da informação é uma indagação sem resposta ou em plena construção. Entretanto, há uma definição proposta por Borko em 1968 que se mantém atual, sobretudo considerando os problemas causados pela ampla polissemia do termo informação e a diversidade de contribuições de outros campos de pesquisa. Para o autor, a ciência da informação é a disciplina encarregada da investigação de um corpo de conhecimento que congrega a origem, coleta, organização, armazenamento, recuperação, interpretação, transmissão e uso da informação.

Em um estudo empreendido no ano de 1999 pelo *Institute of Information Scientists* da Inglaterra, citado por Robredo (2003), um grupo de cientistas da informação trabalhou em uma proposta de consolidação da ciência da informação, a partir da ampla aceitação da polissemia do termo informação. Esta aceitação se apóia no fato da transformação da ciência da informação de uma disciplina eminentemente acadêmica na década de 50, para uma plataforma voltada para a prática na década de 90 do século passado. Resta ainda, com efeito, discutir se a ciência da informação é ciência ou uma arte prática. Diante disto, Summers *et al*. (1999 *apud* ROBREDO, 2003) questionam o que realmente significa a expressão 'cientista da informação'. A resposta parece estar nos critérios para a ciência da informação representados na *Tabela 1*, que proporcionam uma ênfase, com a qual concorda Robredo (2003), sobre os aspectos práticos em contraste com a ausência de ênfase na compreensão das bases teóricas da profissão.

Utilizando a *Tabela 1*, pode-se identificar aonde é factível encaixar a atividade da análise da informação, que parece estar bem assentada como uma das atividades núcleo de pesquisa para a ciência da informação.

- A área núcleo congrega as subáreas de pesquisa que foram, desde os primórdios, centrais e delimitadoras da empresa científica proposta para a ciência da informação, ou seja, coleta, processamento e disseminação da informação. Daí a relevância que possuem todas as atividades que são partes integrantes do trabalho de análise da informação:
- *Construção de linguagens documentárias* – atividade que subsidia sistemas e unidades de informação na descrição do conteúdo dos documentos, a fim de prepará-los para a armazenagem e posterior recuperação das informações que contêm;
- *Análise documentária* - parte sempre do princípio da compreensão do significado e envolve a descrição de conteúdo que é o conjunto de operações que descreve os assuntos de um documento e os produtos dessas operações; e
- *Descrição temática da informação* - principal objetivo da análise da informação. Está diretamente associada à recuperação da informação, pois a busca pelos termos que representam de fato a informação analisada é um fator crítico de sucesso para a sua recuperação.

A *Figura 4* apresenta, de modo espacial, a posição da atividade de análise da informação no contexto da área núcleo de pesquisa em ciência da informação proposta por Summers *et al.* (1999, *apud* ROBREDO, 2003).

Tabela 1 – Critérios para a Ciência da Informação

Seção 1 (Área Núcleo: Ciência da Informação)

Teoria e prática da geração, aquisição, avaliação e validação, organização, armazenagem, transmissão, recuperação e disseminação da informação

Informação: características, fornecedores e usuários
Fontes de informação
Armazenagem e recuperação da informação
Análise da informação
Teoria da ciência da informação

Seção 2 (Gestão da informação)

Gestão de todos os recursos de informação da organização

Planejamento
Comunicações
Informação gerencial e sistemas de controle
Gestão de recursos humanos
Gestão financeira
Promoção, economia e *marketing*
Fatores políticos, éticos, sociais e legais

Seção 3 (Tecnologia da informação)

Tecnologia que pode ser usada em Ciência da Informação e Gestão da Informação

Sistemas computadorizados: *hardware* e *software*
Telecomunicações
Aplicações da tecnologia da informação
Meio ambiente

FONTE: Summers et al. (1999 apud ROBREDO, 2003).

Figura 4 – A análise da informação no contexto da pesquisa em ciência da informação

4. CONSIDERAÇÕES FINAIS

Considerando-se a amplitude de significados e aplicações da expressão "análise da informação", procurou-se nesse capítulo contextualizá-la no âmbito da ciência da informação, abordando aspectos relacionados às suas dimensões conceitual, estratégica e operacional, bem como às ações que desembocam no processo, propriamente dito, de análise da informação, ações essas análogas ao ciclo documentário, e que, na fase da organização, resultam na descrição e representação do objeto informacional, tornando-o passível de ser recuperado, disponibilizado, reproduzido, disseminado, reutilizado e relacio-

nado a outros objetos, de acordo com critérios que melhor atendam as necessidades dos usuários. De fato, é o conceito de representação que melhor traduz o objetivo de análise da informação, enquanto mediação entre a informação e seu usuário.

Embora a mediação se faça necessária tanto na organização (pelos critérios, instrumentos e linguagens de representação) como na recuperação (negociação usuário-sistema; usuário-bibliotecário, etc.), o foco concentrou-se na organização, já que é nessa fase que as linguagens documentárias adquirem maior importância, na medida em que é por meio delas que se descrevem suportes e se representam conteúdos temáticos. A esse propósito, é interessante observar que a tendência atual é considerá-los como elementos indissociáveis na representação de um objeto. A partir do conjunto de dados e informações que o descrevem torna-se possível não só sua identificação unívoca e sua localização, como também a apreensão mais completa de conteúdos, significados e ainda de relacionamentos possíveis entre diferentes objetos, tanto em nível formal como semântico.

Nessa perspectiva, poder-se-ia argumentar que a análise da informação é por demais vasta e abrangente, não comportando uma delimitação conceitual clara, ou ainda, que talvez fosse mais prático substituí-la por conceitos específicos como aqueles relacionados à indexação, elaboração de resumos, classificação, catalogação, na medida em que essas atividades constituem desde sempre processos reconhecidamente típicos da biblioteconomia, e que sobre cujas finalidades não pairam maiores dúvidas.

Contudo, ao considerar-se o escopo mais amplo da ciência da informação, que hoje em dia procura conectar, entre outros, uma série de conhecimentos oriundos da tecnologia da informação e da biblioteconomia, a análise da informação assume a dimensão de um amplo conjunto não fechado de processos que busca, em última análise, encurtar o caminho que vai da busca à obtenção efetiva da informação que interessa ao usuário. Este, por sua vez, se torna cada vez mais exigente, justamente em função das múltiplas opções de acesso à informação que a tecnologia parece viabilizar nos dias de hoje.

Se no ambiente da biblioteca tradicional a descrição bibliográfica/representação descritiva, que sempre englobou as atividades citadas, se mostrou durante longo tempo eficaz na identificação, seleção, e principalmente na localização de documentos, no mundo da comunicação eletrônica da atualidade tais práticas se tornaram isoladamente insuficientes, razão pela qual a análise da informação precisa ser necessariamente interdisciplinar. Ao articular tecnologia com lingüística e lógica, para citar apenas esses exemplos, ela cria os instrumentos necessários não só à identificação, como à localização, e à navegação segura do usuário em meio ao mar nem sempre sereno da informação abundante, multidirecionada, oculta, de acesso restrito, etc. É que hoje, não se trata mais da localização pura e simples de documentos, porém, sem excluir essa função básica e sempre importante, trata-se principalmente do

intercâmbio de informações que ocorre em nível planetário. Ao representar corretamente os conteúdos temáticos presentes nos mais variados objetos, a análise da informação pode dar o leme seguro ao navegador.

5. REFERÊNCIAS DO CAPÍTULO 3

ARAÚJO, Carlos Alberto Ávila. Fundamentos teóricos da classificação. *Encontros Bibli,* Florianópolis, n.22, 2006. Disponível em: <http://www.periodicos.ufsc.br/index.php/eb/article/viewFile/296/368> Acesso: março 2009.

ARAÚJO Junior, R.H. de. *Precisão no processo de busca e recuperação da informação*. Brasília: Thesaurus, 2007.

BORKO, H. Information science: what is? *American Documentation* v.1, n.19, p. 3-5, January, 1968.

CINTRA, A.M.M *et al*. *Para entender as linguagens documentárias*. São Paulo: Polis, 2002. (Coleção Palavra-Chave, 4)

DAHLBERG, Ingetraut. Teoria do conceito. *Ciência da Informação,* Brasília, v.7, n.2, p. 101-107, 1978.

DERVIN, Brenda. Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of Knowledge Management,* Ohio State University v. 2, n. 2, Dec. 1998. Disponível em: http://communication.sbs.ohio-state.edu/sensemaking/zennez/-zennezdervin98km.pdf Acesso: maio 2009.

FUJITA, Mariângela Spotti Lopes. A identificação de conceitos no processo de análise de assunto para indexação. *Revista Digital de Biblioteconomia e Ciência da In formação*, Campinas, v.1, n.1, jul/dez 2003. Disponível em: http://server01.bc.unicamp.br/seer/ojs/viewarticle.php?id=9&layout=abstract Acesso: abril 2009.

GUIMARÃES, José Augusto Chaves. A análise documentária no âmbito do tratamento da informação: elementos históricos e conceituais. *In:* RODRIGUES, Georgete Medleg; LOPES, Ilza Leite, (orgs.). *Organização e representação do conhecimento na perspectiva da ciência da informação.* Brasília: Thesaurus, 2003 (Estudos avançados em ciência da informação, v. 2), p.100-117

KOBASHI, Nair Y. Análise documentária e representação da informação. *Informare,* Rio de Janeiro, v.2, n.2, p.5-27, jul-dez 1996.

LANCASTER, F. W. Indexação e resumos: teoria e prática. 2. ed. rev. e atual. Brasília, DF: Briquet de Lemos/ Livros, 2004. 452 p. ISBN: 85-85637-24-2

LE COADIC, Yves François. *A ciência da informação*. 2. ed. rev. e atual. Brasília, DF: Briquet de Lemos/Livros, 2004. ISBN: 85-85637-23-4

NBR 6028: Informação e documentação - Resumo - Apresentação. Rio de Janeiro: ABNT, 2003.

PAEZ-URDANETA, Iraset. Ges*tión de la inteligência, aprendizaje tecnológico y modernización del trabajo informacional: retos y oportunidades.* Caracas: Unversidad Simón Bolívar, 1992.

ROBREDO, J. *Da ciência da informação revisitada aos sistemas humanos de informação*. Brasília: Thesaurus; SSRR Informações, 2003.

ROBREDO, Jaime. *Documentação de hoje e de amanhã.* 4ª ed. rev. e ampl. Brasília, DF: Edição de autor, 2005. ISBN 85-905920-1-4

ROCHA, Décio; DEUSDARÁ, Bruno. Análise de conteúdo e análise do discurso: aproximações e afastamentos na construção de uma trajetória. *Alea*, v.7, n.2, jul/dez 2005. Disponível em: http://www.scielo.br/pdf/alea/v7n2/a10v7n2.pdf Acesso: abril 2009.

ROWLEY, J. *Abstracting and indexing*. 2th edition. London: Clive Bingley, 1988.

SAUSSURE, F. *Curso de Lingüística Geral.* 8ª ed. São Paulo: Editora Cultrix, 1977. 279 p.

SILVA, Armando Malheiro da. Conhecimento/Informação: sinonímia e/ou diferenciação? *In*: RODRIGUES, Georgete Medleg; LOPES, Ilza Leite, (orgs.). *Organização e representação do conhecimento na perspectiva da ciência da informação*. Brasília: Thesaurus, 2003 (Estudos avançados em ciência da informação: v. 2), p.23-41

SOUZA, J.S. de*. Classificação: sistemas de classificação bibliográfica*. Rio de Janeiro: Imprensa Nacional, 1943. 162 p.

TAYLOR, A. G.; JOUDREY, D.N. *The Organization of Information*. 3ª ed. London: Libraries Limited. 2009. 512 p

WERSIG, G.; NEVELING, U. The phenomena of interest to information science. *Information Scientist*, v.9, n.4, p.127-140, December, 1975.

WHITE, Marilyn Domas; MARSH, Emily E. Content analysis: a flexible methodology. *Library Trends*, v. 55, n.1, p. 22, Summer 2006.

Capítulo 4

$FRBR_{OO}$ – Conceitos Básicos

João Alberto de Oliveira Lima[1]

SUMÁRIO DO CAPÍTULO 4

Resumo do Capítulo 4, p. 82
Como citar o Capítulo 4, p. 82

1. INTRODUÇÃO DO CAPÍTULO 4, *p. 83*
2. MODELO CIDOC CRM, *p. 83*
2. POUCAS PALAVRAS SOBRE "ONTOLOGIA", *p. 84*
4. NOTAÇÃO DO CIDOC CRM, *p. 86*
5. VISÃO GERAL DO CIDOC CRM, *p. 86*
6. MODELO FRBRoo, *p. 87*
7. HARMONIZAÇÃO DAS ENTIDADES DO GRUPO 1 DO $FRBR_{ER}$, *p. 88*
8. CONCLUSÃO DO CAPÍTULO 4, *p. 91*
9. REFERÊNCIAS DO CAPÍTULO 4, *p. 92*

[1] Doutor em Ciência da Informação. Programa de Pós-graduação em Ciência da Informação (PGCInf), - Faculdade de Ciência da Informação da Universidade de Brasília (UnB/FCI}). Senado Federal. CV Lattes: http://cnpq.br/5218258152780642. E-mail: joaolima@senado.gov.br.

Resumo do Capítulo 4

O modelo conceitual FRBR (*Functional Requirements for Bibliographical Records*) da IFLA (*International Federation for Library Associations and Institutions*) vem mudando a forma de criação e organização dos catálogos bibliográficos. Na sua versão original, o modelo foi estruturado segundo a notação entidade-relacionamento. Atualmente, está sendo proposta uma nova versão do FRBR que adotou outra técnica de modelagem: a orientação a objeto. Este capítulo apresenta os conceitos básicos desta nova abordagem capaz de modelar as informações não só das bibliotecas, como também dos arquivos e dos museus.

Palavras-chave: FRBR; FRBR$_{OO}$; FRBR$_{ER}$; CIDOC CRM; Representação da informação; Patrimônio cultural; Ontologia; Informação bibliográfica; Informação museológica; Informação arquivística..

Abstract

FRBR$_{oo}$ – Core concepts

The IFLA (International Federation for Library Associations and Institutions) FRBR (Functional Requirements for Bibliographical Records) conceptual model is changing the way we create and organize bibliographic catalogs. In its original version, the model was structured according to the entity-relationship notation. Currently, there is a proposal of a FRBR new version that adopted another modeling technique: the object orientation. This chapter introduces the basic concepts of this new approach which can model not only the library information but also the information from archives and museums.

Keywords: FRBR; FRBR$_{OO}$; FRBR$_{ER}$; CIDOC CRM; Information representation; Cultural heritage; Ontology; Bibliographic information; Museological information; Archival information.

Como citar o Capítulo 4

LIMA, J. A. O. FRBR$_{OO}$ – Conceitos básicos. *In* : Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios pelo bosque da informação*: *Estudos sobre representação e organização da informação e do conhecimento*. Brasília DF: IBICT, 2010. 335 p. ISBN: 978-85-7013-072-3. Capítulo 4, p. 81-92. Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

1. INTRODUÇÃO DO CAPÍTULO 4

É inegável o impacto do modelo FRBR na organização da informação bibliográfica. Esse impacto pode ser percebido pela leitura do novo código de catalogação RDA (*Resource Description and Access*), sucessor do AACR2 (*Anglo-American Cataloguing Rules*), que incorpora as entidades e relacionamentos do modelo da FRBR, ou ainda pela análise da extensa bibliografia (IFLA 2010) disponível no sítio da IFLA.

Após a publicação do FRBR em 1998 (IFLA 1998), o *FRBR Review Group*, entidade vinculada à Seção de Catalogação da IFLA, estabeleceu alguns grupos de trabalho. O "*Working Group on FRBR/CRM Dialogue*", criado em 2003, teve como objetivo a preparação de uma versão orientada a objeto do modelo FRBR compatível com o CIDOC CRM (*Comité International pour la Documentation - Conceptual Reference Model*). O modelo resultante dessa harmonização é muito mais do que uma nova expressão do FRBR. Representa uma visão unificada sobre as entidades do patrimônio histórico-cultural com capacidade para representar as informações das três instituições de memória: bibliotecas, museus e arquivos. É importante ressaltar ainda que essa nova versão corrige inconsistências semânticas e agrega a dimensão temporal que permite modelar todo o processo criativo considerando não só o ciclo de criação da obra pelo autor como também o trabalho do editor.

Este artigo, após uma rápida introdução ao modelo CIDOC CRM, apresenta os conceitos básicos da versão orientada a objeto do FRBR, referenciada a partir deste ponto pela sigla FRBR$_{OO}$. A versão original do FRBR, por utilizar a notação Entidade Relacionamento, passará a ser referenciada pela sigla FRBR$_{ER}$[2].

2. MODELO CIDOC CRM

O modelo CIDOC CRM foi desenvolvido pelo Grupo de Padrões de Documentação do Conselho Internacional de Museus (ICOM - *International Council of Museums*). De acordo com o sítio oficial (CRM, 2010) o modelo CRM: " [...] representa uma 'ontologia' para a informação do patrimônio cultural, isto é, ele descreve em uma linguagem formal os conceitos implícitos e explícitos e os relacionamentos relevantes para a documentação do patrimônio cultural. O papel principal do CIDOC CRM é o de servir como base para mediação de informações do patrimônio cultural."

O modelo CRM começou a ser criado em 1996 utilizando a mesma técnica do FRBR: a Modelagem Entidade Relacionamento (MER). Já no início, o grupo de trabalho havia identificado cerca de 400 entidades com centenas de relacionamentos. Com a adoção do novo paradigma, a orientação a objeto, o modelo passou a ser estruturado na forma de uma ontologia, e, como conseqüência, ocorreu uma redução significativa no número de entidades (80 classes) e relacionamentos (60 propriedades). Apesar da simplificação em relação à quantidade de entidades, a ontologia resultante passou a ser mais genérica do que o modelo anterior, e, conseqüentemente passou a ter uma maior aplicabilidade. A versão atual do CIDOC CRM possui 80 classes e 132 propriedades, e representa uma agregação da semântica de centenas de modelos (Doerr, 2009, p. 468-9). Desde setembro de 2006, o modelo CIDOC CRM foi aceito como padrão ISO 21.127 (2006).

Segundo Crofts (2004, p. 107), o CIDOC CRM foi desenvolvido, baseado em experiências de projetos de integração de informação, com os seguintes objetivos em mente:

a) tratar de todos os aspectos da documentação do patrimônio cultural necessária para o intercâmbio de informações em um contexto global;
b) permitir a documentação de conhecimento contraditório e parcial;
c) permitir a integração e o intercâmbio sem perda semântica entre esquemas relativamente ricos e pobres;

[2] Os atributos das entidades do modelo FRBR$_{ER}$ são apresentados neste livro no capítulo "Atributos dos Requisitos Funcionais para Registros Bibliográficos" de autoria de Fernanda Moreno.

d) prover uma infra-estrutura extensível e claramente definida para desenvolvimento futuro.

Neste artigo não detalharemos o CIDOC CRM (ICOM 2010), pois se trata de um modelo bem documentado no sítio oficial, como também não analisaremos em detalhes as vantagens e desvantagens da abordagem de ontologias *vs*. entidade relacionamento. No entanto, passearemos rapidamente por esses temas com o objetivo de tocar no conceito de "ontologia" e preparar o terreno para expor os diferenciais do FRBR$_{OO}$.

3. POUCAS PALAVRAS SOBRE "ONTOLOGIA"

O termo "ontologia" vem sendo cada vez mais utilizado em diversas áreas. Guarino (1998) afirma que apesar de, no passado, ter sido objeto de estudo apenas da Filosofia, a sua importância tem sido reconhecida nos mais diversos campos de pesquisa, tais como: engenharia do conhecimento, representação do conhecimento, modelagem de informação, integração de informação, análise orientada a objeto, recuperação e extração de informação, organização e gerência do conhecimento, projeto de sistemas baseados em agentes, projeto de banco de dados, entre outros.

Uma das definições mais citadas de "ontologia" é a de Gruber (1993): "ontologia é uma especificação explícita de uma conceitualização". Guarino & Giaretta (1995) ponderam que, dentre sete definições analisadas, esta é uma das mais problemáticas. Guarino (1998) oferece a seguinte definição: "Uma ontologia é uma teoria lógica que corresponde ao significado intencional de um vocabulário formal, ou seja, um comprometimento ontológico com uma conceitualização específica do mundo. Os modelos pretendidos da linguagem lógica usando este vocabulário são delimitados pelo comprometimento ontológico." No mesmo artigo, Guarino (1998) relaciona os seguintes tipos de ontologias:

- Ontologias de Alto Nível – descrevem conceitos gerais como espaço, tempo, matéria, objeto, evento, ação, de forma independente de um domínio ou problema particular;
- Ontologia de Domínio – descreve o vocabulário relacionado a um domínio genérico (como Medicina, Automóveis) pela especialização dos conceitos definidos na ontologia de alto nível;
- Ontologia de Tarefa – descreve o vocabulário relacionado a uma tarefa ou atividade genérica (como Diagnóstico, Vendas) pela especialização dos conceitos definidos na ontologia de alto nível;
- Ontologia de Aplicação – descreve conceitos dependentes, tanto de um domínio como de uma tarefa particular, utilizando a especialização das respectivas ontologias.

Uma ontologia é estruturada utilizando a relação taxonômica, também conhecida como relação gênero-espécie ou *is-A* (é-UM). Esse relacionamento é representado graficamente nos diagramas do CIDOC CRM por uma seta de linha dupla direcionada para a superclasse, conforme *Figura 1*.

Figura 1. Hierarquia de Classes

Um dos diferenciais do uso da relação taxonômica é que todas as definições da superclasse se aplicam, de forma implícita, às subclasses. Esse mecanismo é denominado herança. Nessa abordagem, podemos dizer que a subclasse herda as propriedades da superclasse. O mecanismo de herança permite representar uma rede complexa de relacionamentos de uma maneira relativamente simples (LE BOUEF, 2009, p.24).

A compreensão do mecanismo de herança é importante para o entendimento do trabalho de harmonização do CIDOC CRM com o modelo FRBR pois, na essência, esse trabalho consistiu no posicionamento das entidades do modelo FRBR na hierarquia de classes do CIDOC CRM. Por exemplo, a entidade "Item" do modelo FRBR$_{ER}$ foi harmonizada como a classe "*F5 Item*", subclasse de "*E84 Information Carrier*".
A *Figura 2* apresenta a hierarquia taxonômica da classe "*F5 Item*" desde a classe "*E1 CRM Entity*" que contém todas as outras classes. Cada uma das classes listadas possui uma ou mais propriedades que são herdadas pelas subclasses. Por exemplo, a propriedade "*P105 right held by*" da classe "*E72 Legal Object*" que endereça a classe "*E39 Actor*" permite modelar a relação de propriedade de um "*F5 Item*" por uma Biblioteca (*F11 Corporate Body*) ou uma Pessoa (*F21 Person*).

Figura *2. Hierarquia da Classe "F5 Item"*

Um caso especial de herança, conhecido como herança múltipla, ocorre quando uma subclasse herda propriedades de duas superclasses. Utilizando um exemplo de Le Bouef (2009, p.24), ao se modelar a entidade "Árvore" como subclasse de "Organismo Vivo" e de "Recurso Natural", se declara de forma implícita que a entidade "Árvore" herda todos atributos e propriedades das classes "Organismo Vivo" e "Recurso Natural". A *Figura 3* representa graficamente o exemplo citado.

Figura 3. Herança Múltipla

Ao analisar as entidades do modelo FRBR$_{ER}$ sob esse novo prisma, percebeu-se a necessidade de criação de novas entidades com o objetivo de preencher lacunas e corrigir inconsistências do modelo original. Por exemplo, a entidade "Manifestação" do modelo FRBR$_{ER}$ foi harmonizada como duas classes: "*F4 Manifestation Singleton*" como o objeto físico "original do autor" e "*F3 Manifestation Product Type*" como o tipo de produto, que é um objeto conceitual.

4. NOTAÇÃO DO CIDOC CRM

As classes e propriedades no modelo CIDOC CRM são identificadas pelas letras "E" e "P", respectivamente, seguidas de um número e de um nome, a exemplo da *Figura 2*. As classes e propriedades resultantes do processo de harmonização são identificadas pelas letras "F" e "R", respectivamente, seguidas de um número e de um nome.

A definição de uma classe é realizada pela especificação dos seguintes elementos: Nome da Classe, Lista de subclasses, Lista de superclasses, Nota de Escopo, Exemplos e Propriedades. Por sua vez, a definição de uma propriedade compreende os seguintes elementos: Nome da Propriedade, Classe de Domínio, Classe de Imagem, Lista de subpropriedades, Lista de superpropriedades, Cardinalidade, Nota de Escopo e Exemplos.

5. VISÃO GERAL DO CIDOC CRM

Uma das grandes inovações do modelo CIDOC CRM é a estruturação das informações em torno dos eventos temporais, em oposição aos modelos tradicionais de metadados, como o MARC e o Dublin Core, que têm o "recurso" como entidade central. Nesta abordagem inovadora, os eventos são definidos como entidades que agregam atores (pessoas ou instituições), coisas (físicas e abstratas), local e uma duração no tempo. Múltiplos nomes, identificadores e tipos podem ser atribuídos a todas as entidades do modelo. A *Figura 4* apresenta uma visão geral dessas entidades citadas e percebe-se a centralidade da entidade temporal.

Figura 4. Entidade Temporal como entidade central do CIDOC CRM.
Fonte: Doerr (2002, adap.)

A ênfase nas entidades temporais pode ser questionada ao considerarmos que normalmente o usuário está interessado em um recurso específico, como, por exemplo, uma obra. No entanto, o usuário terá sua necessidade de informação satisfeita se, como resultado da pesquisa de uma obra, ele for apresentado a uma lista de eventos nos quais a obra participa, mostrando, por exemplo, os eventos relacionados à criação da obra e às publicações das diversas edições. Segundo Patrick Le Bouef (2009, p. 33), essa característica restitui à catalogação uma dimensão filológica.

É importante ressaltar que o modelo FRBR$_{ER}$ não possui entidades para modelar os eventos relacionados a uma obra, restringindo-se apenas à definição da entidade "Evento" (Grupo 3) utilizada de forma exclusiva na indexação de assunto de uma Obra. Segundo Bekiari et al. (2008), essa lacuna já havia sido relatada por vários pesquisadores, tais como, Heaney (1997), Fitch (2004), Lagoze (2004) e Doerr et al (2003).

6. MODELO FRBR$_{OO}$

Segundo Le Bouef (2009, p. 31), o modelo FRBR$_{ER}$ oferece uma visão completamente estática do próprio universo do discurso. Não sabemos de onde vêm os objetos que o compõem e só o atributo "data" pode deixar supor, de uma maneira totalmente implícita que a entidade começou a existir em um dado momento. No novo modelo, foram definidos eventos temporais em relação à concepção de uma obra e a criação de sua expressão. Dessa forma, passa-se a se ter uma semântica mais precisa para o atributo "data". A seguir, apresentaremos os eventos que tratam do processo criativo de um autor, representados na linha do tempo pela *Figura 5*.

(A) Evento "*F27 Work Conception*" – ocorre no momento em que o autor tem a primeira idéia em relação à criação de uma obra. Esse evento marca o início do processo criativo que será finalizado quando o autor considerar que a expressão de sua obra está completa.

(B) Evento "*F28 Expression Creation*" – marca o momento em que o autor finaliza o registro da expressão de sua obra em um suporte físico, criando, ao mesmo tempo, a expressão (*F2 Expression*) e a manifestação única (*F4 Manifestation Singleton*) de sua obra, conhecido também como original do autor.

Figura 5. Processo de criação da obra.
Fonte: (BEKIARI et al, 2010, adap.)

O processo de edição de uma obra é representado com semelhante precisão, como veremos a seguir. A partir do original do autor, será criada, por um "*F30 Publication Event*", uma "*F24 Publication Expression*" que conterá todas as informações necessárias para a produção dos suportes de informação. No caso de livros, por exemplo, essa expressão de publicação considera todos os elementos pré-textuais (definições de capa, apresentação, sumário, etc.), pós-textuais (índices), além daqueles inseridos no corpo, tais como ilustrações. O evento que faz a incorporação de uma obra em um objeto físico ("*F5 Item*") é o "*F32 Carrier Production*

Event". Os objetos produzidos devem possuir características comuns representando um determinado tipo de manifestação ("*F3 Manifestation Product Type*"). A *Figura 6* apresenta as classes citadas no processo de edição e publicação de uma obra.

Figura 6 – Processo de Edição e Publicação de uma Obra
Fonte: (BEKIARI et al, 2010, adap.)

7. HARMONIZAÇÃO DAS ENTIDADES DO GRUPO 1 DO FRBR$_{ER}$

Segundo Smiraglia (2003), o conceito de Obra evoluiu desde Panizzi, deixando de ser uma entidade secundária na visão dos primeiros catálogos, que eram mais voltados à função de inventário, e passaram a assumir um papel importante nos catálogos modernos, após a constatação de que o usuário de um sistema de recuperação de informação está mais interessado, normalmente, no conteúdo do que em um suporte ou manifestação específica.

Mesmo antes da publicação do modelo FRBR$_{ER}$, algumas pesquisas discutiam a necessidade de criação de uma entidade que agrupasse obras derivadas de outras obras. Por exemplo, Yee (1995), Carlyle (1996) e Svenonius (2000) defenderam a criação da entidade "*Superwork*". Outros pesquisadores, que defendiam, na essência, a mesma idéia, nomearam esta entidade como "*Bibliographic Family*" (SMIRAGLIA & LEAZER, 1999), "*Textual Identity Network*" (LEAZER & FURNER, 1999) e "*Instantiation Network*" (SMIRAGLIA, 2002).

O novo modelo FRBR$_{OO}$ cristalizou os resultados destas pesquisas, criando a classe "*F15 Complex Work*" que permite agrupar obras de acordo com algum critério. Esse agrupamento é realizado pela propriedade "*R10 has member*" que permite definir quais outras obras participam de um grupo "*F15 Complex Work".* Apesar de "*F1 Work*", subclasse de "*E89 Propositional Object*", ser a principal classe referente à Obra é a classe "*F15 Complex Work*", segundo Bekiari et al (2010), que mais se aproxima da definição da entidade Obra do modelo FRBR$_{ER}$.

Para modelar os conceitos associados com um conjunto específico de símbolos (Expressão), foi definida a classe "*F14 Individual Work*". O novo modelo define ainda seis outras classes relacionadas à entidade Obra. Aparentemente mais complexa, a nova modelagem permite representar cada aspecto que antes estava compreendido em uma única entidade, tornando, assim, o modelo mais claro, preciso e fácil de aplicar.

A *Figura 7* mostra o posicionamento da classe "*F1 Work*" na ontologia CIDOC CRM bem como suas subclasses. Notem-se dois casos de herança múltipla: classes "*F17 Aggregation Work*" e "*F18 Serial Work*".

Figura 7. Classe "F1 Work" e suas subclasses.

Enquanto a substância de uma Obra é composta de conceitos, a substância de uma Expressão é composta de símbolos. Quando uma Obra é realizada de forma completa por um conjunto de símbolos, tem-se uma instância da classe "*F22 Self-Contained Expression*". No outro caso, quando este conjunto está incompleto por qualquer motivo, por exemplo por perda de parte do suporte da informação, ocorre uma instância da classe "*F23 Expression Fragment*". Estas classes são subclasses de "*F2 Expression*" conforme pode ser observado na *Figura 8*.

A entidade Manifestação, que apresentava problemas de inconsistência semântica no antigo modelo, passou a ser representada por duas classes distintas: uma, relacionada à visão conceitual (abstrata), denominada "*F3 Manifestation Product Type*", subclasse de "*E55 Type"* e de "*E72 Legal Object"*; e outra, relacionada à visão concreta (física), denominada "*F4 Manifestation Singleton*". A *Figura 9* apresenta o posicionamento das classes citadas no novo modelo.

Figura 8. Classe "F2 Expression" e suas subclasses.

Figura 9. Harmonização da Entidade Manifestação.

8. CONCLUSÃO DO CAPÍTULO 4

A versão orientada a objeto do modelo FRBR permite uma visão unificada dos acervos das instituições de memórias. Uma análise na bibliografia do FRBR (IFLA 2010) mostra que a quantidade dos trabalhos que endereçam a versão orientada a objeto é pequena quanto comparada à quantidade de trabalhos que referenciam a versão original. No caso da língua portuguesa, é ainda raro encontrar referências à nova versão do modelo FRBR. Entre os motivos que explicam essa situação podemos citar: o modelo FRBR$_{OO}$ é estruturado como uma ontologia, formalismo mais complexo e completo que o Modelo Entidade Relacionamento e ainda não tão difundido no universo bibliográfico; os trabalhos de harmonização iniciaram em 2003 e apenas em 2009 foi publicada a minuta da versão 1.0; e, por fim, até o presente momento da escrita deste trabalho a IFLA ainda não havia aprovado formalmente a nova versão. Apresentamos os conceitos básicos de ontologia, CIDOC CRM e FRBR$_{OO}$ no intuito de trazer o tema à discussão no Brasil e demonstrar o potencial de expressividade desta nova abordagem.

9. REFERÊNCIAS DO CAPÍTULO 4

BEKIARI, C.; DOERR, M.; LE BOEUF P.; International Working Group on FRBR and CIDOC CRM Harmonisation. *FRBR Object-Oriented Definition and Mapping to FRBRER*. (*version 1.0.1*) 2010. Disponível em : <http://cidoc.ics.forth.gr/docs/frbr_oo/frbr_docs/FRBRoo_V1.0.1.pdf>. Acesso em: 01 junho 2010.

CARLYLE, A. Ordering author and work records: An evaluation of collocation in online catalog displays. *Journal of the American Society for Information Science.* V. 47. P. 538-54. 1996.

CROFTS, N. *Museum informatics: the challenge of integration*. Genebra, 2004. 264 p. Tese (Doutorado) – Faculté des sciences économiques et socials, Universidade de Genebra, Genebra, 2004

DOERR, Martin. Ontologies for Cultura Heritage. In Staab, S; Studer, R (Eds) *Handbook on Ontologies*. Berlin: Spring-Verlag, 2009.

DOERR, Martin; HUNTER, Jane; LAGOZE, Carl. Towards a core ontology for information integration. In: *Journal of Digital Information* [on line]. V. 4, n. 1. Disponível em: <http://journals.tdl.org/jodi/article/view/92/91>. Acesso em: 01 junho 2010.

FITCH, Kent. *ALEG Data Model. Inventory*. AustLit Gateway. Acesso em: 01 junho 2010. Disponível em: <_TTP://www.austlit.edu.au:7777/DataModel/inventory.html>.

GRUBER, T.R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In Guarino, N.; Poli, R.: (Eds.) *Formal Ontology in Conceptual Analysis and Knowledge Representation*. Padova: LADSEB-CNR, 1993.

GUARINO, N.; GIARETTA, P. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In: Mars, N. (Ed.). *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*. Amsterdam: IOS Press, 1995. P. 25-32.

GUARINO, N. Formal Ontology and Information Systems –FOIS'98. Trento, Itália. Amsterdam: IOS Press, 1998. P. 3-15.

HEANEY, M. *Time is of the essence: some thoughts occasioned by the papers contributed to the International Conference on the Principles and Future Development of AACR.* Oxford: Bodleian Library, 1997. Disponível em: <_TTP://www.bodley.ox.ac.uk/users/mh/time978a.htm>. Acesso em: 01 junho 2010.

ICOM/CIDOC Documentation Standards Group; CIDOC CRM Special Interest Group. *Definition of the CIDOC Conceptual Reference Model: version 4.0*, 2004. Heraklion, Greece, 2004. Disponível em: < _TTP://cidoc.ics.forth.gr/docs/cidoc_crm_version_5.0.2.pdf>. Acesso em 01 junho 2010.

International Federation of Library Associations and Institutions. Study Group on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records*. München: K . G. Saur, 1998. 144 p.

International Federation of Library Associations and Institutions. FRBR Review Group. *FRBR Bibliography*. Disponível em: <_TTP://infoserv.inist.fr/wwsympa.fcgi/d_read/frbr/FRBR_bibliography.rtf>. Acesso em 01 junho 2010.

International Organization for Standardization. ISO 21127-2006: Information and documentation – A reference ontology for the interchange of cultural heritage information. Geneva, Switzerland: International Organization for Standardization, 2006. 108 p.

LEAZER, G.; FURNER, J. Topological indices of textual identity networks. In L. Woods (Ed.), *Proceedings of the 62nd Annual Meeting of the American Society for Information Science* (pp. 345–358). Medford, NJ: Information Today. 1999.

LAGOZE, Carl. Business unusual: how "event-awareness" may breathe life into the catalog?. In: *Conference on bibliographic control in the new millennium*. Washington: Library of Congress, 19 de outubro de 2000. Disponível em: <_TTP://lcweb.loc.gov/catdir/bibcontrol/lagoze_paper.html>. Acesso em 01 junho 2010.

LE BOEUF, PATRICK. Da FRBRer a FRBRoo: Lectio Magistralis in biblioteconomia: Firenze, Universitá degli studi di Firenze, 19 marzo 2009. Fiesole (Firenze): Casalini Libri, 2009. 36 p.

SMIRAGLIA, R. P. The History of "The Work" in the Modern Catalog. *Cataloging & Classification Quarterly*. v. 35, n. 3/4, pp. 553-567, 2003.

SMIRAGLIA, R. P.; LEAZER, G.H. Derivative bibliographic relationships: The work relationship in a global bibliographic database. *Journal of the American Society and Information Science*, v. 50, p. 493–504, 1999.

SMIRAGLIA, R. P. Further reflections on the nature of 'A Work': An introduction. *Cataloging & Classification Quarterly*. v. 33, n. 3/4, p. 1-11, 2002.

SVENONIUS, E. *The Intellectual Foundation of Information Organization*. Boston: MIT Press. 2000. 255 p.

YEE, M. What is a work? Part 3: The Anglo-American cataloging codes. *Cataloging & Classification Quarterly*. v. 20, n. 1. 1995. p. 25-46.

Capítulo 5

Atributos dos Requisitos Funcionais para Registros Bibliográficos (FRBR)

Fernanda Passini Moreno[1]

SUMÁRIO DO CAPÍTULO 5

Resumo do Capítulo 5, *p. 49*
Como citar o Capítulo 5, *p. 94*

1. **INTRODUÇÃO DO CAPÍTULO 5**, *p. 95*
2. **ANTECEDENTES DO MODELO**, *p. 95*
3. **VISÃO GERAL DO MODELO**, *p. 96*
4. **ATRIBUTOS DO MODELO**, *p. 97*
5. **SOBRE OS VALORES DOS ATRIBUTOS EM RELAÇÃO ÀS TAREFAS E ENTIDADES**, *p 98*
   - **5.1** **Atributos da entidade Obra**, *p. 99*
   - **5.2** **Atributos de Expressão**, *p. 106*
   - **5.3** **Atributos de Manifestação**, *p. 109*
6. **CONSIDERAÇÕES FINAIS DO CAPÍTULO 5**, *p. 113*
7. **REFERÊNCIAS DO CAPÍTULO 5**, *p. 114*

[1] Mestre e Doutoranda em Ciência da Informação. Programa de Pós-Graduação em Ciência da Informação (PPGCInf), Faculdade de Ciência da Informação da Universidade de Brasília (UnB/FCI). CV Lattes: http://lattes.cnpq.br/7471118579735377. E-mail: fpassini@gmail.com.

Resumo do Capítulo 5

Este capítulo apresenta os antecedentes dos Requisitos Funcionais para Registros Bibliográficos (FRBR), publicados pela IFLA (*International Federation of Library Association and Institutions*) em 1998, trazendo uma visão geral do modelo. Dedica-se a explorar as relações entre os atributos das entidades (obra, expressão e manifestação), seus valores em relação às tarefas do usuário definidas no modelo (encontrar, identificar, selecionar e obter) e os campos do formato MARC 21 bibliográfico, baseando-se na literatura. Finalizando, comenta a importância deste conhecimento para uma potencial FRBRização dos catálogos, as atualizações necessárias e espera que o conteúdo apresentado venha a contribuir para futuros estudos sobre o modelo.
Palavras-chave **:** FRBR; IFLA; Atributos; Entidades; Formato MARC 21

Abstract

Functional Requirements for Bibliografic Records (FRBR)
This chapter briefly presents the Functional Requirements for Bibliographic Records: final report, published by IFLA (International Federation of Library Association and Institutions) in 1998, bringing a model's overview. It dedicates to exploring the relationships between the entities attributes (work, expression and manifestation), the values of user tasks defined in the model (find, identifying, select, obtain) and the MARC 21 bibliographic format tags, by means of the revision of recent literature. Concluding, it presents final commentaries about the relevance of this knowledge for a potential catalog FRBRization and the necessity of updates as well it hopes its content brings some contributions to future studies about the model.
Keywords **:** FRBR; IFLA; Attributes; Entities; MARC 21 format

Como citar o Capítulo 5

MORENO, F. P. Atributos dos Requisitos Funcionais para Registros Bibliográficos (FRBR). *In:* Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento.* Brasília DF: IBICT, 2010. 335 p. ISBN: 978-85-7013-072-3. Capíitulo 5, p. 93-114. Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

1. INTRODUÇÃO DO CAPÍTULO 5

FRBR é a sigla para Requisitos Funcionais para Registros Bibliográficos, isto é, uma publicação que apresenta os requisitos mínimos que os registros bibliográficos em formato eletrônico devem possuir. Nesse sentido, os requisitos propõem uma reorganização dos registros bibliográficos e uma compreensão diferenciada dos recursos de informação que estão sendo descritos. É baseado no modelo computacional Entidade-Relacionamento e, portanto, descreve os recursos de informação como entidades, dotadas de atributos, que se relacionam com outras entidades.

FRBR também é considerado um modelo teórico ou conceitual e foi desenvolvido após a comunidade internacional de catalogação reconhecer a necessidade do estabelecimento de um acordo sobre as funções primárias do registro bibliográfico, em resposta às necessidades dos usuários e para um melhor compartilhamento de dados bibliográfico em nível internacional. Participaram de seu desenvolvimento a Divisão de Controle Bibliográfico da IFLA (*International Federation of Library Association and Institutions*) e a CDNL (*Conference of Directors of National Libraries*) e, a partir do rascunho do Termo de Referência para um Estudo de Requisitos Funcionais para Registros Bibliográficos, elaborado por Tom Delsey e Henriette Avram, constituíram um grupo de estudos sob a responsabilidade do Comitê Permanente da Seção de Catalogação da IFLA, que contou com a colaboração de voluntários e consultores de várias nacionalidades.

2. ANTECEDENTES DO MODELO

Promovido pela IFLA com auxílio da UNESCO, o Seminário de Estocolmo (Suécia) foi realizado como prévia à Conferência Anual da IFLA de em 1990. Em meados de 1992, o documento de referência para o grupo de estudos indicava que era preciso determinar as funções do registro bibliográfico e quais os requisitos para as funções e os conteúdos do registro; desenvolver uma estrutura que identificasse e definisse entidades de interesse de usuários potenciais dos registros bibliográficos e a relação entre estas entidades e, para cada entidade, identificar e definir as funções que se esperava de um registro bibliográfico. Na 58ª. Conferência Geral da IFLA, em Nova Déli, o Comitê Permanente aprovou o Termo de Referência com duas modificações: a adição de assuntos à lista de entidades e a inclusão de um nível básico de funcionalidade para a criação de registros por agências bibliográficas nacionais.

Nos dois anos seguintes, o Grupo de Estudos apresentou os resultados iniciais do estudo nas Conferências da IFLA, acolhendo sugestões do Comitê Permanente e dos participantes. Na Conferência Geral de 1994, realizada em Havana (Cuba) a apresentação dos estudos contou com painelistas da Espanha, Canadá e Noruega, além da apresentação da Dra. Tillett, intitulada "Fundamentos Teóricos e Práticos", o que demonstra o alcance intercontinental das discussões para elaboração do modelo. Percebemos, no entanto, a ausência de participações nacionais nestas discussões, ao menos de maneira formal nas Conferências anuais, segundo nosso levantamento.

Durante os anos seguintes, o Grupo expandiu o projeto do relatório, incluindo o refinamento e a finalização das tarefas do usuário, a expansão dos relacionamentos associados às quatro primeiras entidades (MADISON, 2005, p. 25) e a elaboração dos atributos, baseados nas ISBDs e Gare (MADISON, 2005, p. 25; IFLA 1998, p.7) e na consulta a voluntários, consultores e especialistas.

Aprovado o texto final pelo Comitê Permanente em 1997, foi denominado *Functional Requirements for Bibliographic Records: final report* e publicado pela K. G. Saur em 1998, bem como disponibilizado on-line no *site* da IFLA (www.ifla.org), que estabeleceu planos estratégicos para divulgação e promoção do modelo. Atualmente, existem 17 traduções oficiais em variadas línguas.

Este capítulo se baseia no texto do modelo FRBR atualizado em fevereiro de 2009, que guarda pouca diferença da atualização de 2008 (IFLA STUDY GROUP..., 2008), que compreendeu, entre outras, alterações nos limites entre as entidades do Grupo 1, deixando-os mais claramente definidos.

3. VISÃO GERAL DO MODELO

O modelo apresenta as 10 entidades divididas em: Grupo 1 (Obra, Expressão, Manifestação e Item), Grupo 2 (Pessoa e Entidade Coletiva), Grupo 3 (Conceito, Objeto, Evento e Lugar), além de um grupo adicional de entidades (agregadas e componentes), estas menos exploradas nos FRBR. Um mesmo recurso de informação pode ser desdobrado nas quatro entidades do Grupo 1 e estas são os maiores objetos de estudos relacionados ao modelo. Assim uma obra só é realizada através da expressão, que está contida em um suporte, uma manifestação, que é exemplificada por um item. Estas são as relações bibliográficas primárias.

A primeira figura representa uma Obra literária.

Logo abaixo temos letras do alfabeto romano, que podem exemplificar a língua original da Obra, caracteres em chinês, que podem exemplificar uma tradução para o Mandarim e a figura que representa um registro sonoro não musical – isto é, os sons da leitura de um livro. Estas letras, sons e caracteres são as maneiras que a Obra pode ser realizada – suas Expressões.

No último agrupamento estão figuras que representam o suporte onde aquelas Expressões da Obra estão contidas: o conteúdo em áudio pode estar em um CD, o disquete pode conter o texto escrito em PDF, por exemplo, e o livro em sua forma tradicional, em papel. Estas são as Manifestações. Um Item é a exemplificação da manifestação sendo, geralmente, o mesmo que a Manifestação em si.

Entre as entidades ocorrem ainda outros relacionamentos: de responsabilidade entre as entidades do primeiro e segundo grupo ou entre entidades do primeiro grupo. Os relacionamentos de responsabilidade indicam que uma obra é criada, uma expressão é realizada, a manifestação é produzida e um item é possuído por uma pessoa ou entidade coletiva. Relacionamentos entre as entidades do primeiro grupo podem ser exemplificados pelas relações entre obras: uma obra pode sofrer adaptação e dar origem a outra obra. Neste caso, a adaptação é um relacionamento. Recomenda-se a leitura de Moreno (2006) e Mey e Silveira (2009) para visualização gráfica e exemplos dos relacionamentos.

O modelo FRBR foi definido em relação às tarefas genéricas realizadas pelos usuários, chamadas *user tasks*, no intuito de atender suas necessidades informacionais. Espera-se que os usuários façam uso dos dados contidos em um registro para encontrar, identificar, selecionar e obter os recursos desejados. De forma mais detalhada, segue a descrição contida no Relatório Final do modelo FRBR. Para outras visões sobre as tarefas do usuário encontradas na literatura da área, consulte Moreno (2009).

Tarefas do usuário – FRBR

Encontrar	Encontrar uma única entidade ou um conjunto de entidades em um arquivo ou base de dados como o resultado de uma busca usando um atributo ou o relacionamento da entidade.
Identificar	Confirmar que a entidade descrita corresponde à entidade procurada, ou para distinguir entre duas ou mais entidades com características similares.
Selecionar	Selecionar uma entidade adequada às necessidades do usuário, isto é, para escolher uma entidade que vá ao encontro das exigências do usuário em relação ao conteúdo, formato físico, etc., ou à rejeição de uma entidade como sendo imprópria às necessidades do usuário.
Obter	Encomendar, adquirir, ou obter acesso à entidade descrita, isto é, para adquirir uma entidade por meio de compra ou empréstimo, etc., ou para acessar eletronicamente uma entidade por meio de uma conexão em linha a um computador remoto.

De acordo com Svenonius (2000, p. 15-16), o primeiro objetivo pode ser entendido como o objetivo de encontrar (*finding objective*), supondo que o usuário tem em mãos um autor, título ou assunto e está procurando um documento. O segundo seria o objetivo de organizar (*collocating objective*), onde se assume que o usuário tem em mãos informações semelhantes, mas necessita de um conjunto de documentos, seja de um dado autor, assunto ou gênero. O terceiro seria o objetivo da escolha (*choice objective*), no qual supostamente o usuário estaria diante de um número similar de documentos e necessitaria fazer uma escolha efetiva entre eles como, por exemplo, entre diversas edições de uma obra.

No modelo FRBR foram atribuídos valores a estas tarefas: alto, médio ou baixo, em relação a cada entidade, atributo e relacionamento entre as entidades. Neste capítulo buscamos detalhar os elementos chave para a compreensão das entidades e, por conseguinte, do papel das tarefas do usuário: os atributos. Por definição, os atributos servem para distinguir as entidades e servem para demonstrar as diferenças de conteúdo (intelectual ou artístico), pois são propriedades ou características próprias das entidades. Este é o tema da próxima seção.

4. ATRIBUTOS DO MODELO

Como um modelo teórico, a análise dos requisitos propostos no modelo FRBR não pode ser implementada em um banco de dados ou catálogo da forma como estão apresentados. Dito de outra, forma, os atributos são apresentados em um nível genérico e não como um modelo passível de implementação. Os atributos abarcam mais que os elementos de descrição em si, pois foram definidos em um nível lógico. Em certos casos, o atributo e o elemento de dado comumente encontrado nas descrições é coincidente, mas na maioria das vezes um atributo do modelo FRBR representa um agre-

gado de elementos de dados individuais. Um exemplo de elemento de dado coincidente é o atributo "identificador da manifestação", que corresponde ao número normalizado internacional para livros – ISBN. No caso do atributo do modelo ser mais amplo, um exemplo é o título da manifestação, que pode ter vários elementos de dados associados (título equivalente, título principal, título normalizado, etc).

Os atributos podem ser inerentes às entidades ou a ela imputados. Os primeiros se referem às características físicas, aspecto formal que caracterizam uma manifestação, ou outros identificados através do exame do item (informações na capa, na página de rosto, etc). Os atributos externos compreendem, por exemplo, os identificadores daquela entidade e informações contextuais, o número no catálogo temático, o número de chamada, ou o contexto em que a obra foi realizada, e que, geralmente, requerem o uso de outras fontes para estabelecê-los.

A seguir são apresentados os atributos por entidade do primeiro grupo: obra, expressão e manifestação. Os atributos descritos são relacionados com o formato MARC, isto é, cada atributo pode estar representado por um ou mais campos e subcampos do formato, mas não apresentamos um trabalho exaustivo neste sentido, já que há casos controversos e desdobramentos de atributos nas fontes utilizadas para esta compilação. A convenção cifrão ($) foi adotada para indicação de subcampos e omitida a definição de campo e subcampo (quando de mesmo conteúdo) se já apresentados em outro atributo.

Os atributos são relacionados com as tarefas do usuário acima descritas, ou seja, um atributo pode ser muito importante, ter uma importância mediana ou ter baixa importância para o usuário desempenhar determinada tarefa, de acordo com o modelo FRBR. Estes valores são comentados logo após as tabelas a partir do mapeamento do modelo FRBR (IFLA STUDY GROUP..., 2008, p.79-92)[2].Ressaltamos que nos restringimos à definição e correlação dos atributos para livros, excluindo portanto os atributos definidos no modelo para mapas cartográficos, recursos contínuos e notações musicais.

5. SOBRE OS VALORES DOS ATRIBUTOS EM RELAÇÃO ÀS TAREFAS E ENTIDADES

A avaliação da importância de cada atributo-entidade-tarefa, segundo declaração do modelo (IFLA STUDY GROUP..., 2008, p. 80) se deu em grande parte pelo conhecimento e experiência do grupo de estudos que o desenvolveu, complementado pela literatura e opinião de especialistas. Os critérios são variáveis dependendo da natureza da tarefa em relação ao atributo.

Para *encontrar* uma entidade, o valor alto foi assinalado quando o atributo, por definição serve como identificador da entidade, ou é usado como principal termo para localizá-la em uma busca. O valor moderado, quando aquele atributo proporciona um meio útil para refinar a busca ou é um termo secundário de busca. O valor baixo, para aqueles que, em circunstâncias mais limitadas, podem ser utilizados para refinar a pesquisa.

Para *identificar* uma entidade, o valor alto foi assinalado quando o atributo por definição serve como identificador da entidade, como em encontrar, ou no caso da ausência de um identificador único, aqueles atributos que sirvam para diferenciar entidades de características comuns. O valor moderado foi assinalado para atributos que sob certas circunstâncias servem para identificar entidades ou o atributo é especialmente associado a um subtipo de entidade (gravações de som, por exemplo). O valor baixo, quando o atributo é usado em casos mais limitados que os anteriores.

Para *selecionar* uma entidade, o valor alto foi assinalado quando o atributo é um indicador significativo do conteúdo intelectual ou artístico, ou o atributo é relacionado a algum requisito técnico específico (o sistema requerido para recursos eletrônicos, etc). O valor moderado quando a representação do conteúdo intelectual ou artístico é usado em alguns casos (como exemplo, o publico a que se destina uma obra) ou para o atributo
.

[2] Delsey (2006), nos documentos da análise funcional supra citada, apresenta novas tarefas do usuário, que não foram considerados neste capítulo.-

que em certas circunstâncias vai necessitar de requisitos técnicos. O valor baixo, para aqueles que são de menor importância ou só marginalmente servem para indicar o conteúdo intelectual ou requisitos técnicos especiais.

Para *obter* uma entidade, os valores foram assinalados quase à mesma maneira da tarefa identificar, e é uma tarefa aplicável apenas às entidades manifestação e item.

5.1 Atributos da entidade Obra

A obra é definida como uma entidade abstrata, só reconhecida através das suas expressões. No entanto, é possível descrever os atributos da entidade obra, como se vê a seguir.

ATRIBUTO	Definição	CAMPO MARC	Definição e subcampos
Título	O título da obra é uma palavra, frase ou grupo de caracteres que nomeia a obra. Inclui todas as variações possíveis sobre o título, e não apenas o título uniforme.	130	Campo para títulos uniformes, utilizado nos casos de obras anônimas, sagradas, tratados identificados por um título coletivo, manuscritos ou grupo de manuscritos, etc., quando constituem uma entrada principal. $a - Título uniforme $n - Número da parte/seção da publicação $p - Nome da parte/seção da publicação
		240	Campo para título uniforme /original, quando há entrada principal assinalada nos campos 100, 110 ou 111, isto é, autoria pessoal, entidade coletiva ou evento. $a - título uniforme $d – data de assinatura do tratado $k – forma do subcabeçalho (subcabeçalho utilizado como título uniforme, tais como manuscritos, protocolos, etc e seleções) $m - meio de execução (para música) $n - $p -
		243	Campos para título uniforme coletivo, geralmente um título construído pelo catalogador (portanto representado entre colchetes), quando as obras de um autor prolífico são reunidas. $a, $d, $m, $n, $p $ r - Escala musical
		245	Campo para título principal e indicação de responsabilidade ($c, não relacionado aqui). $a – título principal

			$g - conjunto de datas (se a obra foi escrita em um determinado período de tempo e este deve ser indicado na descrição) $k - forma do material[3] (Termo que descreve a forma dos materiais descritos, determinados por uma análise do seu aspecto físico, seu conteúdo intelectual ou outras informações sobre eles, por exemplo, agendas, diários, diretórios, revistas, memorandos, etc) $n $p
		700	A faixa de campos 70X a 75X é relativa às entradas secundárias. Este campo é utilizado para nome pessoal como entrada secundária. Segundo a literatura, o segundo indicador deve ter como valor 2, isto é, que haja a indicação de que a entrada secundária é uma analítica (entrada para parte de um item, já registrado sob uma entrada abrangente). O mesmo valor de indicador e subcampos são usados para todos os campos relacionados nesta faixa. $t - Título da publicação (que pode ser o título, o título uniforme ou um título de série utilizado na entrada secundária autor/título). $n – $p –
		710	Utilizado para nome da Entidade como entrada secundária, seguida das unidades subordinadas
		711	Utilizado para nome do evento usado como entrada secundária, bem como o número do evento, o ano e o lugar onde foi realizado.
		730	Utilizado para títulos uniformes como entrada secundária
		740	Títulos relacionados e analíticos não controlados em listas de autoridades ou outro registro bibliográfico. Para títulos relacionados e analíticos controlados em listas de autoridades, usamos o campo 730 (Entrada Secundária - Título Uniforme). O campo pode conter parte do título de uma obra relacionada que foi informada numa secundária autor/título nos campos 700, 710, 711, no subcampo $t. No caso das coleções sem título coletivo, pode-se informar neste campo os outros títulos subsequentes ao primeiro título registrado no campo 245.
Forma	Classe à qual obra pertence (por exemplo, novela, peça teatral, poema, ensaio, biografia, sinfonia, concerto, sonata, mapa, desenho, pintura, fotografia etc.)	130	$k - subcabeçalho utilizado como título uniforme
		240/245	Para estes campos relacionados ao título, há a indicação de:

[3] Não confundir com forma da obra (atributo) e com meio, relativo à designação Geral do Material (DGM). Para estas definições, veja Moreno (2006, p. 135)

			$k
		700/710/711/730	Para estes campos de entradas analíticas, há a indicação de: $k – subcabeçalho
		380	Campo recentemente criado (janeiro de 2010), nomeado como forma da obra, a ser publicado no Update n. 11 do formato MARC.
Data	A data da obra é a data (normalmente o ano) que a obra foi originalmente criada.	130/240/243	Para estes campos relacionados à entrada principal e título, há a indicação de: $d – data de assinatura do tratado
		245	$g - Conjunto de datas (também utilizado para material misto, pertencente a uma coleção)
		700/710/711/730	Para estes campos de entradas analíticas, há a indicação de: $d - data
Outra característica distintiva	Uma característica distintiva é qualquer característica que sirva para diferenciar uma obra de outra obra com o mesmo título.		Para todos os campos relacionados a este atributo, há a indicação de $s – versão (nome, código, ou descrição de uma cópia do material descrito que foi gerado em épocas diferentes ou para diferentes audiências)
		130/240/243	Para estes campos relacionados à entrada principal e título, há a indicação de: $s
		600	Campo que contém um nome pessoal utilizado como assunto atribuído à obra. $s
		610	Campo que contém um nome de entidade que será utilizado como assunto atribuído ao documento. $s
		611	Campo utilizado para nome do evento como assunto atribuído à obra. $s
		630	Campo utilizado para título uniforme como assunto atribuído à obra. $s
		700/710/711/730	Para estes campos de entradas analíticas, há a indicação de: $s
		800	Campo para entrada secundária de série, neste caso, entrada para: Autor/título como entrada secundária de série, isto é, a entrada do autor é um nome pessoal. $s
		810	Campo para entrada secundária de série, neste caso, entrada para: "Autor/título como entrada secundária de série; sendo a entrada do autor o nome da Entidade (associações,

			instituições, firmas comerciais, empresas sem fins lucrativos, governos, órgãos estatais, entidades religiosas, igrejas locais, etc.), seguida das unidades subordinadas." (DIVISÃO DE BIBLIOTECAS..., 2008) $s
		811	Campo para entrada secundária de série, neste caso, entrada para "Autor/título como entrada secundária de série; sendo a entrada do autor o nome do evento - conferências, congressos, seminários, encontros, colóquios, exposições, feita, etc.-, bem como o número do evento, o ano e o lugar onde foi realizado." (DIVISÃO DE BIBLIOTECAS..., 2008) $s
		830	Utilizado para títulos uniformes como entrada secundária de série. $s
Público a que se destina	A audiência pretendida, ou público a que se destina, é a classe de usuários para a qual a obra foi criada, definida por faixa etária, nível educacional ou outra categorização.	006	Campo 006 para livros. Campo de dados codificados que não podem estar no campo 008 (que contém informações codificadas sobre o registro como um todo), sem subcampos e indicadores, contém posições de 00 a 17. Para este atributo, utiliza-se a posição 05 (público-alvo) e recomenda-se observar a posição 22 do campo 008, que identificará a faixa etária a partir do campo 521, descrito abaixo. $s
		521	Campo da área de notas, para indicação de público alvo, identifica o público específico ou nível intelectual para o qual o conteúdo do item descrito é apropriado. Não foi encontrada indicação de subcampo, mas presume-se que seja $a – nota de público-alvo.
Contexto da obra	O contexto pode ser histórico, social, intelectual, artístico ou outro contexto no qual a obra foi originalmente concebida.	-	Não foi encontrada correspondência direta
ATRIBUTO	Definição	CAMPO MARC	Definição e subcampos
Título	O título da obra é uma palavra, frase ou grupo de caracteres que nomeia a obra.	130	Campo para títulos uniformes, utilizado nos casos de obras anônimas, sagradas, tratados identificados por um título coletivo, manuscritos ou grupo de manuscritos, etc., quando constituem uma entrada principal. $a - Título uniforme $n - Número da parte/seção da publicação $p - Nome da parte/seção da publicação

	Inclui todas as variações possíveis sobre o título, e não apenas o título uniforme.	240	Campo para título uniforme /original, quando há entrada principal assinalada nos campos 100, 110 ou 11, isto é, autoria pessoal, entidade coletiva ou evento. $a - título uniforme $d – data de assinatura do tratado $k – forma do subcabeçalho (subcabeçalho utilizado como título uniforme, tais como manuscritos, protocolos, etc e seleções) $m - meio de execução (para música) $n - $p -
		243	Campos para título uniforme coletivo, geralmente um título construído pelo catalogador (portanto representado entre colchetes), quando as obras de um autor prolífico são reunidas. $a, $d, $m, $n, $p $ r - Escala musical
		245	Campo para título principal e indicação de responsabilidade ($c, não relacionado aqui). $a – título principal $g - conjunto de datas (se a obra foi escrita em um determinado período de tempo e este deve ser indicado na descrição) $k - forma do material[4] (Termo que descreve a forma dos materiais descritos, determinados por uma análise do seu aspecto físico, seu conteúdo intelectual ou outras informações sobre eles, por exemplo, agendas, diários, diretórios, revistas, memorandos, etc) $n $p
		700	A faixa de campos 70X a 75X é relativa às entradas secundárias. Este campo é utilizado para nome pessoal como entrada secundária. Segundo a literatura, o segundo indicador deve ter como valor 2, isto é, que haja a indicação de que a entrada secundária é uma analítica 9entrada para parte de um item, já registrado sob uma entrada abrangente). O mesmo valor de indicador e subcampos são usados para todos os campos relacionados nesta faixa. $t - Título da publicação (que pode ser o título, o título uniforme ou um título de série utilizado na entrada secundária autor/título). $n – $p –
		710	Utilizado para nome da Entidade como entrada secundária, seguida das unidades subordinadas
		711	Utilizado para nome do evento usado como entrada secundária, bem como o número do evento, o ano e o lugar onde foi realizado.
		730	Utilizado para títulos uniformes como entrada secundária
		740	Títulos relacionados e analíticos não controlados em listas de autoridades ou outro registro bibliográfico. Para títulos relacionados e analíticos controlados em listas de autoridades, usamos o campo 730 (Entrada Secundária - Título Uniforme). O campo pode conter parte do título de uma obra relacionada que foi

[4] Não confundir com forma da obra (atributo) e com meio, relativo à designação Geral do Material (DGM). Para estas definições, veja Moreno (2006, p. 135)

			informada numa secundária autor/título nos campos 700, 710, 711, no subcampo $t. No caso das coleções sem título coletivo, pode-se informar neste campo os outros títulos subsequentes ao primeiro título registrado no campo 245.
Forma	Classe à qual *obra* pertence (por exemplo, novela, peça teatral, poema, ensaio, biografia, sinfonia, concerto, sonata, mapa, desenho, pintura, fotografia etc.)	130	$k - subcabeçalho utilizado como título uniforme
		240/245	Para estes campos relacionados ao titulo, há a indicação de: $k
		700/710/711/ 730	Para estes campos de entradas analíticas, há a indicação de: $k – subcabeçalho
		380	*Campo recentemente criado (janeiro de 2010), nomeado como forma da obra, a ser publicado no Update n. 11 do formato MARC.*
Data	A data da obra é a data (normalmente o ano) que a obra foi originalmente criada.	130/240/243	Para estes campos relacionados à entrada principal e titulo, há a indicação de: $d – data de assinatura do tratado
		245	$g - Conjunto de datas (também utilizado para material misto, pertencente a uma coleção)
		700/710/711/ 730	Para estes campos de entradas analíticas, há a indicação de: $d - data
Outra característica distintiva	Uma característica distintiva é qualquer característica que sirva para diferenciar uma obra de outra obra com o mesmo título.		Para todos os campos relacionados a este atributo, há a indicação de $s – versão (nome, código, ou descrição de uma cópia do material descrito que foi gerado em épocas diferentes ou para diferentes audiências)
		130/240/243	Para estes campos relacionados à entrada principal e titulo, há a indicação de: $s
		600	Campo que contém um nome pessoal utilizado como assunto atribuído à obra. $s
		610	Campo que contém um nome de entidade que será utilizado como assunto atribuído ao documento. $s
		611	Campo utilizado para nome do evento como assunto atribuído à obra. $s
		630	Campo utilizado para título uniforme como assunto atribuído à obra. $s
		700/710/711/ 730	Para estes campos de entradas analíticas, há a indicação de: $s
		800	Campo para entrada secundária de série, neste caso, entrada para: Autor/título como entrada secundária de série, isto é, a entrada do autor é um nome pessoal. $s
		810	Campo para entrada secundária de série, neste caso, entrada para: "Autor/título como entrada secundária de série; sendo a entrada do autor o nome da Entidade (associações, instituições, firmas comerciais, empresas sem fins lucrativos, governos, órgãos estatais, entidades religiosas,

			igrejas locais, etc.), seguida das unidades subordinadas.” (DIVISÃO DE BIBLIOTECAS..., 2008) $s
		811	Campo para entrada secundária de série, neste caso, entrada para “Autor/título como entrada secundária de série; sendo a entrada do autor o nome do evento - conferências, congressos, seminários, encontros, colóquios, exposições, feita, etc.-, bem como o número do evento, o ano e o lugar onde foi realizado.” (DIVISÃO DE BIBLIOTECAS..., 2008) $s
		830	Utilizado para títulos uniformes como entrada secundária de série. $s
Público a que se destina	A audiência pretendida, ou público a que se destina, é a classe de usuários para a qual a obra foi criada, definida por faixa etária, nível educacional ou outra categorização.	006	Campo 006 para livros. Campo de dados codificados que não podem estar no campo 008 (que contém informações codificadas sobre o registro como um todo), sem subcampos e indicadores, contém posições de 00 a 17. Para este atributo, utiliza-se a posição 05 (público-alvo) e recomenda-se observar a posição 22 do campo 008, que identificará a faixa etária a partir do campo 521, descrito abaixo. $s
		521	Campo da área de notas, para indicação de público alvo, identifica o público específico ou nível intelectual para o qual o conteúdo do item descrito é apropriado. Não foi encontrada indicação de subcampo, mas presume-se que seja $a – nota de público-alvo.
Contexto da obra	O contexto pode ser histórico, social, intelec-tual, artístico ou outro contexto no qual a obra foi originalmente conce-bida.	**-**	Não foi encontrada correspondência direta

O título é um atributo de alto valor para o usuário encontrar, identificar e selecionar a obra, posto que grande parte das pesquisas em catálogos eletrônicos é feita para itens conhecidos, geralmente por autor e título.

Observando o quadro acima, percebe-se que o responsável pela obra não figura como atributo. O(s) responsável (eis) estará (ão) ligado(s) à obra por meio do relacionamento de responsabilidade, como afirmado anteriormente. Assim mesmo, é possível relacionarmos os campos MARC associados ao(s) responsável pelo conteúdo intelectual da obra, quais sejam, os campos de entrada principal: 100 – nome pessoal, ou 110 – entidade, ou 111 – evento, o mesmo válido para entradas analíticas relacionadas à responsabilidade (700, 710, 711).

A forma da obra é mais importante para o usuário selecionar a entidade procurada, mas também é relevante para identificar e encontrar. A data da obra e as outras características distintivas tem baixo valor para o usuário encontrar, identificar, selecionar e obter. Como esperado, o público a que se destina é mais importante para selecionar uma obra do que para identificá-la ou encontrá-la, que tem valor moderado. Já o contexto que a obra foi produzida só tem valor moderado assinalado para selecionar uma obra e o modelo não atribui valores para as outras tarefas.

É importante lembrar que, no contexto da entidade obra, a tarefa obter não é considerada ou correlacionada com nenhum valor, posto a obra ser uma entidade abstrata, e só pode-se obter o item, que é (geralmente) o mesmo que a manifestação que contém a expressão da obra. A este tipo de relacionamento dá-se o nome de relações bibliográficas primárias, ou, dito de outra forma: uma obra é realizada pela expressão, contida na manifestação, exemplificada pelo item.

5.2 Atributos de Expressão

A maneira mais comum de visualizar as expressões das obras são as traduções para outras línguas. Os campos MARC oferecem uma variedade de informação codificada sobre a língua dos documentos que estão sendo descritos, mas este não é seu único atributo, como pode ser visto a seguir.

ATRIBUTO	Definição	CAMPO MARC	Definição e subcampos
Título	O título de uma expressão que faça parte de uma expressão maior pode consistir apenas de um número, ou outra designação genérica, dependente do título da expressão maior.	-	Sem correspondência exata na literatura relacionada. A ferramenta conversora FRBR *Display Tool* seleciona os campos de título (245, 243, 240) para associar a obra à expressão da obra, no caso de expressões que são traduções[5].
Forma	O atributo forma da expressão, intimamente ligado à obra que ela expressa, é entendido como "o meio pelo qual a obra se realiza (por exemplo, através de notação alfanumérica, notação musical, palavra falada, som musical, imagem cartográfica, imagem fotográfica, escultura, dança, mímica etc.)".	Posição 06 do líder	Contém informações sobre o tipo de registro, por exemplo"a", material textual, impresso – usado para livros ou recursos contínuos.
		006	Campo para dados codificados que não podem estar no campo 008. Na posição 06 está a indicação da forma do documento.
		008	Campo que contém informações codificadas sobre o registro como um todo, assim como aspectos bibliográficos do documento que está sendo catalogado. A posição 23 deste campo indica a forma do documento.
		336	Campo recentemente criado (outubro de 2009), já aprovado no *Update* n. 10 do formato MARC. Designa a forma de comunicação através da qual a obra é expressa.
Data	A data da expressão é a data que a expressão foi criada (por exemplo, a data em que um texto em particular foi escrito ou revisto).	-	Sem correspondência, mas pode ser usada a data da obra (copyright) ou a data da manifestação, se aplicáveis.
Língua	"A língua da expressão é a língua em que a obra se expressa. A língua da expressão pode compreender mais de uma língua, cada uma pertencendo a um componente individual da expressão."	008 –	Nas posições 35-37, contém 3 caracteres que indicam o idioma do texto
		041	Campo para língua do documento $a - código da língua do texto $h - código da língua original e/ou texto intermediário

[5] Para discussões a respeito da expressão, veja Moreno (2006, p. 119-127)

		130/240/243 -	Para estes campos relacionados à entrada principal e título, há a indicação de: $l – idioma da obra.
		546	Campo para nota de idioma através de informação textual 9 e não codificada como nos campo 008 e 006, por exemplo) $a – nota de idioma
Outra característica distintiva	Qualquer característica que sirva para diferenciar a expressão de outra expressão da mesma obra, por exemplo, os nomes usados para diferenciar as várias versões do texto em inglês da Bíblia, ou uma designação de "edição" ou versão relativa ao conteúdo intelectual da expressão, como "2ª revisão".	130 /240/243	Para estes campos relacionados à entrada principal e título, há a indicação de: $ o -
		245	$s - versão
		562	Campo para nota de identificação de cópia ou versão
		381	*Campo recentemente criado (janeiro de 2010), nomeado como outras características distintivas da expressão, a ser publicado no Update n. 11 do formato MARC.*
Capacidade de revisão da expressão	Esta capacidade reflete a expectativa de que o conteúdo intelectual ou artístico da expressão seja revisto (por exemplo, expressões que são versões preliminares de um relatório, um diretório que seja atualizado periodicamente, etc).	**-**	Não foi encontrada correspondência direta.
Extensão da expressão	Relacionado a quantificação do conteúdo intelectual ou artístico da expressão, como, por exemplo, número de palavras de um texto, dados em um programa de computador, imagens em uma história em quadrinhos, etc. Mais aplicável a obras expressas em som ou movimento.	006	Campo 006 para material visual, posições 01-03 - tempo de execução
		306 a	Campo para tempo de duração $a- Tempo de duração (pode ser usado para música impressa ou manuscrita, assim como para gravações de som)
		256 a	Campo para características do arquivo de computador $a- Características do arquivo de computador (por exemplo, número de registros)
Sumarização do conteúdo	A sumarização do conteúdo de uma expressão é um resumo, uma sumarização, uma sinopse, ou a lista de capítulos, músicas, partes, etc. incluídas na expressão.	505	Campo para nota de conteúdo $g Informações adicionais por exemplo, volume, paginação, etc).
		520	Campo para nota de resumo, informação não padronizada $a – nota de resumo
Contexto para a *expressão*	O contexto pode ser histórico, social, intelectual, artístico ou outros contexto no qual a expressão foi realizada (por exemplo) período Art Deco, etc.)	**-**	Não foi encontrada correspondência direta
Resposta crítica à *expressão*	É a recepção dada à expressão por críticos, revisores, etc.	586	Campo para nota de premiação, por exemplo informações sobre prêmios recebidos. $a^nota de premiação
Restrições de uso da expressão	São restrições de acesso e uso da expressão, por exemplo, as baseadas no copyright.	**-**	Não foi encontrada correspondência direta. A correlação deste atributo é feita na entidade manifestação.

O título da expressão é relativamente importante para as tarefas de encontrar, identificar e selecionar. Já a forma da expressão, é um atributo de alto valor para o usuário identificar e selecionar esta entidade, enquanto a data tem valor moderado para as mesmas tarefas. A língua da expressão tem valor moderado para encontrar, mas alto para identificar e selecionar, desde que o conteúdo lingüístico da expressão seja significante. Em outras palavras, os usuários devem conseguir reconhecer ou confirmar que aquela é a entidade procurada, ou que por meio destes elementos, consiga distinguir entre duas entidades com características similares.

Outras características distintivas, como atributo, tem valor moderado para encontrar a expressão, e alto para sua identificação e seleção. Tanto a capacidade de revisão da expressão quanto a sumarização de conteúdo tem valores baixos para identificar a expressão e moderado para selecioná-la. O atributo extensão da expressão tem valores baixos para identificar e selecionar, mas serão importantes se a descrição for de recursos de informação como áudio e vídeo, isto é, quando expressam duração. Já os atributos resposta crítica e contexto são assinalados apenas como de valor baixo para selecionar. As restrições de uso da expressão, indicadas nos FRBR como de alto valor para selecioná-la e que podem ser relativas ao copyright, por exemplo, não encontram correspondência no formato MARC, isto é, só são encontradas restrições de *acesso*, relativos a manifestação e ao item que contém a expressão. Em termos abstratos, a expressão é uma entidade diferente da manifestação, já que é a realização intelectual da obra (em palavras, letras, símbolos) mas, por outro lado, é impossível seu uso sem o acesso (ou obtenção, ou aquisição, etc.) da manifestação/item, que são entidades concretas.

5.3 Atributos de Manifestação

A manifestação é o mais próximo dos registros bibliográficos que temos hoje. Sua descrição, como se vê abaixo, inclui as tradicionais áreas de distribuição, publicação, etc., pois é uma entidade relacionada ao conteúdo físico, e não ao intelectual como as anteriores. É interessante notar que alguns atributos aparecem assinalados com valor alto para todas as entidades, mesmo em se tratando estritamente da manifestação, como se vê no extrato retirado do modelo (IFLA STUDY GROUP..., 2008, p. 89):

Table 6.3 Attributes and Relationships of Manifestations

	Find				Identity				Select				Obtain			
	Work	Expression	Manifestation	Item	Work	Expression	Manifestation	Item	Work	Expression	Manifestation	Item	Work	Expression	Manifestation	Item
Attributes of a Manifestation																
Title of the manifestation	■	■	■		■	■	■		□	□	□				■	
Statement of responsibility	□	□	□		□	■	■		□	■	■				■	

Mas se o quadro traz apenas os atributos e relacionamentos da manifestação, porque assinalar as outras entidades? A explicação razoável consiste em que esta entidade contém as outras e, de toda forma, pode conter mais de uma de cada tipo. Tome-se como exemplo uma coletânea de diversos autores, isto é, diversas obras, títulos e indicações de responsabilidade (em um conceito amplo – não aquele preconizado pelas regras catalográficas como a "regra de três"). Esta coletânea pode estar em mais de uma língua ou ainda ser uma obra de um autor, com o texto original e o traduzido (expressão), ou seja, uma obra,suas expressões e, portanto, títulos diferentes da obra para as expressões e, dependendo do editor, um terceiro título – o da manifestação (presente na capa da coletânea).

A seguir, da mesma maneira das entidades anteriores, serão expostos os atributos, sua definição, os campos e subcampos MARC e, abaixo a importância de cada um deles é comentada.

ATRIBUTO	Definição	CAMPO MARC	Definição e subcampos
Título	Inclui um agregado de títulos possíveis, como: o título na página de rosto; no quadro; de capa; da página de rosto secundária; da falsa página de rosto; título corrente; da lombada; título explícito; do invólucro; no cabeçalho de microficha.	245	$a $n $p
		242	Campo para título traduzido pela agência catalogadora $a $n $p
Indicação de responsabilidade	Inclui os diversos tipos de responsabilidade concernentes à criação (autor, compositor, etc.), mas não somente à relativa ao conceito de autoria, incluindo também a responsabilidade indireta (o autor de uma novela na qual se baseia o roteiro de um filme). Podem ser incluídos, também os responsáveis pela expressão da obra contida na manifestação (tradutores, intérpretes etc.), ou aqueles responsáveis pela compilação das obras contidas na manifestação (editor, compilador etc.).	245	$c – Indicação de responsabilidade
		242	$c - Indicação de responsabilidade
Designação de edição/ impressão	Uma palavra ou frase que aparece na manifestação e indica normalmente uma diferença qualquer no conteúdo ou na forma entre uma manifestação e a manifestação a ela relacionada previamente editada pelo mesmo publicador/distribuidor (por exemplo, segunda edição, versão 2.0, etc.), ou editada simultaneamente pelo mesmo publicador/distribuidor ou por outro publicador/distribuidor (por exemplo, edição de massa, edição britânica, etc.).	250	Campo para indicação da edição $a - edição $b – outras informações sobre a edição

ATRIBUTOS DOS REQUISITOS FUNCIONAIS PARA REGISTROS BIBLIOGRÁFICOS (FRBR)

Lugar de publicação/ distribuição	O lugar de publicação ou distribuição de uma manifestação é a cidade ou outra localidade associada ao editor ou distribuidor da manifestação.	260	Campo para publicação, impressão, distribuição, tiragem ou produção de uma obra, anteriormente chamado de imprenta. $a - Lugar de publicação, distribuição, etc
Publicador/ distribuidor	O publicador ou distribuidor de uma manifestação é o responsável pela publicação, edição, distribuição ou lançamento da manifestação, nela indicado, podendo ser uma ou mais pessoas, grupos ou uma organizações.	260	$b - nome do editor, distribuidor, etc.
Data de publicação/ distribuição	A data de publicação da manifestação é normalmente um ano, relacionado ao lançamento público da manifestação.	260	$c – data de publicação $g data de impressão (quando esta aparece no lugar da data de publicação)
Fabricante	Uma manifestação pode ter um ou mais fabricantes a ela associados, seja uma ou mais pessoas, grupos ou organizações.	260	$f - Nome do impressor
Indicação de série	A indicação de série é uma frase, palavra ou grupo de caracteres que aparece na manifestação nomeando a série que a manifestação pertence e pode incluir um número que designa uma posição seqüencial daquela manifestação dentro da série, assim como subsérie.	490	Atual campo para série $a – título da série $v – designação numérica da série
Forma do suporte	A forma do suporte é a classe específica do material a que o suporte físico da manifestação pertence (por exemplo, fita cassete, videodisco, cartucho do microfilme, transparência, etc.)	240/242/243245/ 246/247	Para estes campos relacionados aos títulos, há indicação de: $ h -meio
		533	Campo para nota de reproduções, $e – descrição física da reprodução
		610/611/630	Para estes campos relacionados aos assuntos, há indicação de: $ h –meio
		700/710/ 711/730/740	Para estes campos relacionados às entradas de ligação, há indicação de: $ h –meio
		800/810/ 811/830	Para estes campos relacionados às entradas secundárias, há indicação de: $ h -meio
Extensão do suporte	A extensão é a quantificação do número das unidades físicas que fazem o suporte (por exemplo, número das folhas, dos discos, dos rolos, etc.).	300	Campo para descrição física $a - extensão
		533	$e
Meio físico	É o tipo de material de que o suporte é produzido (por exemplo, papel, madeira, plástico, metal, etc.).	007	Variados usos deste campo, principalmente para materiais que não o livro.
		300	$b – detalhes físicos adicionais

		340	Fornece informação textual referente a características físicas dos materiais. $a – base e configuração do material $ c – materiais aplicados à superfície $e – suporte
Dimensões do suporte	As dimensões do suporte são relativas às medidas dos componentes físicos, que podem compreender medidas de altura e largura, por exemplo.	300	$c - dimensões
		533	$e
		340	$b - dimensões $c
		007	Variados usos deste campo, principalmente para materiais que não o livro.
Identificador	O identificador pode ser atribuído como parte de uma numeração internacional ou um sistema de código, como parte de um sistema nacional, ou podem ser atribuídos independentemente pelo publicador ou pelo distribuidor da manifestação, como por exemplo, número da publicação do governo.	020	Campo para International Standard Book Number (ISBN) $a – número do ISBN $z- ISBN cancelado ou inválido
		027	Campo para número do relatório técnico (International Standard Technical Report) $a - Standard Technical Report Number
Fonte para aquisição/ autorização de acesso	São as fontes indicadas na manifestação pelas quais ela pode ser adquirida ou acessada, por exemplo, o endereço do publicador.	020	$c – modalidades de aquisição
		270	$a - endereço $b - cidade $c - estado ou província $d - país $e - código postal $m - endereço eletrônico $z - nota pública
		530	Campo de nota de outros formatos disponíveis $b – fonte
		856	Campo que indica localização e acesso eletrônico $a – nome do servidor $b – número de acesso $n – nome da localização do servidor mencionado em $a
Termos de disponibilidade	São os termos indicados na manifestação sobre a fonte de acesso ou autorização que normalmente faz a manifestação disponível (por exemplo, acesso livre aos membros de uma associação em particular) ou o preço pelo qual a manifestação é vendida.	020	$c
		024	Campo para outros números ou códigos normalizados $c
		530	$c - condições (de acesso, por exemplo, através de senha)

ATRIBUTOS DOS REQUISITOS FUNCIONAIS PARA REGISTROS BIBLIOGRÁFICOS (FRBR)

Restrições de acesso à manifestação	São restrições de acesso e uso da manifestação, por exemplo, as baseadas no copyright.	355	Campo para controle de classificação de segurança, que pode conter restrições de segurança e acesso externo. $a - classificação de segurança $b - instruções de manuseio $c - informação de disseminação externa $e - sistema de classificação $f - código do país de origem $g - data da mudança de categoria $h - data da liberação de acesso $j - autorização
		357	Campo que indica o controle do autor sobre a publicação $a - Termos de controle do autor $b - Instituição criadora $c - Usuários autorizados do material
		506	Campo para nota de restrição de acesso $a - condições do acesso $b - jurisdição $c – critérios para acesso $d –indicação de usuários autorizados $e - autorização
		540	Campo para nota de condições de uso e reprodução, por exemplo, reativo a copyright $a - Termos que gerenciam o uso e a reprodução $b - $c - $d -
		856	$2 – modo de acesso

O título da manifestação é muito importante para as tarefas de encontrar, identificar e obter e de importância moderada para selecionar aquela manifestação.

A identificação de responsabilidade, que aparece aqui como atributo próprio da manifestação, tem valor moderado para encontrar a manifestação, mas alto para as outras tarefas, assim como a responsabilidade pela expressão aparece como importante para identificar e selecionar a expressão contida na manifestação.

A designação de edição, como atributo, [e importante para identificar e selecionar a expressão e a manifestação que a contém, assim como para obter a manifestação. O lugar de publicação é importante para obtenção do recurso de informação, e de importância moderada para outras tarefas. O distribuidor ou publicador tem valor alto para identificação e obtenção da manifestação, moderado para seleção. o que certamente coaduna com os critérios de desenvolvimento de coleções, ou acervamento.

O fabricante tem valor baixo para encontrar, identificar e selecionar uma manifestação e nenhum valor relacionado à obtenção, mas pode ter valor alto nos casos das primeiras edições de uma obra, ou produções gráficas limitadas. Já a indicação de série tem valor moderado para encontrar e selecionar a manifestação, mas alto para sua identificação e obtenção.

A forma do suporte não é imprescindível para identificação, tanto que sequer é assinalado algum valor a respeito, mas é importantíssimo para as demais tarefas e acreditamos que particularmente para selecionar a manifestação, isto é: o suporte pode vir a influenciar o tipo de acesso ao recurso, já que pode ser dependente de terminado equipamento tecnológico ou configurações de software, etc. O atributo extensão suporte tem valor moderado para identificação e seleção da manifestação, mas pode ter um alto valor nos casos em que assinalam diferenças entre manifestações. Já o meio físico tem valor moderado para a primeira e baixo para a segunda, mas pode ter valor alto nos casos em que esta informação é importante para o usuário, como filmes com base em nitrato. As dimensões do suporte são assinaladas com valor baixo para identificação, mas moderado para seleção e obtenção.

Um identificador da manifestação é muito importante para encontrar, identificar e obter um recurso de informação, mas não para selecioná-lo em relação a outros. Destaca-se que hoje, dez anos depois da publicação do modelo, este argumento ganha força com os diversos identificadores digitais. Fonte de aquisição, tem baixo valor para seleção e aquisição, mas pode ter um alto valor para aquelas manifestações que não são fáceis de obter pelos caminhos normais de aquisição.O atributo termos de disponibilidade tem baixo valor para selecionar e obter e as restrições de acesso um alto valor para obtenção do recurso, moderado para sua seleção

6. CONSIDERAÇÕES FINAIS DO CAPÍTULO 5

Neste capítulo, apresentamos o histórico do modelo FRBR, exemplificamos suas principais entidades e demonstramos a relação dos atributos com as tarefas do usuário e os campos do formato MARC.

Conhecendo estas conexões, compreende-se melhor os conceitos das entidades, pois é através de seus atributos que percebemos as diferenças entre elas. Entendendo as tarefas dos usuários, isto é, o que se espera da relação do usuário com os dados do catálogo, temos condições de organizar melhor a coleção representada, seja ela física ou virtual. A tão conhecida catalogação, também nomeada como representação descritiva, é essencialmente isso: representar um recurso de informação, organizá-lo e relacioná-lo aos demais. Conhecer os atributos proporciona riqueza na descrição, atenção aos detalhes importantes, independente de regras ou normas. Esta é a função e a inovação do modelo conceitual FRBR: reconhecer a complexidade intrínseca à descrição, tomando por base não apenas o item, mas o conteúdo intelectual e as características físicas que podem ser importantes (ou não) para os usuários.

Para levar a cabo uma exibição baseada nestes conceitos e nesta nova forma de organização dos registros é imprescindível o conhecimento dos campos MARC relacionados. É possível contar com uma ferramenta que exibe os registros pesquisados em uma interface baseada no modelo, FRBR *Display Tool*, já anteriormente explorada em pesquisa, aplicada a um conjunto de registros bibliográficos previamente selecionados (MORENO, 2006). Na ocasião, reconhecemos as limitações da ferramenta, mas também a liberdade de edição e manipulação da mesma para melhor atingir os objetivos de "FRBR-ização" dos catálogos.

Destacamos que em decorrência da publicação prevista do novo Código de Catalogação RDA – *Resource Description and Access*, alguns campos MARC (Bibliográfico e de Autoridade) estão sendo modificados e/ou criados para acomodar atributos de obras, expressões e relacionamentos entre os recursos. Como nem todas as atualizações foram publicadas (isto é, aprovadas), deixamos aqui a sugestão de que as mudanças sejam incorporadas a

partir deste estudo inicial. Ademais, como não há compatibilidade completa entre o modelo e o formato, sempre há a possibilidade de inclusão ou modificação da compilação aqui apresentada.

Finalmente, esperamos que a visão fornecida neste capítulo venha a contribuir para futuros estudos sobre o modelo, tão reconhecidamente teórico e carente de implementações, mesmo passado tanto tempo de seu lançamento.

7. REFERÊNCIAS DO CAPÏTULO 5

DIVISÃO DE BIBLIOTECAS E DOCUMENTAÇÃO DA PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO. *MARC 21: formato bibliográfico*. Disponível em: http://www.dbd.puc-rio.br/MARC21/index.html. Acesso em: 4 de jul. 2009.

JOINT STEERING COMMITEE FOR REVISION OF AACR; FEDERAÇÃO BRASILEIRA DE ASSOCIAÇÕES DE BIBLIOTECÁRIOS, CIENTISTAS DA INFORMAÇÃO E INSTITUIÇÕES. *Código de Catalogação Anglo-Americano*. 2. ed. São Paulo: Febab, 2004. 1 v.

LIBRARY OF CONGRESS (LC). *Displays for Multiple Versions from MARC 21 and FRBR* (2003). Prepared by Tom Delsey. Disponível em: http://www.loc.gov/marc/marc-functional-analysis/multipleversions.html.

DELSEY, T. *Functional analysis of the MARC 21 bibliographic and holdings formats.* 2nd. rev., 2003 (update 2006). Prepared for the Network Development and MARC Standards Office, Library of Congress. Disponível em: <www.loc.gov/marc/marc-functional-analysis/functional-analysis.html>. Acesso em: 11 dez. 2009.

IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS. *Functional requirements for bibliographic records: final report.* UBCIM Publications - New Series, vol. 19. München: K. G. Saur, 1998.

IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS. *Functional requirements for bibliographic records: final report.* Current text, feb. 2009. Including amendments and corrections to date. Disponível em: <http://www.ifla.org/VII/s13/frbr/frbr_2008.pdf >. Acesso em: 10 mar. 2008.

LIBRARY OF CONGRESS (LC). *MARC Format overview*. Disponível em: http://www.loc.gov/marc/status.html. Acesso em: 08 de fev. 2010.

LIBRARY OF CONGRESS (LC). MARC 21 *Format for Bibliographic Data*. http://www.loc.gov/marc/bibliographic/ecbdhome.html. Acesso em: 19 de set. 2009.

MEY, E. S. A. ; SILVEIRA, N. C. *Catalogação plural*. Brasília: Briquet de Lemos/Livros, 2009.

MORENO, F. P. *Requisitos funcionais para registro bibliográficos - FRBR: um estudo no catálogo da Rede Bibliodata*. Brasília, 2006. 199 f. Dissertação (mestrado) - Universidade de Brasília. Disponível no E-LIS.

MORENO, F. P. O modelo conceitual FRBR: discussões recentes e um olhar sobre as tarefas do usuário. *Enc. Bibli: R. Eletr. Bibliotecon*. Ci. Inf., ISSN 1518-2924, Florianópolis, v. 14, n.27, 2009. Disponível em: http://www.periodicos.ufsc.br/index.php/eb/issue/view/286. Acesso em: 04 ago. 2009.

Capítulo 6

Organização da informação: Conceitos básicos e breve fundamentação teórica

Lígia Maria Arruda Café[1] **e Rodrigo de Sales**[2]

SUMÁRIO DO CAPÍTULO 6

Resumo do Capítulo 6, *p.116*
Como citar o Capítulo 6, *p.116*

1. PORQUE ORGANIZAR?, *p.117*
2. O CONCEITO DE ORGANIZAÇÃO DA INFORMAÇÃO, *p.118*
3. BREVE FUNDAMENTAÇÃO TEÓRICA PARA ORGANIZAÇÃO DA INFORMAÇÃO, *p.119*
3.1 A catalogação de assuntos de Cutter, *p.121*
3.2 A indexação de Kaiser, *p.122*
3.3 *A classificação de Ranganathan*, *p.123*
3.3.1 O Classification Research Group (CRG), *p.125*
3.3.2 A International Society for Knowledge Organization (ISKO), *p.125*
3.4 Análise documental, *p.126*
4. CONCLUSÃO DO CAPÍTULO 6, *p.126*
5. REFERÊNCIAS DO CAPÍTULO 6, *p.128*

[1] Doutora em Lingüística pela Université Laval (Canadá). Mestre em Ciência da Informação e Documentação pela Universidade de Brasília (UnB). Professora do Departamento de Ciência da Informação da Universidade Federal de Santa Catarina (UFSC). CV Lattes: http://lattes.cnpq.br/9103205400963118. E-mail: ligia@cin.ufsc.br.
[2] Mestre em Ciência da Informação pela Universidade Federal de Santa Catarina (UFSC). Doutorando em Ciência da Informação na Universidade Estadual Paulista (UNESP) – Campus de Marília. Bolsista da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). CV Lattes: http://lattes.cnpq.br/0144377022254487.

Resumo do Capítulo 6 **:**

A organização da informação é abordada na ciência da informação sob dois aspectos, enquanto espaço investigativo que fornece os pressupostos teóricos e metodológicos ao tratamento da informação, e enquanto atividade operacional inerente ao fazer profissional relativo ao tratamento da informação. Os saberes da organização da informação no espaço investigativo alicerçam os fazeres dessa organização como atividade aplicada. Empenhados em explicitar conceitualmente a organização da informação no âmbito da ciência da informação, apresentamos conceitos e abordagens teóricas que compõem a epistemologia da organização da informação, sobretudo no que diz respeito à organização temática. Constatamos que os fundamentos da catalogação de assuntos, da classificação, da indexação e da análise documental são abordagens do tratamento temático da informação responsáveis pela formação da organização da informação como a conhecemos hoje.
Palavras-Chave **:** Organização da informação; Tratamento temático da informação; Catalogação de assuntos. Classificação; Indexação; Análise documental.

Abstract

The organization of information: basic concepts and a brief discussion of its theoretical principles.
The organization of information is analyzed in Information Science in two ways: as an area of research that offers the theoretical and methodological bases to the information treatment and as an activity inherent to the professional labor related to the information treatment. The knowledge of the information organization as an investigative area found the doings of this organization as an applied activity. We aim, in this paper, to explain conceptually the organization of information in the context of Information Science. We present concepts and theoretical approaches that are part of the organization of information epistemology, mainly with respect to thematic organization. We discovered that the foundations of subject cataloguing, classification, indexation and documentary analysis are approaches of the thematic information treatment responsible for the conception of organization of information as we know presently.
Key words: Organization of information; Thematic information treatment; Subject cataloguing; Classification; Indexation; Documentary analysis.

Como citar o Capítulo 6 **:**

CAFÉ, Lígia; SALES, R. Organização da informação: Conceitos básicos e breve fundamentação teórica. *In:* Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento.* Brasília DF: IBICT, 2010. 335 p. ISBN: 978-85-7013-072-3. Capíitulo 6, p. 115-129. Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

1. PORQUE ORGANIZAR?

Segundo o Dicionário Houaiss, organizar significa "dar determinada ordem a (série de itens)". Trata-se de uma característica essencial humana presente nos nossos dias desde o início de nossas vidas. Nesse sentido, Taylor e Joudrey (2008, p. 1), mencionam que para os psicólogos

> [...] a mente dos bebês organiza imagens em categorias como *faces* ou *comidas*. Crianças pequenas constroem muitas organizações e associações enquanto brincam. Na medida em que crescemos, os humanos desenvolvem habilidades cognitivas mais sofisticas para categorizar, para reconhecer padrões, ordenar, para relacionar e criar grupos de idéias e coisas.

Em uma visão mais ampla, podemos dizer que precisamos organizar para poder compreender o mundo e nos comunicarmos melhor. Por exemplo, quando queremos explicar a alguém o significado de determinada palavra normalmente recorremos a sua classificação, ou seja a incluímos em um grupo maior que a define. Assim, ao explicarmos o que significa a palavra *cadeira* diremos que é um *móvel* (classe) que possui um assento e um encosto. Neste momento, organizamos nosso pensamento sobre o conceito desta palavra focando a classe dos móveis e excluindo outras classes.

A organização dos armários de nossas casas ou das prateleiras de um supermercado são outros exemplos de organização. Eles têm como principal função encontrar os itens que necessitamos da forma mais rápida possível e assim ganharmos tempo.

No contexto dos sistemas de informação, a função não é diferente. Organizamos um acervo para compreendê-lo melhor e assim podermos recuperar objetos informacionais, isto é informações registradas nos mais variados suportes (textos, imagens, registros sonoros, representações cartográficas e páginas web). Segundo Chowdhury e Chowdhury (2007, p.3, *tradução nossa*), "Por meio da organização geramos uma ordem de itens, ou arranjamos itens em um sistema de forma que possamos encontrá-los e recuperá-los como e quando necessitarmos sem grande dificuldade".

Taylor e Joudrey (2008, p. 5-7), ampliando a lista de funções do controle bibliográfico proposta por Ronald Hagler, apresentam os principais objetivos que envolvem a organização dos objetos informacionais:

- identificar a existência de todos os tipos de recursos informacionais assim que eles estiverem disponíveis;
- identificar trabalhos contidos nestes recursos informacionais ou partes deles;
- agrupar sistematicamente esses recursos informacionais em coleções de bibliotecas, arquivos, museus, arquivos da internet, e outros tipos similares de repositórios;
- produzir listas desses recursos informacionais preparadas de acordo com regras padronizadas para citação;
- oferecer nome, título, assunto e outro acesso útil para esses recursos informacionais;
- oferecer os meios de localização de cada recurso informacional ou uma cópia dele.

Para atingir essas funções, os sistemas devem descrever cada objeto informacional adequadamente refletindo as necessidades de informação da comunidade para a qual estão destinados. Estas comunidades estão presentes em diversos contextos: bibliotecas, arquivos, museus, internet etc. A organização da informação nesses contextos adota padrões e técnicas estudados e definidos há muitos anos pela Ciência da Informação[3]. O grande e atual desafio é: como adaptá-los para a organização da informação no mundo digital e da Internet?

[3] Em período anterior à consolidação do termo 'Ciência da Informação', considerando que o termo foi cunhado no período Pós-Segunda Guerra, M. Dewey (em 1876), C. A. Cutter (em 1876, 1891-3) e P. Otlet (a partir de 1895) já se ocupavam com as definições de padrões e técnicas relativas à organização sistemática da informação em contexto bibliográfico (documental).

2. O CONCEITO DE ORGANIZAÇÃO DA INFORMAÇÃO

A organização da informação[4] é um processo de arranjo de acervos tradicionais ou eletrônicos realizado por meio da descrição física e de conteúdo (assunto) de seus objetos informacionais. Estas descrições, segundo Svenonius (2000), se constituem em um enunciado de propriedades de um 'objeto' ou das relações desse objeto com outros que o identificam. De acordo com Bräscher e Café (2008), o "produto desse processo descritivo é a **representação da informação**, entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico".

A descrição física de um objeto informacional se dá pelo processo de catalogação cujo resultado é a representação do suporte físico ou documento. Pode utilizar linguagens específicas, normas e formatos que padronizam este tipo de descrição. Um exemplo é a AACR2 (Anglo-american Cataloguing Rules 2. ed.).

A descrição do assunto recebe várias denominações na literatura como Análise documentária (CUNHA, 1989), Análise Temática (CAVALCANTI, 1978), Análise de Assunto (DIAS; NAVES, 2007), Descrição de Conteúdo (GUINCHAT; MENOU, 1994) e Tratamento Temático da Informação (Foskett, 1973). Ela é realizada pelos processos de classificação, indexação e resumo ou condensação documental[5].

Segundo Piedade (1977, p. 8), a classificação "é um processo mental habitual ao homem, pois vivemos automaticamente classificando coisas e idéias, a fim de compreendê-las e conhecê-las". Vickery (1980, p. 23), nesta mesma linha de pensamento, relata que "Classificar, na acepção mais simples do termo, é reunir coisas e idéias que sejam semelhantes entre si, e separar as que apresentam diferenças."

A classificação gera uma representação da informação no formato de números ou símbolos de classificação. Geralmente, ela descreve o conteúdo do documento de forma mais abrangente. Os instrumentos ou linguagens utilizados neste caso são os Sistemas de Classificação, cujos exemplos clássicos são a Classificação Decimal Dewey (CDD) e a Classificação Decimal Universal (CDU).

A indexação é "a operação pela qual se escolhe os termos mais apropriados para descrever o conteúdo de um documento e da pergunta do usuário" (GUINCHAT; MENOU, 1994). As representações da informação resultantes deste processo são as palavras-chave ou descritores que reunidos podem formar índices. A indexação tem a função de complementar a organização da informação realizada pela classificação, pois, ao contrário da abordagem mais genérica da classificação, ela segue os princípios de especificidade e exaustividade.

O princípio da especificidade diz respeito à identificação dos assuntos de um documento da forma mais específica possível. A este respeito Lancaster (2004, p.34) relata que "O princípio que, isoladamente, é o mais importante da indexação de assuntos, e que remonta a Cutter (1876), é aquele segundo o qual o tópico deve ser indexado sob o termo mais específico que o abranja completamente".

O princípio da exaustividade traz como orientação que os assuntos devem ser incluídos na sua totalidade em função das necessidades informacionais do usuário e do sistema de informação.

Os princípios de especificidade e exaustividade propiciam aquilo que Lancaster (2004) chama de 'coerência' e 'qualidade' da indexação, e que Gil Leiva (2008) chama de 'consistência' da indexação. Essa preocupação com uma indexação precisa e pertinente se deve ao fato de que um sistema de informação deve sempre ser organizado com vistas à recuperação eficiente da informação. A indexação pode ser realizada em linguagem natural, ou seja, sem adotar uma linguagem documentária[6]. Neste caso, denomina-se de indexação livre.

---

[4] Essa definição diz respeito à organização da informação enquanto atividade operacional inerente ao 'fazer profissional' da Ciência da Informação..

[5] A expressão condensação documental foi criada sob J. C. Gardin na década de 60.

[6] \Linguagem documentária: "[...] linguagem artificial utilizada para o registro ou indicação dos assuntos contidos nos documentos, dotada de vocabulário controlado e regida por sintaxe própria" (CAVALVANTI, 1978, p. 18). Também denominada de linguagem de indexação.

Ela pode também descrever o assunto dos documentos seguindo uma linguagem documentária como os tesauros. Neste caso, receberá a denominação de indexação controlada.

A indexação livre, apesar de ser mais prática e rápida, não controla algumas ambigüidades da língua como sinônimos[7] e homonímias[8] que provocam efeitos negativos na recuperação da informação.

A indexação controlada, ao contrário da anterior, por seguir uma linguagem documentária, padroniza a representação do assunto, minimizando os efeitos provocados na recuperação pelos fenômenos lingüísticos mencionados. Grande parte das bases de dados de resumos disponíveis no Portal Capes adota a indexação controlada. Um exemplo delas é a base LISA (*Library Information Science Abstract*) que utiliza um tesauro específico para a área da Ciência da Informação.

A elaboração de resumos é o terceiro processo de descrição de conteúdo da organização da informação. Sua função é complementar a informação dada pela classificação e pela indexação. Este é o único que não segue uma linguagem específica de padronização, mas pode obedecer a determinados métodos e normas. No Brasil, a norma atualmente adotada é a NBR 6028 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2003). Ela determina três tipos de resumos:

- **Resumo indicativo**: indica apenas os pontos principais do documento, não apresentando dados qualitativos, quantitativos, etc. De modo geral, não dispensa a consulta ao original.
- **Resumo informativo**: informa ao leitor finalidades, metodologia, resultados e conclusões do documento, de tal forma que este possa, inclusive, dispensar a consulta ao original.
- **Resumo crítico**: resumo redigido por especialistas com análise crítica de um documento. Também chamado de resenha. Quando analisa uma determinada edição entre várias, denomina-se recensão.

Tanto os processos de classificação como os de indexação e elaboração de resumos produzem a chamada informação documentária (KOBASHI, 1994). A informação documentária pode ser considerada, portanto, o elo entre o documento e o usuário e, como toda representação, não é o próprio objeto (documento). Em outras palavras, trata-se do resultado de um processo de tradução semântica daquilo que é dito pelo autor por uma representação do conteúdo do documento, levando em consideração fatores condicionantes como necessidade do usuário, domínio tratado, entre outros (KOBASHI, 1994).

Na Seção seguinte, apresentamos os principais autores nacionais e internacionais que abordaram os conceitos relacionados a organização da informação ao longo da história.

3. BREVE FUNDAMENTAÇÃO TEÓRICA PARA ORGANIZAÇÃO DA INFORMAÇÃO

A Ciência da Informação, enquanto campo científico, fornece, por meio de seus pressupostos teórico e metodológico, os alicerces que sustentam os fazeres do profissional da informação. Entendendo que a Biblioteconomia constitui-se como um dos fazeres da Ciência da Informação, é possível afirmarmos, de acordo com Pinheiro (1999) e Smit & Barreto (2002) que os saberes da Ciência da Informação fornecem as bases conceituais da prática biblioteconômica.

[7] Sinônimo: palavra com o mesmo sentido, ou aproximadamente o mesmo sentido e com formas diferentes (DUBOIS, *et. alii*, 1993 p.556)
[8] Homonímia: identidade fônica (homofonia) ou identidade gráfica (homografia) de dois morfemas que não têm o mesmo sentido, de um modo geral (DUBOIS, *et. alii*, 1993, 326).

O fazer biblioteconômico, esteja ele situado em qualquer unidade de informação (bibliotecas convencionais e digitais, arquivos, centros de documentação etc.) perpassa por um conjunto mínimo de atividades composto por produção, organização e uso da informação documental, conjunto esse chamado por Kobashi (1994) de ciclo de operações documentárias. Ao entendermos que toda informação produzida visa à sua efetiva utilização, temos na atividade de organização da informação um papel fundamental de caráter mediador, já que essa atividade é responsável pela comunicação entre a produção e o uso de informações.

A organização da informação é notadamente composta por duas dimensões relativas ao tratamento da informação – a dimensão descritiva, voltada aos elementos relativos à forma dos documentos (como na catalogação descritiva) e a dimensão temática, voltada aos conteúdos informacionais (como na catalogação de assuntos, na classificação, na indexação e na análise documental). Em que pese a segunda dimensão, Tratamento Temático da Informação (*Subject Approach to Information*) é uma denominação que vem sendo veiculada na literatura a partir de Foskett (1973), que o define como uma atividade de natureza mediadora da organização da informação voltada para o acesso ao conteúdo informacional. Segundo Barité (1998, p. 124), o Tratamento Temático da Informação (TTI) centra-se nas questões relativas "à análise, descrição e representação do conteúdo dos documentos, bem como suas inevitáveis interfaces com as teorias e sistemas de armazenamento e recuperação da informação".

A interlocução entre os contextos de produção e de uso da informação se revela nitidamente na dimensão temática, especialmente na era da internet, em que o estímulo por buscas de informações reside preponderantemente no conteúdo informacional. Focaremos, portanto, na dimensão temática da organização da informação. O tratamento temático, segundo Guimarães (2008, 2009) desenvolveu-se sob três abordagens: a catalogação de assunto (*subject cataloguing*) de matriz norte-americana (a partir do final do século XIX), a indexação (*indexing*) de matriz inglesa (a partir de meados do século XX) e a análise documental (*analyse documentair*e) de matriz francesa, notadamente a partir da década de 1960. Embora essas três vertentes se diferenciem quanto à ênfase dada aos seus fazeres – catalogação de assunto focada no desenvolvimento de produtos (como catálogos), indexação focada no desenvolvimento de instrumentos (como tesauros) e análise documental focada no desenvolvimento de referenciais teórico-metodológicos para os procedimentos envolvidos no TTI (GUIMARÃES, 2008 e 2009) – elas constroem juntas um arcabouço conceitual que, conjuntamente às idéias classificacionistas, formam a base teórica do tratamento temático da informação.

Percorrendo a literatura, deparamo-nos com nomes fundamentais que contribuíram de maneira efetiva à construção da organização da informação como a conhecemos hoje.

A história conceitual da organização da informação nos moldes atuais, manifestada notadamente na catalogação, na classificação, na indexação e na análise documental, pode começar a ser contada a partir do ano de 1876, quando do surgimento da Classificação Decimal de Dewey - CDD e da obra *Rules for a Dictionary Catalog* de Cutter.

Em momento histórico em que a idéia de informação estava diretamente ligada ao livro, o norteamericano Melvin Dewey, por meio de sua classificação decimal, se ocupou em desenvolver uma lógica organizacional para acervos de bibliotecas baseada no assunto tratado nos livros, de modo a possibilitar a consulta (recuperação) dos mesmos de maneira uniforme. A principal contribuição de Dewey para a organização da informação reside na possibilidade de dar ordem aos assuntos (conhecimentos) sob uma perspectiva hierárquica, assuntos mais gerais e assuntos mais específicos.

Em se tratando de organizar assuntos informacionais por meio de notações decimais, merece destaque o trabalho realizado pelos belgas Paul Otlet e Henri La Fontaine, criadores da Classificação Decimal Universal - CDU (1905). Lançando mão de codificações contidas de elementos semióticos, Otlet e La Fontaine concederam ao universo da organização temática, por meio da CDU, as primeiras formas de se efetuar uma organização (classificação) de assuntos com base na análise e síntese, cabendo a eles o pioneirismo da organização análitico-sintética, que surge conscientemente apenas na década de 1930 com Ranganathan.

3.1 A catalogação de assuntos de Cutter

A grande contribuição do bibliotecário norteamericano Charles Ammi Cutter (1837-1903) para o tratamento temático da informação encontra-se na obra *Rules for a Dictionary Catalog* (1876), que se situa na abordagem da Catalogação de Assunto. A catalogação de assunto enquanto abordagem teórica reflete uma tradição pragmática anglo-saxônica centrada na geração de produtos advindos do TTI em bibliotecas. Fundamentalmente de matriz norteamericana, tem seus princípios fundados na catalogação alfabética de Cutter e nos cabeçalhos de assunto desenvolvido pela *Library of Congress* (GUIMARÃES, 2009).

Ressalta-se que foi no contexto dos cabeçalhos de assunto que surgiram as reflexões iniciais sobre a ordem de citação. Segundo Piedade (1977, p. 18), "a ordem de citação determina a seqüência em que serão citados os vários conceitos presentes nos assuntos, na construção de descritores (símbolos de classificação, cabeçalhos de assunto etc)". Neste campo, encontra-se além de Cutter (1876), Kaiser (1911), Ranganathan (1930) e o Classification Research Group (1950), citados posteriormente nesta sessão.

As regras para um catálogo dicionário de Cutter definem princípios que o levou à grande notoriedade no campo da Biblioteconomia (BARBOSA, 1969; FOSKETT, 1973; MALTBY, 1975; PIEDADE, 1977, e COATES, 1988). A referida obra de Cutter recebeu esse nome porque "os cabeçalhos (autor, título, assunto e forma) são organizados como as palavras em um dicionário, em ordem alfabética" (CUTTER, 1904, p. 19. Tradução livre).

As regras para elaboração de catálogos alfabéticos de Cutter foram criadas em um contexto em que, segundo Chaumier (1988), a vertente dos estudos de linguagens documentais se volta para os padrões de consistência de índices e elaboração de esquemas de estrutura alfabético-combinatória. Quando Cutter publicou as *Rules for a Dictionary Catalog* o universo da catalogação ganhou nova roupagem, passou a ser tratado com maior rigor mediante um verdadeiro tratado de catalogação (BARBOSA, 1969 e PIEDADE, 1977), ou pilar fundamental da catalogação de assunto norteamericana (FOSKETT, 1973) e, com um caráter pragmático voltado ao usuário (MEY, 1987). Fujita (1989) lembra que as regras estabelecidas por Cutter diziam respeito não somente à construção dos cabeçalhos, mas também de seu arranjo.

As regras de Cutter para a elaboração dos cabeçalhos eram fundamentadas em três princípios básicos:

1. **Princípio específico** – Os assuntos devem dar entrada pelo termo mais específico e não pela classe a que está subordinada. Apesar de parecer óbvio, este princípio causou impacto porque na época era comum a adoção de cabeçalhos bastante gerais, que pouco representavam o assunto da obra em questão (CESARINO; PINTO, 1978, p. 274)
2. **Princípio de uso** – para ele "os cabeçalhos serão aqueles sob os quais é provável que a maioria dos americanos educados irão procurar, com referências cruzadas para outras formas de cabeçalhos relacionados". É o princípio da conveniência, de acordo com a necessidade dos usuários (CESARINO; PINTO, 1978, p. 275)
3. **Princípio sindético** – Por se basearem no alfabeto dos cabeçalhos de assunto fazem aproximações absurdas de assuntos e ao mesmo tempo, separam assuntos relacionados. Assim Cutter propôs o desenvolvimento nas listas de cabeçalhos de assunto de estruturas sindéticas que, através de uma rede bem construída de referências cruzadas, poderiam ajudar ou mesmo superar este problema (CESARINO; PINTO, 1978, p. 275).

Quando Cutter define que *os assuntos devem dar entrada pelo termo mais específico e não pela classe a que está subordinada* (princípio específico), deve-se ater ao fato de que, como afirma Foskett (1973), Cutter apresentava razões mais econômicas e menos teóricas, motivo pelo qual sua abordagem alfabêtica não contemplava as remissivas ao tipo transcendente – de assuntos gerais para específicos.

Quanto ao princípio de uso, Cutter apresentava uma postura dedutiva, pontuando que o cabeçalho deveria ser escolhido devacordo com a dedução de quais os termos seriam provavelmete relacionados pelo usuário no momento da busca.

O princípio sindético de Cutter, o qual sustenta que há a necessidade de uma rede cruzada de referências entre os cabeçalhos de assunto, demonstra um pioneirismo fundamental em que pese à questão do 'relacionamento entre assuntos'. Verifica-se assim que Cutter fincou pilares iniciais para a futura aparição da classificação facetada, tornando-se, portanto, um dos precursores da representação do conhecimento baseada em princípios voltados ao relacionamento entre assuntos.

3.2 A Indexação de Kaiser

O bibliotecário alemão Julius Otto Kaiser (1968-1927) tem seu nome gravado nos escritos oriundos da classificação e indexação devido à publicação da obra *Systematic Indexing* (1911). Com a referida obra, em que o autor apresenta uma maneira sistemática para a realização do processo de indexação de assuntos de documentos, o universo do tratamento da informação, até então familiarizado com a classificação e com a catalogação alfabética de assuntos, direciona também a atenção para a indexação alfabética, prática esta que foi revigorada com o sistema de Kaiser.

Atento às coleções de conteúdos heterogêneos, Kaiser, de certa forma, deu continuidade aos estudos de Cutter no que se refere à catalogação de assunto (COATES, 1988; RODRIGUEZ, 1984). Assim como Cutter procurou resolver o problema da inconsistência na catalogação, definindo regras para a elaboração de cabeçalhos de assunto, Kaiser elaborou uma forma sistemática, também permeada por regras, para a construção de sentenças terminológicas que melhor representassem o assunto de um livro. O que na época de Cutter era denominado cabeçalhos de assunto (*subject headings*), Kaiser chamou de enunciado (*statement*). O enunciado, a exemplo do cabeçalho de assunto, nada mais era do que uma palavra ou grupo de palavras que expressavam o conteúdo de um documento. Porém, o avanço de Kaiser reside notadamente nos elementos que compõem tal enunciado.

Partindo do pensamento de que "no mundo existem *coisas* e que essas coisas podem ser *ditas* ou *descritas*" (RODRIGUEZ, 1984, p.164), Kaiser preconizou que todos os assuntos podem ser mais bem representados e descritos por meio de duas categorias fundamentais: Concretos e Processos. O assunto propriamente dito, ou seja, *a coisa da qual um livro trata é o Concreto*, ao passo que *o que é dito a respeito dessa coisa (concreto) é o Processo.*

A importância das duas categorias fundamentais, cocreto e processo, é explicada pelo próprio Kaiser por meio da afiamação de que "concretos são conhecidos por nós apenas superficialmente [...] somos incapaces de dar uma descrição comapleta a qualquer concreto" (KAISER, 1911, p.54, tradução nossa). E segue: "não podemos dizer o qoe são os concretos, somos obrigados a a dar maior atenção aos seus processos, para o que eles fazem ou o que podemos fazer com eles" (p.55, tradução nossa).

Preocupado com o nível de especificidade dos enunciados, Kaiser definiu também a categoria Localidades (ou lugar). Foskett (1973) e Coates (1988) expllicam que caso esteja incluído um lugar no assunto do documento, o sistema de Kaiser define que seja feita uma entrada dupla, ou seja, um enunciado a partir do concreto e outro a partir da localidade.

Além de estabelecer categorias,, Kaiser, a exemplo de Cutter, definiu regras para a formação dos enunciados. São seis as regras estabelecidas com o objetivo de obter êxito no processo de indexação. A saber:

- Selecione o que é realmente importante para seu objetivo sem considerar forma ou extensão;
- Concentre-se na informação relativamente específica;
- Lide com cada item independentemente;

- Não adultere o nome dos concretos;
- Evite inversão, proposições e plurais sempre que possível;
- Teste a exactidão de cada enunciado pelo ponto de vista tanto do indexador como do usuário (KAISER, 1911, p.348, tradução nossa).

Observamos que Kaiser se assemelha a Cutter nos focos bordados em seus princípios de especificidadede, sindético e de uso. Enquanto Cutter que os assuntos "devem dar entrada pelo termo mais específico e não pela classe a que está subordinado", Kaiser previu a necessidade que havia de possibilitar refrências de assuntos específicos para os assuntos gerais, inexistente no catélogo dicionário de Cutter (FOSKETT, 1986). Segundo Foskett (1973), a ligação dos conceitos em redes de remissivas coordenadas de forma superordenada e subordinada,prevista no sistema de Kaiser, é o mais adequado.

Mills (1960) explica que os termos eram relacionados na rede de remissivas de Kaiser por meio de cartões guias que especificavm, além das relações de termos super ordenados e subordinados, também termos sinônimos e qualquer outro termo/concreto que apresentasse alguma conexão encontrada no momento da indexção da coleção. Notamos, assim, que essas relações entre termos, tecidas no *systematic index*, são continuidade do *princípio sindético* preconizado por Cutter.

Kaiser volta-se tmbém à preocupação apresentada no *princípio de uso* de Cutter. Ambos os sistemas estão voltados não apenas para o êxito da representação (catalogação e indexação), mas principalmente para a satisfção das necessidades dos usuários. E, nesse ponto, mais uma vez é possível verificar um avanço de Kaiser com relação ao trbalho iniciado por Cutter. Enquanto este apresentava uma postura mais dedutiva, pontuando que o cabeçalho deveria ser escolhido de acordo com a dedução de quais termos seriam provavelmente relacionados pelo usuário, aquele buscava extrair do próprio usuário a exatidão dos enunciados.

As idéias de Kaiser são estudadas tanto pelo universo investigativo da classificação quanto da indexação, fato que torna possível afirmar que ele estabeleceu o elo entre as classificações filosóficas e bibliográficas, sistemas nos quais prvelecem as visões e as lógicas dos criadores dos sistemas, e a indexação de assuntos, na qual o que impera são as visões dos utilizadores do sistema. Lançando mão de princípios de classificação, Kaiser concebeu uma nova forma de indexar.

3.3 A classificação de Ranganathan

Entre 1933 e 1960, o bibliotecário e matemático indiano Shiyali Ramamrita Ranganathan (1892-1972), por meio da criação e do desenvolvimento da *Colon Classification* (CC), transcende o papel das classificações lineares e rigorosamente hierárquicas existentes até então, configurando um novo modo de classificar assuntos. Algo semelhante, porém nem tão repercutido, ocorreu, como já mostrado anteriormente, com Cutter – que criou um novo modo de formar cabeçalhos – e com Kaiser – que criou um novo modo de indexar assuntos. Portanto, a transformação ocorrida no universo das classificações diante da publicação da *Colon Classification* é o efeito não do surgimento de uma nova classificação, mas sim da emersão de um novo método.

Embora Dewey já tenha antes incluído alguns elementos de síntese em suas notações decimais, e Otlet e La Fontaine já tivessem lançado mão de um conjunto semiótico nas notações da CDU, Ranganathan foi quem fortaleceu o desenvolvimento de um sistema puramente analítico-sintético, pois, como informa Foskett (1973), ele conseguiu demonstrar que análise e síntese podem ser aplicadas a qualquer classe de assunto, e assim ser siste matizadas. Ademais, Ranganathan desenvolveu sob rigor científico uma terminologia própria utilizada até hoje para os estudos de classificação. Termos

como facetas (*facets*), focos (*focus* ou *foci*), ciclos (*rounds*), níveis (*levels*), coordenação e subordinação de assuntos (*arrays and chains*) foram incorporados no universo investigativo das classificações a partir de Ranganathan.

A lógica de Ranganathan era baseada no agrupamento de assuntos que apresentavam as mesmas características e os mesmos tipos de relacionamentos. As subdivisões do conhecimento podiam se originar por meio de diversas relações, entre elas, gênero-espécie, todo-parte, propriedade-possuidor, ação-paciente ou agente etc.

Ranganathan concebeu, para além de um sistema de classificação, bases teóricas que propiciaram a construção de códigos capazes de classificar qualquer assunto, por meio de regras e esquemas classificatórios.

Para alcançar seu modelo teórico, Ranganathan desenvolveu o princípio das categorias fundamentais: Personalidade (P), Matéria (M), Energia (E), Espaço (S) e Tempo (T), conhecido como PMEST. É possível observarmos uma relação direta entre as categorias Personalidade, Energia e Espaço com as categorias Concreto, Processo e Lugar de Kaiser. Esse princípio zelava pela ordem de citação de concretude decrescente. Como ressaltam Foskett (1973) e Coates (1988), a análise baseada nas categorias PMEST pode não solucionar por completo a questão da ordem de citação, mas muitas vezes são uteis para o estabelecimento da ordem de citação adequada de assuntos. Além do mais, não há, segundo os autores, registro de esquemas classificatórios que tenham tratado tal questão com tanto empenho e clareza até então.

Para Ranganathan (1976), cada faceta de qualquer assunto é considerada como uma manifestação de uma das cinco categorias fundamentais: Personalidade (objeto estudado), Matéria (materiais e substâncias), Energia (ações e processos), Espaço (local) e Tempo (época). Para ele as cinco categorias fundamentais são dadas sob uma seqüência crescente de concretude com os respectivos tipos de ligação: apóstrofe invertido (‘) Tempo, ponto (.) Espaço, dois pontos (:) Energia, ponto e vírgula (;) Matéria, vírgula (,) Personalidade (p. 1.25).

A característica principal da *classificação de Ranganathan* é a sua subdivisão em facetas e focos. *Faceta* pode ser entendida como um aspecto (característica) particular de um assunto, e cada faceta é dividida por *focos*, que segundo Foskett (1973) e Lima (2004), são termos que servem para formar um assunto em cada contexto.

Imaginemos a classe de assunto ‘Biblioteconomia’ subdividida segundo a faceta ‘Item Bibliográfico’ de coleções de bibliotecas (*Quadro 1*):

Quadro 1. – Faceta e Focos

CLASSE	FACETA	FOCOS
Biblioteconomia	Item Bibliográfico	Livro Periódico Dicionário Tese

(Elaboração pelos autores)

Os termos Livro, Periódico, Dicionário, Anais e Tese são *focos* porque estão inseridos no contexto da *faceta* ‘Item Bibliográfico’. Ao concentrarmos a atenção somente em um dos focos, como por exemplo, no termo Livro, pode-se afirmar que na faceta ‘Item Bibliográfico’ o foco atingido foi o Livro.

A exemplo de Cutter e Kaiser que respectivamente se preocuparam em conceber uma nova maneira de catalogar e indexar assuntos, Ranganathan, com sua classificação facetada, alavancou uma nova maneira de classificar assuntos. Indo além da concepção de um sistema classificatório, ele criou

uma metodologia para a construção de classificações, propiciando esquemas autoperpetuantes. Sua fundamentação pautada na análise e síntese de facetas e focos fez de seu trabalho um marco no compromisso de extrair os aspectos mais significativos de assuntos de documentos.

As bases lógicas de Ranganathan amadureceram os princípios de uso, de especificidade, de relacionamento entre assuntos e de estabelecimento de categorias e de regras presentes ora em estudos de Cutter ora em estudos de Kaiser, consolidando uma teoria – a teoria da Classificação Facetada.

3.3.1 O Classification Research Group (CRG)

Notória importância tem o trabalho realizado pelos pesquisadores do *Classification Research Group* (CRG), primeiro grande grupo de pesquisa no campo das classificações, formado em Londres, em 1952, pelos professores A.J. Wells e B.C. Vickery. A equipe era constituída por professores, bibliotecários e documentalistas que se ocuparam com a discussão e trocas de experiências no que diz respeito à classificação e à organização do conhecimento, tendo assim a responsabilidade de trilhar caminhos que avançassem as questões metodológicas referentes à organização da informação especializada. Os membros do CRG avançaram a teoria de Ranganathan. Contribuíram para o surgimento de pesquisas que propiciaram a sedimentação de uma Teoria da Classificação. Para o pessoal do CRG, o universo do conhecimento era o assunto específico (especializado) a ser classificado. E a partir daí elaboraram esquemas de classificação que cobriam assuntos especializados, além de aprofundarem princípios e métodos acerca das classificações.

O CRG definiu doze categorias possíveis para representação de assuntos: tipos de produto final, partes, todo, materiais, constituintes, processos, propriedades, operações, agentes, espaço (lugar), tempo e formas de apresentação. Percebemos aqui, um avanço das categorias do CRG com relação às categorias de Ranganathan refletindo a mudança no universo do conhecimento a ser organizado, visto que o foco dos estudos do CRG eram os assuntos especializados.

3.3.2 A International Society for Knowledge Organization (ISKO)

Dando continuidade ao contexto da organização da informação, há que se destacar a *International Society for Knowledge Oragnization* (ISKO), criada em Frankfurt, em julho de 1989 por Ingetraut Dahlberg, a partir dos trabalhos realizados pela *Society for Classification* (GUIMARÃES, 2008). Sua base teórica reside especialmente nos princípios da classificação e nas pesquisas sobre tesauros, com particular destaque para os trabalhos de Ranganathan e do CRG. A ISKO foi criada com base em duas vertentes da Biblioteconomia internacional, a primeira ligada ao estudo e desenvolvimento de tesauros como instrumentos alfabéticos de tratamento temático e a segunda vinculada à aplicação das teorias facetadas à elaboração de tesauros.

Os pesquisadores que atualmente compõem o quadro investigativo da ISKO são oriundos das mais variadas áreas do conhecimento: Ciência da Informação, Filosofia, Lingüística, Ciência da Computação, entre outras envolvidas com a organização do conhecimento e da informação. Segundo Guimarães (2008) a criação da ISKO propiciou à área de organização da informação um campo de reflexão e produção teórica, proporcionando um representativo fórum científico internacional.

3.4 Análise documental

No final da década de 1960, na França, J.-C. Gardin e M. Coyaud, preocupados com o desenvolvimento técnico-metodológico dos processos inerentes ao tratamento temático da informação, dão início a uma visão mais abrangente, chamada *Anlyse Documentaire* (análise documental). Consistindo em uma "operação intelectual" de representação dos conteúdos documentais voltada à recuperação (CHAUMIER, 1982; SERRES, s.d.), a

Análise documental ancora-se em uma matriz de influência lingüística. Os produtos da análise documental são os resumos, resultado do processo de condensação, e os índices, resultado do processo de indexação.

A análise documental pode ser entendida como sendo a decomposição das partes de um todo para maior compreensão do conteúdo informacional, para fins de representação e, consequentemente, de recuperação. Segundo a visão de grande parte dos pesquisadores brasileiros, os processos de leitura, análise, síntese e representação de conteúdos informacionais compõem à atividade intelectual e fundamental da análise documental, Guimarães e Sales (2010). Diante de tal constatação, a análise do conteúdo de documentos, como ressalta Guimarães (2009), passa a ser o cerne do tratamento temático da informação.

A partir desta perspectiva francesa para análise documental, e como visto acima, refletida nos estudos brasileiros, o ponto principal passa a ser a própria explicitação dos procedimentos empregados para a realização da análise e da representação do conteúdo dos documentos. Dá-se neste momento maior ênfase ao desenvolvimento teórico-metodológico do tratamento temático em si e menor à geração de produtos e instrumentos. Para Chaumier (1982) a análise documental está no núcleo dos problemas da área, condicionando assim o valor do sistema documental (informacional).

Dialogando com a linha gardiniana, pesquisadores espanhóis como Martínez de Souza (1989), Coll-Vinent & Bernal Cruz (1990), Pinto Molina (1992), Ruiz Perez (1992) e Fox (2005) convergem na idéia de que a análise documental é uma atividade eminentemente intelectual e, complementando a visão francesa, entendem que a análise documental contempla tanto a dimensão física (material) quanto a dimensão temática (conteúdo) do documento.

No Brasil, onde os estudos de análise documental são fortemente influenciados pela abordagem francesa, identifica-se nas pesquisas desenvolvidas pelo Grupo TEMMA (ECA/USP) – grupo criado no início da década de 1980 pela professora Johanna Smit – o espaço acadêmico responsável pela germinação e crescimento dos fundamentos teóricos e metodológicos da Análise documental neste país. Focado nas reflexões tanto teóricas quanto práticas da Análise Documental, o Grupo TEMMA, cientificamente reconhecido nos estudos de Anna Maria Marques Cintra, Mariângela Spotti Lopes Fujita, João Batista Ernesto de Moraes, Marilda Lopes Ginez de Lara, Johanna Wilhelmina Smit, Nair Yumiko Kobashi, José Augusto Chaves Guimarães, Vânia Mara Alves Lima, Maria de Fátima Gonçalves Moreira Tálamo entre outros, vem diversificando seu espectro investigativo entre estudos de terminologia aplicada à organização e transferência da informação, nos processos de leitura em análise documental e linguagem de especialidade da Ciência da Informação.

Em pesquisa acerca da concepção conceitual da análise documental brasileira, Guimarães & Sales (2010) apresentam um panorama da literatura produzida pelo Grupo TEMMA de acordo com as vertentes temáticas, como mostrado no Quadro 2, na página seguinte:

4. CONCLUSÃO DO CAPÍTULO 6

Como foi possível observar, a habilidade que hoje temos nos processos de organização da informação é fruto de uma longa história tecida por profissionais e pesquisadores nacionais e internacionais preocupados em tornar a informação acessível por meio de sua descrição. As iniciativas norte-americana, inglesa e francesa relatadas neste trabalho formam as bases teórica e metodológica que fundamentam as atividades e pesquisas atuais. Todas elas têm como objetivo organizar a informação para entendê-la melhor e, desta forma, recuperá-la de forma mais eficiente.

Para os estudantes de Biblioteconomia, e de demais fazeres da Ciência da Informação, é desejável ter clareza de que o conjunto formado pelos estudos de catalogação de assuntos, classificação, indexação e análise documental forma a base teórica e prática do tratamento temático da informação, consequentemente, uma base substancial da organização da informação.

Quadro 2. – Temas e Pesquisadores (Grupo TEMMA)

Temas tratados na análise documental	Pesquisadores
Bases epistemológicas: no sentido de se resgatar os aspectos históricos que levaram à construção da disciplina bem como à constituição de seus marcos teóricos	Cunha, Kobashi, Lara, Fujita, Guimarães, Liberatore, Lara, Smit, Tálamo e Vogel.
Processos em análise documental, englobando a leitura	Cintra, Fujita, Nardi, Santos, Rubi e Fagundes
Processos em análise documental, englobando a condensação	Guimarães e Kobashi
Processos em análise documental, englobando a representação documental	Lara
Instrumentos de análise documental, notadamente as linguagens documentais	Cintra, Tálamo, Kobashi, Lara e Tálamo
Relações interdisciplinares da análise documental com a Linguística	Cintra, Kobashi, Lara, Smit e Tálamo.
Relações interdisciplinares da análise documental com a Terminologia	Cintra , Lara, Tálamo, Kobashi, Smit e Lenzi
Relações interdisciplinares da análise documental com a Lógica	Cunha, Guimarães e Kobashi
Relações interdisciplinares da análise documental com a Diplomática	Guimarães, Nascimento e Moraes
Análise documental em universos específicos: em Arquivos	Smit, Guimarães e Kobashi
Análise documental em universos específicos: em Museus	Cerávolo e Tálamo
Análise documental da imagem	Boccato, Fujita, Smit, Gonçalves e Macambyra
Análise documental de documentos jurídicos	Guimarães e Barité
Aspectos humanos e sociais: Ética em análise documental	Fernandez-Molina & Guimarães, 2002;, Fernandez-Molina, Guimarães, Lara e Cioffi
Aspectos humanos e sociais : ensino e pesquisa	Danuello, Menezes, Guimarães, Rubi e Fujita
Aspectos humanos e sociais: Políticas de análise documental	Guimarães

Fonte: Guimarães & Sales (2010)

Sabemos que a informação só tem sentido se for transmitida e socializada. No entanto, para que isso ocorra é necessário que ela circule e, para tanto, ela deve ser tratada nos seus aspectos formais e temáticos de forma adequada. Nesse sentido, uma perspectiva histórica e contextualizada deve estar sempre presente nos estudos relacionados ao tratamento dos objetos informacionais da nova era digital e da Internet.

5. REFERÊNCIAS DO CAPÍTULO 6

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR6028:* Informação e documentação: resumo. apresentação. Rio de Janeiro, 2003. 2p.

BARBOSA, A. P. *Teoria e prática dos sistemas de classificação bibliográfica*. Rio de Janeiro: Instituto Brasileiro de Bibliografia e Documentação, 1969. 441 p. (Obras Didáticas; 1).

BARITÉ, M. *Referenciales teóricos vigentes en el área de tratamiento temático de la información y su expresión metodológica.* Porto Alegre: ABEBD, 1998. 7 p. [Relatório técnico do II Encontro de Dirigentes dos cursos superiores de Biblioteconomia dos países do Mercosul, Buenos Aires, nov. 1997].

BRÄSCHER, M. CAFÉ, L. Organização da Informação ou Organização do Conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008. São Paulo. *Anais...*São Paulo: USP, 2008.

CAVALCANTI, C.R. *Indexação e tesauro*: metodologia e técnicas. Brasília: ABDF, 1978. 87 p.

CESARINO, M. A. da N.; PINTO, M. C. M. F. Cabeçalho de assunto como linguagem de indexação. *Revista da Escola de Biblioteconomia*, UFMG, Belo Horizonte, ano 7, n. 2, p. 268-88, set. 1978.

CHAUMIER, J. *Analyse et langages documentaires*: le traitement linguistique de l'information documentaire. Paris:Entreprise Moderne, 1982.

CHAUMIER, J. Indexação: conceito, etapas, instrumentos. Trad. José Augusto Chaves Guimarães. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v.21, n.1/2, p. 63-79, jan./jun. 1988.

CHOWDHURY, G.G.; CHOWDHURY, S. *Organizing information*: from the shelf to the web. London: Facet Publishing, 2007. 230 p.

COATES, E. J. *Subject catagues*: headings and structure. London : The Library Association, 1988. 186p.

COLL-VINENT, R.; BERNAL CRUZ, F. J. *Curso de documentación*. Madrid: Sossat, 1990.

CUNHA, I. M. R.F. (Coord.) *Análise Documentária*: considerações teóricas e experiências. São Paulo:FEBAB, 1989. 191 p.

CUTTER, C. A. *Rules for a dictionary catalog*. 4 ed. Washington: Government Printing Office, 1904. 173 p. (U.S. Bureau of Education. Special Report on Public Libraries; 2)

DEGEZ, D. ; MELLINET, D. *Thésaurusglossaire des langages documentairtes* : un outil de contrôle sémantique. Paris : ADBS, 2001.

DIAS, Eduardo Wense; NAVES, Madalena Martins Lopes. *Análise de assunto.* Brasília: Thesaurus, 2007. (Estudos Avançados em Ciência da Informação, 3)

DUBOIS et. alii. *Dicionário de lingüística*. São Paulo: Cultrix, 1993. 653 p.

FOSKETT, A.C. *A abordagem temática da informação*. Tradução de Antônio Agenor Briquet de Lemos. São Paulo: Polígono; Brasília: Ed.UnB, 1973.

FOSKETT, A.C. *The subject approach to information.* 5. ed. London: Clive Bingley, 1986.

FOX, V. *Análisis documental de contenido*: principios y prácticas. Buenos Aires: Alfagrama, 2005.

FUJITA, M. S. *PRECIS na língua portuguesa*: teoria e prática de indexação. Brasília: Editora da UNB, 1989.

GIL LEIVA, I. Evaluación de la indización. *In*: GIL LEIVA, I. *Manual de indización: teoria y práctica*. Gijón, Espanha: Trea, 2008, p. 385-397. (Biblioteconomía y Administración Cultural, 193).

GUIMARAES, J. A. C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ibero-Americana de Ciência da Informação*, v. 1, p. 77-99, 2008.

GUIMARAES, J. A. C. Abordagens teóricas de tratamento temático da informação (TTI): catalogação de assunto, indexação e análise documental. In: GARCÍA MARCO, F. J. *Avances y perspectivas en sistemas de información y documentación*. Ibersid, Zaragoza, 2009, p. 105-117.

GUIMARÃES, J. A. C.; SALES, R. Análise Documental: concepções do universo acadêmico brasileiro em ciência da informação. *DataGramaZero* - Revista de Ciência da Informação, v.11, n.1, fev., 2010. Disponível em <http://dgz.org.br/fev10/F_I_art.htm>. Acessado em: 21 mar 2010.

GUINCHAT, C.; MENOU, M. A descrição de conteúdo. In*:___. Introdução às ciências e técnicas da informação e da documentação*. 2.ed. Brasília: IBICT, 1994.

HOUAISS, Antônio. *Dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2001.

KAISER, L. O. *Systematic indexing*. London: Isaac Pitman & Sons, 1911 (The Card Sistem Series, 2). Disponível em <http://www.archive.org/stream/systematicindexi00kaisuoft#page/n0/mode/2u>. Acessado em: 28.01.10.

KOBASHI, Nair Yomiko. *A elaboração de informações documentárias*: em busca de uma metodologia. São Paulo, 1994. 195 f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicação e Artes, Universidade de São Paulo, 1994.

LANCASTER, F. W. *Indexação e resumos*: teoria e prática. Trad. Antonio Agenor Briquet de Lemos. 2. ed. Brasília: Briquet de Lemos/Livros, 2004. 452 p.

LIMA, G. A. B. O modelo simplificado para análise facetada de Spiteri a partir de Ranganathan e do Classification Research Group, *Inf. Cult. Soc.,* Bueno Aires, n.11, jul./dic. 2004.

MALTBY, A. *Sayers' Manual of classification for librarians*. London: A. Deutsch / A Grafton Book, 1975. 336 p.

MARTÍNEZ DE SOUZA, J. *Diccionario de Bibliología y ciencias afines*. Madrid : Fund. Germán Sánchez Ruipérez, 1989.

MEY, E. S. A. *Catalogação e descrição bibliográfica:* contribuições a uma teoria. Brasília: Associação dos Bibliotecários do Distrito Federal, 1987, 201p.

MILLS, J. *A morden outline of library classification*. London, Chapman and Hall, 1960. 196 p.

PIEDADE, M. A. R. *Introdução à teoria da classificação.* 2. ed. Rio de Janeiro: Interciência, 1983, 221 p.

PIEDADE, M. A. R. *Introdução à teoria da classificação*. Rio de Janeiro: Interciência, 1977. 185 p.

PINHEIRO, L. V. R. Campo interdisciplinar da ciência da informação: fronteiras remotas e recentes. In.: PINHEIRO, L. V. R. (org.). *Ciência da informação, ciências sociais e interdisciplinaridade*. Barília: IBICT/DDI/DEP, 1999. p. 155-182.

PINTO MOLINA, M. *El resumen documental: principios y métodos*. Madrid; Salamanca: Fundación Germán Sánchez Ruipérez; Madrid: Pirámide, 1992.

RANGANATHAN, S. R. *Colon classification*. 6. ed., 4. reimpr. Bombay; Calcutta; New Delhi; Madras; Lucknow; Bangalore; London; New York: Asia Publishing House, 1976.

RODRIGUEZ, R. D. Kaiser's systematic indexing. *Library resources & technical services*, p. 163-173, apr./jun, 1984.

RUIZ PEREZ, R. *El analisis documental* : *Bases terminológicas, conceptualización y estructura operativa*. Granada: Ed. Universidad de Granada, 1992.

SERRES, A. *Introduction a l'indexation lexique de l'indexation documentaire*, s.d. Disponível em <http://www.uhb.fr/urfist/Supports/Indexation/IndexationLexique.htm>. Acessado em: 18 mar 2010.

SMIT, J.; BARRETO, A. A. Ciência da informação: base conceitual para a formação do profissional. In.: VALENTIM, M. L. P. (org.). *Formação do profissional da informação*. São Paulo: Polis, 2002. p. 9-23.

SVENONIUS, E. *The intellectual foundations of information organization*. Cambridge: The MIT Press, c2000. 255p.

TAYLOR, A. G; JOURDREY, D.N. *The organization of the information*. 3.ed. Westport: Libraries Unlimited, 2008. 513 p.

VICKERY, B.C. *Classificação e indexação nas ciências*. Trad. M.C.G. Pirolla. Rio de Janeiro: BNG/Brasilart, 1980. 274p.

Capítulo 7

Organização da informação: Proposta de elementos de arquitetura da informação para repositórios digitais institucionais, baseados na descrição física e temática

Fernanda de Souza Monteiro[1]**Organização da informação no ambient dig**

SUMÁRIO DO CAPÍTULO 7

Resumo do Capítulo 7, *p. 131*
Como citar o Capítulo 7, *p. 131*

1. **INTRODUÇÃO DO CAPÍTULO 7**, *p. 132*
2. **REPOSITÓRIOS DIGITAIS INSTITUCIONAIS**, *p. 132*
3 . **ORGANIZAÇÃO DA INFORMAÇÃO**, *p. 133*
3.1 **Organização da informação no ambiente digital**, *p. 134*
4. **ARQUITETURA DA INFORMAÇÃO**, *p. 134*
4.1 **Modelo de Rosenfeld e Morville**, *p. 135*
5. **METODOLOGIA**, *p. 137*
6. **RESULTADOS**, *p. 137*
7. **ANÁLISE DOS RESULTADOS**, *p. 138*
8. **CONJUNTO DE ELEMENTOS DE AI, BASEADOS NA DESCRIÇÃO FÍSICA E TEMÁTICA, CONSIDERADOS APROPRIADOS**, *p. 142*
9. **CONSIDERAÇÕES FINAIS DO CAPÍITULO 7**, *p. 144*
10. **REFERÊNCIAS DO CAPÍTULO 7**, *p. 144*

[1] Mestre em Ciência da Informação. Programa de Pós-Graduação em Ciência da Informação e Documentação (PPGCInf), Universidade de Brasília (UnB). Professora da Universidade Federal de Goias (UFGO). CV Lattes: http://lattes.cnpq.br/2216821765371227. E-mail: fernanda.s.monteiro@gmail.com.

Resumo do Capítulo 7

Trata a perspectiva cada vez mais dinâmica da informação no meio digital, favorecida pelo acesso livre e arquivos abertos que, dispostos em repositórios, vislumbram a possibilidade de reunir, divulgar e disponibilizar informações científicas. Neste contexto, essas informações encontram alternativa aos mecanismos tradicionais de comunicação, mas esbarra na problemática ocasionada pelo amplo universo de informação disponível: a dificuldade de organizá-la coerentemente para facilitar sua recuperação. A partir destas considerações, este estudo discorre sobre elementos de Arquitetura da Informação (AI) e procedimentos de descrição física e temática que possam contribuir para a organização dos documentos armazenados em repositórios digitais institucionais.
Palavras-chave **:** Organização da informação, Arquitetura da informação, Repositórios digitais institucionais.

Como citar o Capítulo 7

MONTEIRO, Fernanda. Organização da informação: proposta de elementos de arquitetura da informação para repositórios digitais institucionais, baseados na descrição física e temática. *In:* Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento.* Brasília DF: IBICT, 2010. 335 p. ISBN: 978-85-7013-072-3. Capíitulo 7, p. 130-145. Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

1. INTRODUÇÃO DO CAPÍTULO 7

Os arquivos abertos, dispostos em repositórios na Internet, reúnem, divulgam e disponibilizam uma vasta quantidade de informações. A organização dessas informações interfere na localização e utilização dos documentos, ressaltando sua importância. Neste sentido, esta pesquisa discorre sobre elementos de Arquitetura da Informação (AI) baseados em procedimentos de descrição física e temática que possam contribuir para a organização dos documentos armazenados em Repositórios digitais Institucionais (RI).

RIs são sistemas de informação disponíveis na Internet, auxiliados por ferramentas, estratégias e metodologias que caracterizam um modelo de comunicação científica. Estes assimilam dificuldades encontradas na Internet, tais como a grande quantidade de informações não organizada e a suscetível indisponibilidade de acesso (BERGMAN, 2001; BERNERS-LEE, Tim; HENDLER, James; LASSILA, 2001). Com o auto-arquivamento a problemática dos repositórios digitais aumenta, pois em geral, os autores não possuem conhecimentos específicos para a descrição e organização dos documentos.

No intuito de identificar estes elementos, o modelo de Rosenfeld e Morville (2002) é visto como uma abordagem adequada, pois permite que a amplitude do seu conceito de arquitetura seja aplicada às premissas de um repositório digital. Além disso, o modelo dos autores considera a relação entre o usuário, o contexto e o conteúdo na definição da arquitetura. Assim, justifica-se o interesse em enriquecer o estudo sobre esta temática e contribuir para a organização da informação dos repositórios digitais, fornecendo um conjunto de elementos capazes de nortear o desenvolvimento dos repositórios no que diz respeito à representação de suas informações.

2. REPOSITÓRIOS DIGITAIS INSTITUCIONAIS

Um levantamento sobre repositórios digitais, as características e peculiaridades que influenciam sua organização da informação, permitem delinear nosso objeto de estudo, os repositórios digitais institucionais (RIs). Compostos por repositórios temáticos, desenvolvidos diante da perspectiva do acesso livre a informação, com o intuito de disponibilizar arquivos abertos conforme as especificações da *Open Archives Initiative* (OAI) e demais manifestações citadas, estes repositórios, ainda que sejam provedores de dados, também provêem serviços, agregando e disponibilizando coleções distintas, permitindo sua organização, disseminação e acesso (COSTA, 2007).

Para este estudo, os repositórios digitais institucionais, contemplam diferentes tipos de documentos e ampla diversidade de formatos. Esta característica irá conduzir à definição de uma amostra de repositórios mantidos por instituições universitárias que disponibilizam coleções digitais de departamentos e faculdades. Sendo o objetivo do estudo contribuir para a organização da informação, o processamento automático das discussões entre os pares, a geração de versões de um mesmo documento e a preocupação com a preservação digital, são características possivelmente encontradas nos repositórios estudados, mas não interessam diretamente para os aspectos avaliados.

Quanto à interoperabilidade, o uso do protocolo *Open Archives Initiative - Protocol for Metadata Harvesting* (OAI-PMH) e de seu conjunto mínimo de metadados, o Dublin Core, são relevantes. Estas considerações fornecem diretrizes para a descrição dos documentos, procedimento que influencia a organização da informação. O auto-arquivamento permite inferir que não se pode contar com a expertise do autor sobre o processo de submissão e descrição dos documentos, tornando essencial que os repositórios digitais apresentemm uma estrutura de organização da informação, pois esta representa a abrangência de seu conteúdo, a instituição que lhe dá subsídios e o interesse de seus usuários. A estrutura, tal como o processo de submissão, está condicionada à ferramenta escolhida para o desenvolvimento do repositório. A interface *Web* e os recursos de busca desta ferramenta
.

fornecem acesso ao conteúdo, finalidade da disseminação da informação, objetivo de sistemas de informação tanto tradicionais quanto digitais (CROW, 2002).

3 . ORGANIZAÇÃO DA INFORMAÇÃO

A informação pode ser compreendida como conhecimento registrado em um suporte para sua comunicação (LE COADIC, 2004). Para este autor, a informação comporta um elemento de sentido, é um significado transmitido a um ser consciente por meio de uma mensagem inscrita em suporte material-temporal (impresso, sinal elétrico, onda sonora, etc). Svenonius (2001), ao dizer que informação é uma abstração e os documentos que a contêm estão em um formato, como papel ou arquivos de computador, conclui que o documento é o suporte da informação. O acesso a estes documentos ocorre por intermédio de Unidades de Informação (UI) que os reúnem e disponibilizam. Para facilitar este acesso as UIs têm a preocupação de organizar a informação, utilizando procedimentos que são anteriores ao desenvolvimento das tecnologias atuais para comunicação da informação, como os repositórios digitais.

A organização da informação (OI), tanto no ambiente tradicional, com bibliotecas e documentos impressos, quanto no ambiente digital, com a *Web* e arquivos digitais, ocorre em um sistema de informação constituído de componentes inter-relacionados para coletar (entrada), tratar (processamento) e disseminar (saída) informações, sob gestão de um responsável, que conta com um mecanismo de *feedback*[2] para auxiliar sua avaliação, aprimoramento e adequação ao ambiente no qual se insere (ARAÚJO, 1995). Neste sistema, a coleta de documentos inclui a seleção daquilo que é pertinente e a aquisição de informações que subsidiam as atividades de uma instituição ou grupo de pessoas. Durante o tratamento da informação são realizadas a descrição física (DF) e descrição temática (DT), que consolidam o armazenamento e a organização da informação. Já a disseminação é a etapa que transfere a informação aos usuários.

A OI une aquilo que apresenta características iguais e considera a relação entre a obra, com seu conteúdo intelectual, e o documento, com formato variado, ambos com entidades que os descrevem (SVENONIUS, 2001). Ao descrever um documento pretende-se comunicar, por meio de linguagem específica, desenvolvida e aplicada de acordo com um conjunto de regras, determinados atributos (SVENONIUS, 2001). Este é um procedimento fundamental para a recuperação da informação uma vez que identifica e enumera as características do documento. Para isto, são realizadas representações documentais a partir de referências bibliográficas, resumos, descritores, etc., constituindo unidades mais fáceis de manipular que o documento na íntegra, capazes de promover a avaliação do usuário quanto à relevância do conteúdo integral diante das suas necessidades de informação (DIAS, 2001). A descrição pode ser física e temática.

A descrição física é a distinção das características físicas de um documento com base na análise do seu tipo e identificação das informações descritivas (título, autor, etc.) utilizando padrões e normalizações específicas, pois constituem entradas dos sistemas de informações, posteriormente recuperadas e visualizadas pelos usuários, impactando diretamente na consistência dos mesmos. A exemplo de procedimentos de descrição física temos a definição de tipos de documento e a catalogação.

A descrição temática objetiva representar um conteúdo e a profundidade de sua abordagem, para tanto, é possível utilizar informações extraídas do próprio documento ou de instrumentos capazes de sintetizar o assunto. A descrição temática fornece a síntese do conteúdo dos documentos e esta-

[2] Também denominados documentos eletrônicos.

belece categorias para seu armazenamento. A diversidade do universo de informação dos usuários, bem como os objetivos que se pretendem atingir com a descrição, podem tornar suas regras complexas. Assim, a discussão sobre o nível da descrição temática a ser utilizado é comum em sistemas de informação. Svenonius (2001) destaca instrumentos e procedimentos utilizados na descrição temática, como as classificações, cabeçalhos de assuntos e tesauros. Lancaster (1993) completa que um mesmo documento pode ser objeto de diversas descrições como a sua indexação por uma dezena de termos ou seu resumo. Essas descrições são sucessivas e interdependentes, além de utilizarem esforços intelectuais similares.

O foco do estudo são os procedimentos e instrumentos que possibilitam a representação do conhecimento que possam dar subsídios a sua organização no ambiente digital. Esta organização pode ser entendida como a estrutura de informação elaborada a partir da descrição física e descrição temática de um documento, para sua posterior disponibilização. Considerando algumas definições e procedimentos de organização da informação enunciados até o momento, aplicados no ambiente tradicional, é pertinente a investigação sobre as especificidades do ambiente digital que possam influenciar sua organização da informação.

3.1 Organização da informação no ambiente digital

Uma primeira consideração a ser feita sobre o ambiente digital em que se inserem os repositórios estudados é que este dispõe de documentos digitais [3i], entendidos como documentos em formato eletrônico, acessados por meio de uma interface ou tecnologia específica. Um documento digital pode ser instável, dinâmico, e sem fronteiras identificáveis, o que dificulta sua descrição e organização (SVENONIUS, 2001).

Como anteriormente definido, a organização da informação compreende a descrição dos documentos de acordo com seus aspectos físicos e temáticos que, no ambiente digital, pode ser realizada com metadados. De uma forma simples, pode-se definir metadados como dados codificados e estruturados que descrevem as características de um recurso, seja este documento, produto ou serviço. Elementos como autor, título e assunto são exemplos de metadados e podem ser usados para descrever tanto um artigo em um repositório digital, quanto uma *homepage*, base de dados ou qualquer outro recurso no ambiente *Web* (SOUZA *et al.*, 2000). Os metadados podem seguir padrões como o Dublin Core, mencionado como o conjunto mínimo de metadados requerido pelo OAI-PMH, planejado para facilitar a descrição de documentos digitais. Suas principais características são a simplicidade do uso, entendimento semântico universal, escopo internacional e extensibilidade, que permite a adaptação às necessidades adicionais de descrição (DUBLIN CORE METADATA INITIATIVE, 2004).

Ao discorrer sobre o ambiente digital, sua estrutura de informação e o uso de metadados, cabe uma analogia ao ambiente tradicional com informações descritas em formulários seguindo padrões específicos como o Machine Readable Cataloging (MARC), Anglo American Cataloguing Rules (AACR2) e Classificação Decimal Universal (CDU). A principal diferença está no acesso aos conteúdos, uma vez que na biblioteca tradicional uma classificação remete a localização física de um documento impresso disposto na estante, e a classificação da informação no ambiente digital indica *links* para acessar documentos digitais em um repositório digital, por exemplo. Com isso, é observado o importante papel da interface gráfica *Web* ao servir de intermédio para o acesso as coleções digitais destes sistemas.

4. ARQUITETURA DA INFORMAÇÃO

Rosenfeld e Morville (2002), autores de importante trabalho na área, o livro "*Information Architeture for the World Wide Web*", editado em 1998, 2002 e 2006, definem Arquitetura da Informação (AI) como a combinação entre esquemas de organização, nomeação e navegação em um sistema de informação. Estudos sobre essa área ressaltam o favorecimento da organização da informação, com destaque para estes autores e outros como Garrett

[3] Também denominados documentos eletrônicos.

(2003), que discutem importantes elementos de uma AI e a necessidade de assegurar o mapeamento e controle da experiência do usuário diante de um ambiente informacional para compreender suas expectativas.

Os autores já citados, Rosenfeld e Morville (2002), têm contribuições importantes para este contexto e afirmam que AI é o "*design* estrutural de um espaço de informação a fim de facilitar a realização de tarefas e o acesso intuitivo a conteúdos". O acesso e a realização de tarefas no ambiente digital se dão a partir de esquemas de organização ou lógicas de agrupamento das informações que procuram ressaltar suas características comuns.

A visão destes bibliotecários, embora voltada para o desenvolvimento de *sites*, aponta princípios que se aplicam a outras coleções de documentos e mostra sua relação com a Ciência da Informação.

Com a preocupação pautada essencialmente na definição de um conjunto elementos de AI para repositórios digitais institucionais, são explanados os elementos do modelo da arquitetura de Rosenfeld e Morville (2002) que correspondem à organização da informação.

4.1 Modelo de Rosenfeld e Morville

Para facilitar a escolha dos elementos de interação dos usuários com a informação apresentada na *Web*, Rosenfeld e Morville (2002) dividem a AI em quatro grandes sistemas, cada um composto por suas regras e aplicações, embora sejam interdependentes:

- **Sistema de Organização (Organization System):** elementos e instruções para classificar o conteúdo informacional;
- **Sistema de Rotulagem (Labeling System):** define as formas de representação e apresentação da informação a partir da atribuição de rótulos aos conteúdos;
- **Sistema de Navegação (Navegation System):** especifica as rotas de navegação, de movimentação pelo espaço informacional e hipertextual;
- **Sistema de Busca (Search System):** determina as questões que o usuário pode fazer ao sistema e o conjunto de respostas a serem obtidas.

Destes, o sistema de organização e seus elementos são especialmente interessantes para subsidiar a elaboração da estrutura de informação. Os elementos de organização da informação são compostos por esquemas e estruturas de organização, constituídas a partir das características do conteúdo, dos objetivos da instituição e das necessidades de seus usuários (ROSENFELD; MORVILLE, 2002).

Os esquemas de organização definem as características dos itens de informação e a influência destas características para o agrupamento dos mesmos. A partir daí são estabelecidas categorias. Estes esquemas podem ser exatos, ambíguos e híbridos (ROSENFELD; MORVILLE, 2002). Esquemas de organização exatos são facilmente implementados e agrupam informações a partir de características óbvias e mutuamente excludentes. Embora sejam adequados para o caso em que os usuários sabem exatamente o que estão procurando, eles não permitem flexibilidades e ambigüidades. Podem ser alfabéticos, cronológicos e geográficos. Já os esquemas de organização ambíguos são arbitrários, ou seja, dependem de um esforço intelectual individual ou coletivo para elaborar categorias e subcategorias representativas, que sirvam para agrupar conteúdos informacionais com proximidade semântica. Estas categorias são definidas a partir da identificação de características semelhantes e do próprio tratamento da informação que requer acuidade na determinação do assunto e apreciação do contexto. Esquemas ambíguos são significativos quando não se sabe exatamente o que se está procurando, uma vez que classificam o que pode ser encontrado. Podem ser do tipo tópico, orientado à tarefa, ao público ou à metáfora. Os esquemas de organização híbridos utilizam elementos dos esquemas exatos e ambíguos, pois, como afirmam os autores, é difícil optar por apenas um esquema.

As estruturas de organização definem tipos de relações entre o conteúdo e os agrupamentos aos quais eles pertencem, estabelecendo os principais caminhos de navegação apresentados para os usuários. Geralmente as estruturas de *sites* incluem hierarquias, bases de dados relacionais e hipertextos,

cada qual com pontos positivos e negativos. Em alguns casos faz sentido utilizar uma ou outra, mas na maioria das vezes o ideal é utilizar todas as três de maneira complementar (ROSENFELD; MORVILLE, 2002).

Quadro 1 - Elementos de organização da informação segundo Rosenfeld e Morville (2002).

Elementos de organização da informação			
Elementos			**Descrição**
Esquemas	Exatos	Alfabético	Apresenta informações ordenadas alfabeticamente
		Cronológico	Apresenta informações ordenadas por datas ou períodos temporais
		Geográfico	Apresenta informações ordenadas segundo aspectos espaciais
	Ambíguos	Tópico	Organiza informações por assunto
		Orientado a tarefa	Organiza informações segundo processos, funções ou ações que usuários podem executar.
		Orientado ao público	Organiza informações de acordo com os tipos de usuário e seus respectivos interesses.
		Orientado a metáfora	Organiza informações a partir de metáforas conhecidas pelo usuário
	Híbridos		Combina elementos de organização exatos e ambíguos
Estruturas	Hierárquica (*top-down*)		Estrutura taxonômica com classes e subclasses, partindo de assuntos mais gerais para os mais específicos
	Base de dados relacional (*bottom-up*)		Estrutura baseada em registros com descrições de conteúdo, partindo de assuntos específicos para os mais gerais
	Hipertexto		Estrutura que agrupa conteúdos de acordo com nível de semelhança ou critérios pré-definidos, de forma não linear e com componentes interligados por *links*

A organização da informação é um processo complexo e requer tomada de decisões que impactam na localização e uso efetivo dos conteúdos. A variedade de esquemas exatos e ambíguos e as diferentes estruturas de organização devem ser consideradas no contexto de desenvolvimento de um *site* ou sistema de informação com interface gráfica *Web*. A partir da análise das possibilidades, é importante elaborar um projeto que apresente as necessidades e preveja soluções, integrando-as. O *Quadro 1* sintetiza os elementos de organização da informação propostos por Rosenfeld e Morville (2002).

5. METODOLOGIA

A metodologia tem a preocupação de minimizar o caráter subjetivo da análise, pois sistematiza esta ação, pautada na observação e interpretação, e evita impasses no estabelecimento de princípios comuns e extensíveis a outros contextos. Assim, é possível classificar este estudo de acordo com seus objetivos, como uma *pesquisa exploratória*, pois o assunto tratado de maneira ainda incipiente na literatura, facilitando o seu entendimento; e quanto aos procedimentos técnicos, como *pesquisa bibliográfica*, pois utilizou como fonte de consulta, material publicado sobre os temas investigados, servindo de embasamento para análise de RIs e proposta de elementos de AI, estimulando a compreensão da organização da informação nestes. Desta forma, apresentam-se a seguir, as diretrizes metodológicas para o andamento e realização da pesquisa:

1) Revisão bibliográfica: respaldar o estudo e guiar escolhas que melhor atendem seu objetivo;
2) Procedimentos iniciais: adquirida a base teórica e identificados elementos de AI e procedimentos para descrição física e temática, o aprendizado foi aplicado na correlação destes, definição dos parâmetros para a análise da pesquisa (*Figura 1*) e elaboração do roteiro para guiá-la;
3) Análise dos repositórios digitais institucionais: momento para realizar a análise em si, a partir da definição da amostra, análise da ocorrência de elementos de AI correlacionados a procedimentos de descrição física e temática e análise da ocorrência de elementos e procedimentos não previstos.
4) Procedimentos finais: propor um conjunto de elementos de AI, baseados na descrição física e descrição temática, adequados à organização da informação dos repositórios digitais institucionais, com o resultado das etapas anteriores.

Figura 1. – Parâmetros da análise: com a formulação apresentada nesta figura é possível observar que os parâmetros da análise (PA) dos repositórios institucionais pesquisados devem ser compostos de procedimentos para descrição (D) física (DF) e temática (DT), correlacionados aos elementos (E) de arquitetura da informação, ou seja, esquemas (Esq) e estruturas (Est) de organização

6. RESULTADOS

No primeiro momento são estabelecidos parâmetros da análise com a correlação de procedimentos para descrição física e temática e elementos de AI. Esta correlação visa o entendimento sobre quais procedimentos podem embasar a definição de um elemento que orienta a disposição das informações na interface gráfica *Web* do repositório. Desta forma, o *Quadro 2* sintetiza os parâmetros da análise, observando que um elemento pode

ser baseado em mais de um procedimento e que não se extingue a possibilidade de serem encontrados elementos e correlações não previstas durante a análise dos repositórios.

Quadro 2 – Parâmetros de análise dos repositórios digitais institucionais.
(*DF: Descrição Física; DT: Descrição Temática (DT*).

Correlação entre elementos de arquitetura da informação e procedimentos para descrição física e temática	
Elementos	**Procedimentos**
Esquema Exato Alfabético	DF - Catalogação DF - Tipologia documental
Esquema Exato Cronológico	DF - Catalogação DF - Tipologia documental
Esquema Exato Geográfico	DF - Catalogação DF - Tipologia documental
Esquema Ambíguo Tópico	DT - Classificação DT - Indexação
Esquema Ambíguo Orientado a tarefa	DT - Classificação DT - Indexação T
Esquema Ambíguo Orientado ao público	DT - Classificação T DT - Indexação
Esquema Ambíguo Orientado a metáfora	DT - Classificação DT - Indexação
Esquema Híbrido	DF - Catalogação DF - Tipologia documental DT - Classificação DT - Indexação
Estrutura Hierárquica (top-down)	Padrão, Instrumento ou Esquema
Estrutura Base de dados relacional (bottom-up)	Padrão, Instrumento ou Esquema
Estrutura Hipertexto	Padrão, Instrumento ou Esquema

De acordo com as definições propostas a amostra deve ser parte do universo de repositórios digitais institucionais, desenvolvidos diante da perspectiva do acesso livre a informação e contemplando uma pluralidade de assuntos. Os repositórios devem, ainda, disponibilizar arquivos abertos conforme as especificações da OAI e demais manifestações, garantir sua interoperabilidade e primar pelo auto-arquivamento.

Cada vez mais são criados e disponibilizados repositórios de acesso livre, considerando a facilidade de implementação e o incentivo por parte da comunidade científica. Desta forma, existe uma expressiva quantidade de repositórios sem controle e com suas *homepages* e URLs inacessíveis. Diante disto, o universo da pesquisa foi representado pelos repositórios cadastrados no *Registry of Open Access Repositories* (ROAR)[4] considerando, primordialmente, a confiabilidade de suas informações e a organização de seu cadastro, que permite diferentes formas de organização e busca de repositórios, além de estatísticas que contribuem com a delimitação da amostra deste estudo.

O universo de repositórios institucionais de acesso livre, cadastrados no ROAR no dia 16 de fevereiro de 2008, era de 530, com ressalva a este universo para 36 repositórios que não são respaldados pela OAI (pouco mais de 6%), segundo informações obtidas em seus próprios cadastros, e por conseqüência, excluídos do estudo. Aos 494 restantes é estabelecido o filtro referente ao *software* empregado para escolher os dois mais utilizados para representar a amostra, considerando que o *software* tem potencial interferência na organização da informação nos repositórios. Sendo assim, temos DSpace (167 entre 494) e EPrints (146 entre 494), que juntos somam 313 dentre os 494 repositórios institucionais respaldados pela OAI. Foi definido como critério para compor a amostra a seleção de três repositórios que usam o DSpace e três que usam o E-prints, e, dentre esses, um deve ser aquele com maior número de registros, um deve ter a maior intensidade de depósitos e um deve ser o mais antigo em língua portuguesa. A amostra então, favorece a abrangência e relevância dos repositórios que a compõe, contemplando, ainda, os de língua portuguesa, dado a sua importância para as pesquisas sobre repositórios realizadas no Brasil.

Na lista de repositórios apresentada de acordo com estes requisitos, foram averiguados os demais requisitos e características da amostra que incluem, principalmente, o fato dos repositórios serem universitários e provedores de dados e serviços. Sendo assim, a amostra da pesquisa é composta por:

- *DSpace at Cambridge* [5]: RI com maior número de registros que utiliza o DSpace;
- *University of Twente Repository* [6]: RI com maior número de registros que utiliza o E-prints;
- *The University of Adelaide Digital Library* [7]: RI com maior intensidade de depósitos que utiliza o DSpace;
- *Queensland University of Technology - ePrints Archive* [8]: RI com maior intensidade de depósitos que utilizam o E-prints;
- *RepositoriUM* [9]: RI mais antigo em português que utiliza o DSpace;
- *E-Prints Universidad Complutense Madrid* [10]: RI mais antigo em espanhol que utiliza o E-Prints, visto que, não houve sucesso na busca pelo RI mais antigo em português.

[4] http://roar.eprints.org/index.php.
[5] http://www.dspace.cam.ac.uk/
[6] http://doc.utwente.nl/
[7] http://digital.library.adelaide.edu.au/dspace/
[8] http://eprints.qut.edu.au/
[9] https://repositorium.sdum.uminho.pt/
[10] http://www.ucm.es/eprints

7. ANÁLISE DOS RESULTADOS

Existem diferenças na organização da informação dos repositórios diretamente relacionadas às ferramentas utilizadas. O DSpace mostrou-se melhor estruturado, facilitando inclusive a identificação e análise de elementos de AI. Nos repositórios que utilizam o DSpace foi possível discernir uma organização principal evidenciada por suas comunidades, subcomunidades e coleções, que indicam as categorias para o armazenamento dos documentos. Esta organização principal, na maioria das vezes, segue a classificação das unidades de pesquisa da universidade responsável pelo repositório.

As comunidades e subcomunidades dos repositórios que utilizam esta ferramenta são criadas com o propósito de estabelecer hierarquias para agrupar coleções de documentos. As coleções, em uma analogia às bibliotecas tradicionais, devem segmentar o acervo de acordo com seu tipo e a necessidade da instituição, adotando, por exemplo, tipologias documentais, como ocorre no *RepositóriUm*. Contudo, em certos repositórios como o *DSpace at Cambridge*, foi observado que as coleções seguem o entendimento das comunidades e subcomunidades, sendo apenas mais um nível nesta hierarquia.

As comunidades e coleções estão em ordem alfabética, um recurso do próprio DSpace que facilita a visualização de conteúdos específicos. Porém, em alguns casos, isto gera desorganização, pois não é possível, no modelo padrão da ferramenta, agrupar comunidades e coleções semelhantes. São nas coleções de um repositório que os documentos são armazenados e, algumas vezes, estas são definidas apenas para executar este papel. Isto fica claro quando são encontradas coleções únicas para representar todo o conteúdo da coleção, como no *Adelaide Research & Scholarship*, ou quando são observadas coleções com siglas e outros rótulos pouco expressivos que somente reduzem o nome de sua respectiva comunidade, como no *DSpace at Cambridge*.

Não foi encontrado nos repositórios documentação substancial sobre suas comunidades e coleções. O que se pode perceber pela análise, é que por vezes estas são criadas de maneira aleatória, sem preocupação com sua organização, como no *DSpace at Cambridge*. Isso, sem dúvida, reflete no grande número de registros deste repositório. Quando as comunidades não seguem regras particulares, são criadas apenas como pontos de acesso, independente de seu nível de especificidade, para atender o interesse de grupos de usuários. Como conseqüência é necessário a permissão de acesso às coleções, fazendo com que o repositório seja utilizado de maneira limitada, como interface para publicação e comunicação entre os membros da própria universidade, sem facilitar o acesso de usuários externos.

Além da organização principal, o DSpace apresenta organizações secundárias que permitem a visualização dos documentos segundo pontos de vista distintos. Estas organizações são definidas a partir das informações preenchidas em formulários de metadados no padrão Dublin Core, ordenadas alfabética ou numericamente. Desta forma, é possível percorrer listas de autores, títulos, assuntos, datas, etc., em ordens crescentes e decrescentes. A qualidade do preenchimento dos metadados é fundamental para a qualidade destas listas e, por exemplo, quando o autor não é informado ou quando o assunto é indicado de maneira errada, estas opções de organização da informação ficam comprometidas. Tais observações são evidenciadas quando se encontra uma quantidade muito pequena de autores listados[11] diante de um grande número de registros.

A estrutura de organização observada nos esquemas da organização principal do DSpace geralmente é hierárquica e tem relação com as categorias deste esquema. A estrutura de organização observada nos esquemas da organização secundária é hipertextual, seguindo o nível de semelhança pré-definido por estes esquemas.

[11] São encontrados 1278 autores listados e 189950 títulos. Título é um metadado obrigatório e autor não. Na lista de títulos observa a ausência do preenchimento de autores (*DSpace at Cambridge*).

Nos repositórios que utilizam a ferramenta E-prints não se distinguem organizações principais e secundárias. São identificados esquemas de organização que apresentam os documentos sem necessariamente apresentar todos estes, ou seja, diferentes esquemas em um mesmo repositório organizam diferentes grupos de documentos, enquanto alguns outros esquemas organizam todos os documentos do repositório. Assim, no *Queensland ePrints Archive*, os esquemas ambíguos orientado ao público que organizam institutos (*Institutes*) e centros de pesquisa (*Research Centres*) não são opções de organização, pois os documentos que estão no institutos (*Institutes*) não se repetem no centros de pesquisa (*Research Centres*). Já o esquema exato alfabético que organiza autores (*Person*) é uma opção de organização da informação e apresenta todos os documentos do repositório segundo um ponto de vista.

A estrutura de organização observada nos esquemas do E-prints geralmente é hipertextual e segue o nível de semelhança pré-definido por estes esquemas.

O esquema exato geográfico, esquema ambíguo orientado a tarefa e esquema ambíguo orientado a metáfora, não foram identificados em nenhum dos repositórios da amostra, e a estrutura base de dados relacional também não. A esses elementos de AI não cabem observações sobre sua relevância para os repositórios.

Quanto aos procedimentos de descrição física e temática que podem embasar os elementos de AI, a catalogação dos documentos, realizada em formulários de metadados, é evidentemente um procedimento importante para a definição de esquemas de organização. Embora nos repositórios desenvolvidos com o E-prints não tenha sido possível identificar com exatidão os metadados utilizados, indicando-se apenas seu campo. Na maioria dos repositórios analisados, as informações fornecidas durante a descrição dos documentos indicam a qualificação de um documento em um esquema. Em outras palavras, ao catalogar um documento e descrever no metadado específico seu autor, é fornecida uma informação que o caracteriza e posiciona em um esquema que destaca o autor como ponto de acesso em sua organização. O procedimento de descrição da tipologia documental, que segue critérios próprios de cada repositório, foi verificado e representa um recurso importante na definição de esquemas e estruturas de organização da informação. Os procedimentos de classificação e indexação são observados como o foco de organizações que destacam o assunto dos documentos. Contudo, a indexação não costuma se basear em instrumentos específicos, sendo, em geral, realizada em linguagem natural.

O uso de padrão, instrumento ou esquema de descrição na estrutura de organização, não foi comumente observado. Com exceção do esquema de descrição, os outros dois não foram identificados em nenhum repositório. O que se pode observar é que, como a estrutura é visualizada em um esquema, geralmente esta segue a sua organização.

Os esquemas exatos são comumente identificados nos repositório e isto se deve principalmente, à facilidade de implementá-los. Apenas um dos repositórios não utiliza o esquema exato alfabético definido a partir da catalogação em formulários de metadados. Este, o *University of Twente Repository*, apresenta um esquema exato a partir da tipologia documental e é notável a dificuldade de percorrer listas de tipos de documento, como teses e dissertações, sem ordenação alfabética ou cronológica.

O procedimento de definição da tipologia documental, embora identificados em apenas dois repositórios, o *University of Twente Repository* e o *RepositoriUm*, é útil quando aplicado em esquemas exatos alfabéticos e cronológicos, oferecendo uma ordenação para a tipologia de documentos que pode ser estabelecida pela instituição responsável pelo repositório, tal como observado no *RepositoriUm*.

O esquema exato cronológico, também comum a todos os repositórios analisados, é utilizado de acordo com a data de publicação, *copyright*, ou data de aceite no repositório, valendo-se de informações fornecidas na catalogação dos documentos. A ordem estabelecida, quando não é permitido variar, como em todos aqueles que utilizam o DSpace, é do ano mais recente para o mais antigo, priorizando o acesso aos documentos recentes, mais

procurados em instituições que realizam pesquisa científica, ressaltando, inclusive, a importância deste esquema. O *Queensland ePrints Archive* é o único que permite visualizar seus documentos apenas ordenado dos mais antigos para os mais recentes.

O esquema ambíguo tópico só foi observado em metade dos repositórios analisados. Este tipo de organização, que faz referência ao conteúdo dos documentos armazenados, talvez não seja muito utilizado devido à subjetividade de sua definição. Como os repositórios pesquisados favorecem o auto-arquivamento e sua filosofia concede certa liberdade de criação, isto os distancia da exigência de um profissional de informação ou mesmo da adoção de padrões e regras consistentes que contribuam com a qualidade da descrição temática de seus documentos. Esta observação, embora incentive que os próprios autores sejam capazes de descrever e indicar o assunto dos documentos pode comprometer a qualidade de um esquema de organização que tenha por base seus assuntos ou temas.

Como observado, um esquema ambíguo tópico pode se basear em procedimentos de classificação e indexação. Quando o esquema de organização é baseado na classificação, é comum que esta siga um esquema ou padrão de descrição específico, contribuindo com o entendimento consensual de seus assuntos ou temas, tornando-o de grande utilidade. Vale ressaltar a importância de que a classificação utilizada seja evidente, como observado apenas no *Queensland ePrints Archive* [12]. Quando um esquema ambíguo tópico se baseia no procedimento de indexação, como no *DSpace at Cambridge* e no *RepositoriUm*, o uso de tesauros ou taxonomias é uma sugestão e não uma regra. Estes esquemas não são homogêneos devido à diversidade de descrições de assunto que podem ocorrer. Caso um mesmo esquema ambíguo tópico se baseie em procedimentos tanto de classificação quanto de indexação, como no *RepositoriUm*, pode haver divergências em sua organização, ora apresentando notações de esquemas de classificação, ora apresentando descritores muito específicos, tornando-o desestruturado. O esquema ambíguo orientado ao público, segundo classificação das unidades de pesquisa da universidade, é o mais observado nos repositórios pesquisados, apresentando-se oito vezes. O interesse em organizar os documentos segundo institutos, escolas, faculdades, departamentos e centros de pesquisa, por exemplo, tem relação com a proposta dos repositórios digitais institucionais de dar visibilidade a sua produção intelectual. Porém, uma ressalva é feita para a necessidade de conhecer a estrutura orgânica da universidade que mantém o repositório para facilitar a utilização deste esquema.

As estruturas de organização influenciam a navegação nos esquemas e nos repositórios pesquisados que utilizam o DSpace, estas foram observadas na organização principal. No E-prints, que não distingue uma organização principal, as estruturas foram verificadas em todos os esquemas de organização observados causando, inclusive, certa repetitividade no resultado. De modo geral, sempre é identificada uma estrutura de hipertexto estabelecida de acordo com o nível de semelhança pré-definido nos esquemas de organização utilizados. A estrutura hierárquica é muito útil no entendimento dos níveis de assuntos dos documentos do repositório, como o *Queensland ePrints Archive* que utiliza esta estrutura em seu esquema ambíguo tópico, e também para hierarquizar as unidades de pesquisa da universidade representadas no repositório, mais comum em esquemas ambíguos orientado ao público, como visto no *RepositoriUm*.

8. CONJUNTO DE ELEMENTOS DE AI, BASEADOS NA DESCRIÇÃO FÍSICA E TEMÁTICA, CONSIDERADOS APROPRIADOS

Diante do que foi analisado e discorrido, os esquemas e estruturas de organização, com seus procedimentos de descrição, considerados apropriados aos repositórios digitais institucionais estão nos quadros a seguir.

Todos estes esquemas e procedimentos foram observados na pesquisa e à sua utilização cabem as ressalvas discutidas. O esquema exato baseado na tipologia documental, verificado no *University of Twente Repository*, não foi considerado adequado, pois o esquema exato alfabético baseado na tipologia documental, também verificado, é capaz de substituí-lo, agregando a ordenação alfabética, um recurso considerado favorável.

[12] Este repositório destaca, ao apresentar o esquema ambíguo tópico baseado na classificação, qual esquema de descrição utiliza (*Qeensland ePrints Archive*)

Considerando que o intuito da pesquisa é contribuir com a organização da informação nos repositórios digitais institucionais e para tanto foram observadas as áreas de seus sites que dão acesso aos documentos disponibilizados, no que diz respeito às estruturas de organização, o que se pode observar, como mencionado anteriormente, é que estas são relacionadas com os esquemas de organização.

Quadro 3 – Esquemas de organização da informação considerados apropriados.

Esquema	**Procedimento**	**Sugestão de utilização**
Exato Alfabético	DF-Catalogação	Segundo autor, autor entidade, editor, título, etc.
Exato Alfabético	DF-Tipologia documental	Segundo artigo, anais, dissertação, tese, livro, capítulo de livro, resumo de livro, relatório técnico/científico, etc.
Exato Cronológico	DF-Catalogação	Segundo data de publicação, data de *copyright,* data de defesa de tese, data de submissão, etc.
Ambíguo Tópico	DT-Classificação	Segundo esquemas de classificação como a CDU, Research Fields, Courses and Disciplines /Australian Bureau of Statistics (RFCD/ABS), Classification-JEL, etc.
Ambíguo Tópico	DT-Indexação	Segundo Indexação em linguagem natural e em linguagem controlada.
Ambíguo Orientado ao público	DT-Classificação	Segundo divisões administrativas, institutos, departamentos, faculdades, centros de pesquisa e demais unidades de pesquisa da universidade.

Quadro 4 – Estruturas de organização da informação consideradas apropriadas.

Estrutura	Esquema	Sugestão de utilização
Hipertexto	Esquema exato	Segundo o nível de semelhança pré-definido nos esquemas utilizados, baseados na catalogação e tipologia documental
Hierárquica	Esquema ambíguo	Segundo o nível de semelhança pré-definido nos esquemas utilizados, baseados na classificação e indexação

Sendo assim, uma estrutura pode ser baseada em um procedimento de descrição, mas este procedimento de descrição é aquele utilizado na definição do esquema em que se observa a estrutura. Como exemplo, o *Queensland ePrints Archive* adota uma estrutura hierárquica baseada em um esquema de descrição, o RFCD/ABS. Essa estrutura foi observada em seu esquema ambíguo tópico baseado na classificação. Isto significa que o esquema de descrição (RFCD/ABS) que embasa a estrutura é exatamente aquele utilizado no procedimento de classificação que define o esquema utilizado. Diante disto, é possível concluir que há relação entre o procedimento de descrição utilizado no esquema de organização e sua estrutura.

Sobre o uso de estruturas de organização em repositórios digitais institucionais, de acordo com as considerações descritas, quando se observa um esquema exato, a estrutura considerada apropriada é a de hipertexto, segundo o nível de semelhança observado na utilização de procedimentos de catalogação e tipologia documental. Quando se observa um esquema ambíguo, a estrutura considerada apropriada é a hierárquica, segundo o nível de semelhança observado na utilização de procedimentos de classificação e indexação.

9. CONSIDERAÇÕES FINAIS DO CAPÍTULO 7

Como foi discutido, o crescimento da disponibilização de documentos em meio digital é uma realidade e os repositórios digitais são um importante recurso para este fim. Com a facilidade de desenvolvê-los, diversas universidades recorrem aos repositórios digitais para comunicar pesquisas realizadas em seus domínios e para disponibilizar arquivos abertos de modo geral. Neste estudo, os repositórios digitais institucionais analisados são compreendidos como sistemas de informação com interface gráfica *Web,* mantidos por universidades, que armazenam documentos digitais em categorias originadas a partir da sua descrição física e temática. Foi observado que a descrição ocorre em formulários de metadados preenchidos pelos próprios autores no momento da submissão de seus documentos digitais e nem sempre segue diretrizes específicas, tornando essencial a preocupação com a organização da informação. Concluindo, os elementos de AI, correlacionados a procedimentos de descrição, analisados e sistematizados, orientam e facilitam a organização da informação em repositórios digitais institucionais, mantidos por universidades. Vale salientar que nos repositórios digitais pesquisados foram observados *menus* e opções da sua interface gráfica *Web* que apresentam *links* aos documentos digitais, outras áreas dos seus *sites* podem apresentar diferentes elementos e/ou procedimentos, mas não são o foco deste estudo.

10. REFERÊNCIAS DO CAPÍTULO 7

ARAÚJO, Vânia Maria Rodrigues Hermes de. Sistemas de informação: nova abordagem teórico-conceitual. *Ci. Inf.*, Brasília, v. 24, n. 1. 1995.

BERGMAN, Michael K. The deep web: surface hidden value. *Journal of Electronic Publishing*, v. 7, n. 1, aug. 2001. Disponível em: http://www.press.umich.edu/jep/07-01/bergman.html. Acesso em: jun. 2002.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*. 2001. Disponível em: http://www.scientificamerican.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21. Acesso em: set. 2005.

COSTA, Sely M. S. Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o acesso aberto à informação científica. *Ci. Inf.,* Brasília, v. 35, n. 2, p. 39-50, maio/ago 2006.

CROW, Raym. *The case for institutional repositories*: a SPARC position paper, 2002. 37p. Disponível em: <www.arl.org/sparc/bm~doc/ release_102.pdf> Acesso em: abr. 2007.

DIAS, Eduardo, Wense. Contexto Digital e Tratamento da Informação. *DataGramaZero*, Rio de Janeiro, v. 2, n. 5, out. 2001.

DUBLIN CORE Metadata Initiative. Disponível em: http://dublincore.org/index.shtml. Acesso em: jan. 2008.

GARRETTT, Jesse James. *The Elements of user Experience*: User-Centers Design for the Web. 2. ed. Indianápolis (Indiana), 2003, 189 p. Disponível em: http://www.jjg.net/ia/elements.pdf. Acesso em: out. 2007.

LANCASTER, F. W. *Indexação e resumos:* teoria e prática. Brasília: Briquet de Lemos, 1993. 347p.

LE COADIC, Yves-François. *A ciência da informação*. Brasília: Briquet de Lemos, 2004. 124 p.

ROSENFELD, L.; MORVILLE, P. *Information Architecture for the Word Wide Web*. 2. ed. Sebastopol: O'Reilly, 2002.

SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. *Ci. Inf.*, v. 29, n. 1, p. 93-102, abr. 2000.

SVENONIUS, Elaine. *The intellectual foundation of information organization*. Cambridge: MIT Press, 2001.

III. SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO

Capítulo 8. – Sistemas de organização do conhecimento: Antigas e novas linguagens – Marisa BRÄSCHER e Eliana CARLAN, *p. 147-176*

Capítulo 9. – Categorização linguística como esteio da organização do conhecimento – Fabio José Dantas de MELO, *p. 177-182*

Capítulo 8

Sistemas de organização do conhecimento: Antigas e novas linguagens

Marisa Bräscher[1] e Eliana Carlan[2]

SUMÁRIO DO CAPÍTULO 8

Resumo do Capítulo 8, p. 148
Como citar o Capítulo 8, p. 148

1. INTRODUÇÃO DO CAPÍTULO 8, *p. 149*
2 . O QUE SÃO E PARA QUE SERVEM OS SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO? *p. 149*
3. QUAIS SÃO OS ELEMENTOS QUE COMPÕEM OS SOCs? *p. 155*
4 . O QUE SÃO CLASSIFICAÇÕES, TESAUROS, TAXONOMIAS E ONTOLOGIAS? *p. 157*
  4.1 Classificações, *p. 157*
  4.2 Tesauros, *p. 158*
  4.3 Taxonomias, *p. 158*
  4.4 Ontologias, *p. 160*
5 . QUAIS SÃO AS SEMELHANÇAS E AS DIFERENÇAS ENTRE CLASSIFICAÇÕES, TESAUROS, TAXONOMIAS E ONTOLOGIAS? *p. 161*
  5.1 Princípios de classificação, *p. 161*
  5.2 Rótulos ou etiquetas, *p. 168*
  5.3 Padronização terminológica, *p. 168*
  5.4 Especificação do conceito, *p. 169*
  5.5 Relacionamento conceitual, *p. 171*
6. CONSIDERAÇÕES FINAIS DO CAPÍTULO 8, *p. 173*
7. REFERÊNCIAS DO CAPÍTULO 8, *p. 174*

[1] Doutora em Ciência da Informação. Professora Adjunta. Programa de Pós-Graduação em Ciência da Informação (PPGCInf) da Faculdade de Ciência da Informação da Universidade de Brasília (UnB/FCI). CV Lattes: http://lattes.cnpq.br/8951909489273046. E-mail: marisab@unb.br.
[2] Mestre em Ciência da Informação. Consultora OPAS/OMS - área de disseminação da informação. E-mail: eliana.carlan@gmail.com

Resumo do Capítulo 8

O capítulo apresenta os Sistemas de Organização do Conhecimento (SOCs), no contexto da Ciência da Informação, mais especificamente na área de Organização e Representação do Conhecimento. Discute conceitos e funções dos SOC na visão dos principais autores sobre o assunto. Os conceitos, termos e relações semânticas são apontados como os elementos que compõem esses sistemas. Dos diversos tipos de SOC, as classificações, tesauros, taxonomias e ontologias (CTTOs) foram contempladas nesse estudo com suas descrições. Destacam-se as semelhanças e diferenças entre esses sistemas, por meio de exemplos, os quais ilustram as peculiaridades de cada um.

Palavras-chave: Sistemas de Organização do Conhecimento. Classificação. Tesauro. Taxonomia. Ontologia.

Abstract

Knowledge Organization Systems: Old and New Languages
This chapter presents an overview of the Knowledge Organization Systems (KOS), in the context of Information Science, more specifically in the field of Knowledge Organization and Representation. The concepts and functions of the KOS are discussed and the opinion of the most known authors about the subject. The concepts, terms and the semantic relations are pointed as fundamental elements of these systems. From the different types of the KOS, classification, thesaurus, taxonomy and ontology (CTTO) were contemplated on this research with their description. The similarities and the differences among these systems were pointed trough examples and peculiarities of each system were analyzed.

Keywords: Knowledge organization systems; Classification; Thesaurus; Taxonomy; Ontology

Como citar o Capítulo 8

BRÄSCHER, Marisa; CARLAN, Eliana. Sistemas de organização do conhecimento: antigas e novas linguagens. *In:* Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento.* Brasília DF: IBICT, 2010, 335 p. ISBN: 978-85-7013-072-3. Capíitulo 8, p. 147-176 Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

1. INTRODUÇÃO DO CAPÍTULO 8

Sistemas de organização do conhecimento (SOCs)[3] são objetos de interesse particular para a Ciência da Informação, uma vez que cumprem importante papel de padronização da terminologia adotada para organização e recuperação de informações, ao delimitar o uso de termos e definir conceitos e relações de alguma área do conhecimento, de forma compartilhada e consensual. Esses sistemas são fruto do processo de organização do conhecimento.

Entendemos a organização do conhecimento (OC) como processo de modelagem que visa construir representações do conhecimento (RC). A OC tem por base a análise do conceito e de suas características para o estabelecimento da posição que cada conceito ocupa num determinado domínio, bem como das suas relações com os demais conceitos que compõem um sistema conceitual. (BRÄSCHER; CAFÉ, 2008)

Linguística, Filosofia, Psicologia, Ciência da Informação e Inteligência Artificial são algumas das áreas que se ocupam dos diferentes aspectos da representação do conhecimento, que englobam desde os estudos da cognição humana até os modelos de representação construídos artificialmente para uso em sistemas especialistas. A diversidade de abordagens leva a diferentes conceitos acerca da representação do conhecimento. Nesse trabalho abordamos a RC num escopo restrito às representações construídas pelo homem. Consideramos, portanto, que esses tipos de representações são modelos de abstração do mundo real construídos para determinada finalidade e caracterizamos os Sistemas de Organização do Conhecimento como tipos de representação do conhecimento.

Neste capítulo apresentamos o conceito e funções dos SOCs e estabelecemos algumas características desses sistemas, com o intuito de contribuir para o entendimento dos pontos comuns e daqueles que diferenciam os tipos de sistemas. Centramos nossa atenção nas classificações, tesauros, taxonomias e ontologias, por sua importância na área de organização e recuperação da informação.

2. O QUE SÃO E PARA QUE SERVEM OS SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO?

Antes de abordar especificamente os SOC, é necessário introduzir conceitos relacionados à organização do conhecimento e à organização da informação. Em trabalho apresentado no X Encontro Nacional de Pesquisa em Ciência da Informação, Bräscher e Café (2008) discorreram sobre o tema, por conseguinte, limitar-nos-emos a destacar pontos que auxiliam a situar os SOC no contexto desses processos de organização.

Dahlberg (1993, p. 211) define organização do conhecimento como "a ciência que estrutura e organiza sistematicamente unidades do conhecimento (conceitos) segundo seus elementos de conhecimento (características) inerentes e a aplicação desses conceitos e classes de conceitos ordenados a objetos/assuntos." Em trabalho mais recente, a autora distingue duas aplicações para a organização do conhecimento: *"a)* a construção de sistemas conceituais; e *b)* a correlação ou mapeamento de unidades desse sistema conceitual com objetos da realidade." (DAHLBERG, 2006, p. 12)

A fim de evitar confusões terminológicas advindas do uso do termo para se referir às duas aplicações, preferimos usar o termo organização do conhecimento apenas no sentido da primeira aplicação sugerida por Dahlberg. Para nos referirmos à segunda aplicação, adotamos o termo organização da informação, conforme proposto por Bräscher e Café (2008):

[3] Tradução para o português do termo *knowledge organization systems (KOS)* proposto em 1998 pelo *Networked Knowledge Organization Systems Working Group*, na Conferência ACM Digital Libraries 98. (Hodge, 2000)

> Em nossa visão, temos dois tipos distintos de processos de organização, um que se aplica às ocorrências individuais de objetos informacionais - o processo de organização da informação, e outro que se aplica a unidades do pensamento (conceitos) - o processo de organização do conhecimento. A OI compreende, também, a organização de um conjunto de objetos informacionais para arranjá-los sistematicamente em coleções, neste caso, temos a organização da informação em bibliotecas, museus, arquivos, tanto tradicionais quanto eletrônicos. A organização do conhecimento, por sua vez, visa à construção de modelos de mundo que se constituem em abstrações da realidade." (BRÄSCHER; CAFÉ, 2008, P. 6)

Considerando a primeira aplicação sugerida por Dahlberg (2006), os SOC são tipos de sistemas conceituais, ou, como afirmamos na introdução deste trabalho, tipos de representações do conhecimento, frutos do processo de organização do conhecimento.

Conforme a segunda aplicação destacada por Dahlberg (2006) os SOC são aplicados para mapear objetos informacionais, ou seja, para representar os assuntos dos documentos num sistema de informação. Nesse contexto, os SOC são instrumentos usados nos processos de classificação e indexação. A qualidade obtida na recuperação da informação dependrá substancialmente desses instrumentos. Portanto, os padrões de organização devem ser definidos desde a concepção do sistema para permitir que a informação seja encontrada posteriormente.

Num sistema de informação, a qualidade obtida na recuperação da informação depende substancialmente dos procedimentos e instrumentos utilizados para organização da informação. Os padrões de organização devem, portanto, ser definidos desde a concepção do sistema para permitir que a informação seja encontrada posteriormente.

Na década de 60, Montgomery (1969) já afirmava que o módulo que controla a representação do conteúdo é o coração de um sistema de recuperação da informação. Esta ideia também é reforçada por Salton (1990) quando afirma que qualquer modelo avançado de recuperação da informação deve lidar com o problema da análise da linguagem, uma vez que o conteúdo dos textos e dos documentos necessariamente controla as atividades de recuperação.

Zeng (2006) considera a busca por assunto a tarefa mais importante do usuário, como evidenciam os recursos utilizados por ferramentas de busca, diretórios, bases de dados e sites para prover acesso aos conteúdos.

Vickery (1986) lembra que a representação do conhecimento é objeto antigo de pesquisa na área de informação e retrata a evolução nessa área.

> A representação do conhecimento é uma questão que preocupa a Documentação desde sua origem. O problema agora é relevante em muitas outras situações além dos documentos e índices. A estrutura de registros e arquivos de bases de dados; a estrutura de dados nos programas de computador; a estrutura sintática e semântica da linguagem natural; a representação do conhecimento em inteligência artificial; os modelos de memória humana: em todos esses campos é necessário decidir como o conhecimento pode ser representado de forma que estas representações possam ser manipuladas. (VICKERY, 1986, p. 145)

Diante do destaque atribuído por autores renomados da área de Ciência da Informação aos aspectos de representação e recuperação de conteúdos, concluímos que os SOCs cumprem importante papel nesse contexto. Para Hodge (2000), os SOCs são "mecanismos de organização da informação", e, como prefere Vickery (2008), "instrumentos complementares que ajudam o usuário a encontrar seu caminho no texto". Hodge (2000) chega a qualificar esses sistemas como "o coração de toda biblioteca, museu e arquivo".

Se por um lado há concordância dos autores quanto à importância dos SOCs para a organização e recuperação de informações, parece não haver tanto consenso quanto à abrangência do conceito de SOC e aos tipos de sistemas considerados. Vickery (2008) indica índices e sumários de livros como

as formas mais simples de SOCs. No entanto, o autor ressalta que, com o tempo, esses sistemas tornaram-se mais complexos e assumiram funções mais amplas, havendo, em consequência, o surgimento de novas denominações, como taxonomias, categorizações, tesauros ou ontologias. Os SOCs são esquemas que organizam, gerenciam e recuperam informação e a base dos modernos SOCs são termos e indicações de relacionamentos entre eles. Ainda para o autor supracitado, essa definição abrange dicionários, glossários, redes semânticas, *slots* e *frames* (representações do conhecimento usadas pela inteligência artificial), mapas conceituais e listas de termos.

Segundo perspectiva histórica, Vickery (2008) reúne os SOCs em quatro grupos:

1) Era da pré-coordenação: os SOCs eram estruturas estáticas e atendiam às necessidades dos sistemas manuais de organização e recuperação da informação, como índices e catálogos. Incluem-se aqui as listas de cabeçalhos de assunto e as classificações.
2) Era da pós-coordenação: os SOCs tornam-se mais dinâmicos e possibilitam que cada um de seus elementos (termos) sejam manipulados de forma independente para representar os assuntos de cada documento. Exemplos de SOCs dessa era são vocabulários controlados (listas de termos autorizados para uso na indexação e recuperação da informação) e tesauros.
3) Era da Internet: os SOCs que se destacam são as classificações hierárquicas que orientam o usuário na escolha do termo que melhor expressa sua questão de busca; os elos estabelecidos por meio de URL entre itens da Web e os índices das ferramentas de busca, compostos de palavras extraídas dos conteúdos dos objetos informacionais.
4) Era da Web Semântica: os SOCs dessa era diferenciam-se dos demais por serem projetados para uso por agentes inteligentes. O principal exemplo são as ontologias.

Pode-se ou não concordar com a classificação sugerida por Vickery (2008). No entanto, observamos que a característica que o autor privilegia em sua proposta é a função de organização e recuperação de informações que os SOCs cumprem em diferentes momentos que marcam a evolução dos sistemas de informação. Nesse aspecto, concordamos com a visão do autor e destacamos que a Ciência da Informação deve se ocupar principalmente dos SOCs que cumprem essa função.

A classificação de tipos de SOC proposta por Hodge (2000) é utilizada por vários autores (SHIRI, MOLBERG, 2005; MCCULLOCH, MACGREGOR 2008; HJORLAND, 2008; ZENG, 2006) e também adotada pelo Networked Knowledge Organization Systems and Services – NKOS[4].

Segundo Hodge (2000),

> Sistemas de Organização do Conhecimento englobam todos os tipos de instrumentos usados para organizar a informação e promover o gerenciamento do conhecimento. Incluem os esquemas de classificação que organizam materiais em nível geral (como livros em estantes), cabeçalhos de assunto que provêm acesso mais detalhado e listas de autoridade que controlam versões variantes de chaves de acesso à informação (nomes geográficos e nomes de pessoas). Incluem, ainda, esquemas menos tradicionais, tais como redes semânticas e ontologias. (HODGE, 2000)

Em sua classificação, Hodge (2000) agrupa os SOCs em três categorias:

[4] http://nkos.slis.kent.edu/KOS_taxonomy.html

1) Listas de termos: englobam listas de termos geralmente acompanhados de suas definições. Compreendem:
   - Lista de autoridades: listas de termos que controlam as variações de nomes para entidades, como nomes de países, indivíduos ou instituições. Ex.:http://www.senado.gov.br/sf/Biblioteca/servicos/VCB_AUTR.asp
   - Glossários: lista de termos com definições, geralmente de um assunto ou domínio específico. Ex.: http://www.bcb.gov.br/?glossario
   - Dicionários: lista alfabética de palavras e suas definições. Seu escopo é mais geral que o dos glossários. Ex.:http://www.dicionariodoaurelio.com/
   - *Gazetteers*: dicionário de nomes de lugares e acidentes geográficos, tais como cidades, rios, vulcões. Quando georreferenciados, os *gazetteers* apresentam as coordenadas para a localização de lugares na superfície da Terra. Ex.: http://www.world-gazetteer.com/

2) Classificações e categorias: reúnem SOCs que enfatizam a criação de classes de assuntos. São eles:
   - Cabeçalhos de assunto: conjunto de termos controlados que representam os assuntos de uma coleção. Ex.:
   - http://memory.loc.gov/hlas/portugues/assuntos.html
   - Esquemas de classificação, taxonomias e esquemas de categorização: esquemas utilizados para agrupar entidades em classes mais gerais. Os esquemas de categorização são geralmente usados para reunir os termos de tesauros em tópicos. As taxonomias são usadas em modelos orientados a objeto e em sistemas de gestão do conhecimento, para indicar grupos de objetos baseados em características particulares. Ex.: http://www.google.com.br/dirhp

3) Listas de relacionamentos: agrupam SOCs que privilegiam a conexão entre termos e conceitos. Englobam:
   - Tesauros: conjunto de termos que representam conceitos e as relações de equivalência, hierárquicas e associativas que se estabelecem entre eles. Ex.: http://www.cnfcp.gov.br/tesauro/
   - Redes semânticas: estrutura de conceitos e termos em forma de rede ou teia, os conceitos são nós e os relacionamentos expandem-se a partir dos nós. Ex.: http://wordnetweb.princeton.edu/
   - Ontologia: estrutura de conceitos e representação dos relacionamentos complexos entre eles, incluindo regras de inferência e axiomas. Ex.: http://wiki.musicontology.com/index.php/Main_Page

Zeng e Salaba (2005) propõem um gráfico (*Fig. 1*) que sintetiza a classificação proposta por Hodge (2000) e ressaltam as características que diferenciam os SOC quanto ao nível de complexidade da estrutura e quanto à maior ou menor aproximação com a linguagem natural.

Zeng (2008) apresenta proposta mais recente de visão geral dos SOCs, segundo sua estrutura e função. No entanto, preferimos adotar a proposta anterior, de 2005, por considerá-la mais coerente em relação às características dos SOCs destacadas no gráfico, onde se torna claro que as listas de termos são menos estruturadas e mais próximas da linguagem natural do que os tesauros e ontologias, que se encontram no outro extremo do gráfico.

Apesar da ampla aceitação da classificação de Hodge (2000), não concordarmos plenamente com a abrangência atribuída ao conceito de SOC, que reúne diferentes tipos de instrumentos sob essa mesma denominação. Predomina na literatura a visão de SOC como "um amplo conjunto de esquemas para organizar a informação e facilitar a descrição e descoberta de recursos" [de informação]. (Shiri e Molberg, 2005, p.605).

No entanto, neste capítulo, preferimos seguir a visão de autores que definem os SOCs num sentido mais restrito, como Broughton et al. (2004, p.143), que os consideram, no "sentido específico da palavra, como ferramentas semânticas que consistem de palavras conceitos e relações semânticas, definidas e selecionadas". Apesar da posição dos autores, preferimos adotar "termos" em vez de "palavras".

Para chegar a uma definição que melhor represente os interesses da Ciência da Informação em relação ao estudo e ao desenvolvimento de SOCs, tomamos por princípio que esses sistemas são ferramentas semânticas, segundo a definição proposta por Hjorland (2008): "os SOCs são ferramentas

que apresentam a interpretação organizada de estruturas do conhecimento, também chamadas de ferramentas semânticas". Segundo o mesmo autor, essas ferramentas contemplam, essencialmente, conceitos e suas relações semânticas, além de informarem o significado dos termos ou símbolos que empregam. Pressupomos, ainda, que, para a Ciência da Informação, os SOCs devem cumprir a função de instrumentos de organização e recuperação da informação.

Figura 1. – Tipos de Sistemas de Organização do Conhecimento
(Fonte: Zeng & Salaba: FRBR Workshop, OCLC 2005)

Diante do exposto, entendemos que, para a Ciência da Informação, os SOCs são representações de domínios do conhecimento que delimitam o significado de termos no contexto desses domínios, estabelecem relações conceituais que auxiliam a posicionar um conceito no sistema conceitual e são utilizadas como instrumentos de organização e recuperação da informação.

Levando em conta essa delimitação conceitual, discutimos a inclusão, entre esses sistemas, das listas de autoridade, dicionários, glossários e *gazetteers*. Apesar de serem instrumentos de organização e recuperação da informação, as listas de autoridade não contemplam a representação em nível conceitual. São instrumentos que têm por essência a padronização de vocabulário e não constituem representações do conhecimento. Os dicionários, glossários e *gazetteers* delimitam conceitos e apresentam alguns tipos de relações conceituais, como as relações de equivalência e associativas. No entanto, não são instrumentos elaborados com os objetivos de organização e recuperação de informações.

Acreditamos ser importante delimitarmos os tipos de SOC àqueles que possuem essas características comuns, em função da dificuldade em traçar princípios teóricos e metodológicos que abranjam uma ampla diversidade de sistemas. Hjorland (2008) chama a atenção para isso e lembra o fato de uma pessoa lhe haver dito que era especialista em alguns tipos de SOC, como, por exemplo, tesauros, mas não em outros, como enciclopédias. Para o autor, essa afirmação leva ao questionamento acerca da abrangência do conceito de SOC: "esse conceito geral é ruim ou está definido de maneira muito geral". Hjorland (2008) nos coloca, ainda, a seguinte questão: "os métodos utilizados para elaboração de tesauros são úteis para a elaboração de enciclopédias?"

Não aprofundaremos essas discussões neste capítulo, deixaremos essa árdua tarefa para outra oportunidade. Apenas levantamos esses questionamentos para justificar a delimitação do conjunto de SOCs que abordamos neste trabalho. Para tanto, recorremos ao escopo definido pelo SKOS - Simple Knowledge Organization System, que se ocupa do desenvolvimento de especificações e padrões para apoiar o uso de sistemas de organização do conhecimento, no contexto da Web Semântica. "Sistema de Organização do Conhecimento é um conjunto de elementos, geralmente estruturado e controlado, que pode ser usado para descrever (indexar) objetos, navegar em coleções, etc. Exemplos típicos de SOCs são tesauros, classificações, listas de cabeçalhos de assunto e taxonomias" (SKOS, 2004).

Delimitada, então, a abrangência dos SOCs adotada neste capítulo, passemos às funções desses sistemas. Soergel (1999) apresenta diversas aplicações dos SOCs, das quais destacamos o auxílio à recuperação da informação. Para o autor, os SOCs fornecem uma base de conhecimento que apoia a busca pelo usuário final, como, por exemplo, menus em árvores, orientação de busca por facetas de determinado assunto, cadeias hierárquicas que orientam a navegação, mapeamento de termos de busca para descritores usados em bases de dados ou para outras expressões em linguagem natural. Além disso, esses sistemas apoiam a expansão da busca por meio da cadeia hierárquica e orientam a ordenação de resultados de busca. Nos processos de tratamento da informação, Soergel (1999) classifica os SOCs como instrumentos de indexação.

Para Vickery (2008), os SOCs explicitam o conhecimento de um domínio e possibilitam: indexação, organização e recuperação de informações ou de outros tipos de objetos; construção de mapas de conhecimento; navegação para busca de informação e criação de novo conhecimento a partir do existente. Segundo o autor, as relações semânticas em SOCs são necessárias para que cumpram as seguintes funções de auxílio à busca:

- buscas genéricas: seleção de um conjunto de itens relevantes para determinado assunto geral. O SOC reúne todos os assuntos específicos em um campo de assunto mais geral, de maneira que possam ser encontrados pelo usuário;
- buscas específicas: seleção de um conjunto de itens do conhecimento relevantes para um tópico específico. Esses assuntos normalmente são descritos por uma combinação de termos, os quais devem estar reunidos de forma compreensível em um SOC;
- arranjo sequencial: arranjo de um conjunto de itens armazenados ou selecionados de maneira compreensível para o usuário;
- escolha de termos de busca: apresentação de termos semanticamente relacionados, de maneira que o usuário possa selecionar aqueles que melhor representem o assunto que deseja.

Na visão de Shiri e Molberg (2005), os SOCs têm o potencial de prover várias perspectivas de uma coleção no ambiente digital, pois oferecem ao usuário uma estrutura conceitual rica e uma estrutura semântica que facilitam a seleção de termos para a formulação, reformulação e expansão da busca. A função dos SOCs em bibliotecas digitais é abordada por Hodge (2000): Bibliotecas digitais podem usar SOCs para estabelecer elos entre recursos digitais e entre esses e objetos físicos. Um exemplo disso é a expansão de códigos e siglas. Os registros podem ser localizados diretamente por meio da navegação nos SOCs ou indiretamente, por meio de uma chave de busca que pode ser usada para acessar outra fonte de informação. Esse conceito coloca o uso dos SOCs um passo à frente, fornecendo informações sobre a localização física de objetos em museus ou catálogos de exposições.

McCulloch e Macgregor (2008) discutem a questão do mapeamento terminológico entre diferentes SOCs em repositórios digitais, pois com o crescimento do número de repositórios, objetos digitais são indexados e organizados segundo uma variedade de diferentes esquemas. Para esses autores, é irreal esperar que os usuários interroguem cada repositório separadamente e que estejam familiarizados com as inúmeras terminologias adotadas. É necessário, portanto, que os usuários possam consultar e navegar simultaneamente múltiplos repositórios distribuídos.

Com base no exposto, concluímos que os SOCs cumprem duas funções principais. Na organização da informação, cuidam da padronização da representação dos conteúdos dos documentos e, na recuperação da informação, orientam e auxiliam o usuário quanto à localização de conteúdos. Essas funções se aplicam tanto ao ambiente tradicional quanto ao digital.

3. QUAIS SÃO OS ELEMENTOS QUE COMPÕEM OS SOCs?

Como nos lembra Dahlberg (1993,p. 211), o aspecto mais importante da base teórica da área de Organização do Conhecimento é o fato de que "qualquer organização do conhecimento deve ser baseada em unidades do conhecimento – que nada mais são do que conceitos". Nesse ponto Hjorland (2007) concorda com Dahlberg e afirma que "a organização do conhecimento é basicamente organização de conceitos". Entendemos, portanto, que os **componentes essenciais** dos Sistemas de Organização do Conhecimento, **são os conceitos**, uma vez que esses são as unidades básicas do conhecimento.

Para Sowa (1984, p. 344), os conceitos "são invenções da mente humana usadas para construir um modelo de mundo". Segundo Dahlberg (1978), os conceitos são unidades de conhecimento, identificadas por meio de enunciados verdadeiros acerca de determinado objeto e representadas por uma forma verbal.

O SKOS considera o conceito como elemento fundamental e o define como "unidades do pensamento – ideias significados, ou categorias de objetos e eventos, que constituem a base de vários sistemas de organização. Como tal, os conceitos existem na mente dos indivíduos como entidades abstratas, as quais são independentes dos termos usados para rotulá-las (SKOS, 2004)".

Na elaboração de SOCs, o principal objeto de trabalho são os conceitos e não a expressão usada para se referir a eles (os termos, códigos, fórmulas ou outros símbolos que os representam). O SKOS denomina essas expressões de **rótulos ou etiquetas** (*labels*), os quais constituem **elementos dos SOCs** usados para fazer referência aos conceitos.

O exemplo a seguir, extraído do *Tesauro de Folclore e Cultura Popular*, ilustra como os elementos rótulo e conceito são apresentados em tesauros:

> **Agogô**
> Instrumento de percussão cujo som é extraído por uma vareta de metal percutida no corpo do instrumento, que é constituído por duas campânulas de metal de tamanhos diferentes.

Nesse exemplo, 'agogô' é o rótulo – nesse caso um termo – que representa o conceito desse tipo de instrumento de percussão. A definição que segue abaixo do termo explicita seu significado.

Para Vickery (3008), além dos termos que representam os conceitos, os SOCs devem, ainda, indicar relações entre eles. O SKOS atribui às relações semânticas papel crucial na definição de conceitos e destaca que esses não são definidos apenas pelas palavras da linguagem natural, mas também pelos elos que os ligam a outros conceitos.

As **relações semânticas ou conceituais** constituem, portanto, o terceiro **elemento dos SOCs** e são associações entre conceitos que representam entidades no contexto de um SOC. Para Bean,Green e Myaeng (2002), "conceitos são os tijolos de uma estrutura conceitual, as relações entre eles são o cimento que os mantêm juntos".

As relações conceituais são estabelecidas por meio da análise das características ou propriedades dos conceitos, as quais permitem identificar diferenças e semelhanças que evidenciam determinados tipos de relacionamentos, como afirma Blis: "Na multiplicidade, diversidade e complexidade dos objetos naturais e suas relações, percebemos que coisas individuais e eventos diferem nas suas individualidades e características específicas, alguns são similares em certas características e relações; e por essas características de similaridade nós os relacionamos em classes e conceitos no processo de aprendizagem e de compreensão mental" (BLISS, 1952, p. 78).

A organização dos conceitos em classes é um processo arbitrário, no qual selecionamos os relacionamentos a serem apresentados, segundo as peculiaridades do domínio que pretendemos representar. Ao analisar um domínio, identificamos facetas e podemos agrupar os conceitos e relacioná-los de diferentes maneiras. Sowa (1984) destaca que "... conceitos e percepções não podem formar modelos perfeitos de mundo, são abstrações que selecionam características importantes para determinada finalidade e que ignoram detalhes ou complexidades que são importantes apenas para determinados objetivos (SOWA, 1984, p.344)."

Sendo assim, a mesma entidade pode ser caracterizada de diferentes maneiras, dependendo do domínio em questão e do tipo de sistema de organização do conhecimento que desejamos construir. Ao analisar o conceito de leite em diferentes domínios, determinadas características são ressaltadas, como ilustrado a seguir:

> **Leite**
> *Para a biologia*: líquido secretado pelas glândulas mamárias das fêmeas dos mamíferos, um pouco antes e principalmente após o parto.
> *Para a química*: mistura complexa constituída de substâncias orgânicas e inorgânicas, na qual encontramos água, gordura, carboidrato, proteínas, sais minerais, vitaminas, certas enzimas e gases. (Fonte: www.dag.uem.br/prof/ptmpintro/material/4b/leite_e_derivados.pdf)

Evidentemente o conceito "completo" de leite contempla essas e outras características. No entanto, cada domínio destaca aquelas que são particularmente importantes para situar um conceito em relação aos demais que compõem o seu sistema conceitual. Com base no exemplo, pode-se estabelecer uma relação entre leite e glândulas mamárias, na área de biologia, já na área de química, entre leite e cada uma das substâncias que o compõem.

A arbitrariedade da seleção dos agrupamentos e relacionamentos não pode ser vista, no entanto, como algo totalmente aleatório. Ela se deve aos aspectos que desejamos destacar numa determinada representação do conhecimento. Devemos atentar que, acima de tudo, salienta Hodge (2000), "deve haver uma correspondência entre o SOC e a realidade que ele pretende representar". Portanto, na construção de SOCs não é aceitável, por exemplo, agrupar baleias sob a classe peixes, pois esse agrupamento não corresponde à realidade, uma vez que baleias não são peixes e sim mamíferos.

Cada um dos elementos dos SOCs demanda estudo de aspectos teóricos e metodológicos específicos. Para o estudo do conceito, devemos recorrer, entre outras à Teoria do Conceito e à Teoria Geral da Terminologia. Para os termos, necessitamos da Linguística, da Terminologia, da Semiótica, da Lexicografia. Já para os relacionamentos recorreremos às abordagens semânticas oferecidas por diferentes áreas do conhecimento que se ocupam desse campo de estudo.

Como não é possível aprofundar o estudo de cada elemento, nos limitaremos, ao final dessa seção, a responder a questão colocada inicialmente, afirmando que conceitos, termos e relacionamentos são os principais elementos que compõem os SOCs. A maneira como esses elementos são tratados e apresentados depende do tipo de sistema, como veremos nas seções seguintes.

4. O QUE SÃO CLASSIFICAÇÕES, TESAUROS, TAXONOMIAS E ONTOLOGIAS?

4.1 Classificações

A classificação é, provavelmente, o método mais simples de ordenar a confusa multiplicidade da natureza. É um processo de seleção de ideias ou objetos em grupos, conforme suas qualidades semelhantes e diferenças específicas. Esse processo, chamado de abstração, é essencialmente mental, nós agrupamos ou separamos coisas de acordo com o conceito ou ideia que temos dos objetos no mundo, fortalecendo a memória e o poder de raciocínio. Sem isso, nada pode ser identificado, portanto, na prática, pode-se dizer que os pensamentos e raciocínios consistem de classificação. Citado por Phillips (1955, p. 10), Mill afirma que a proposta da classificação é primeiramente facilitar as operações da mente concebendo e retendo claramente na memória as características dos objetos em questão.

Enquanto fenômeno social, as pessoas classificam, intuitivamente, as coisas o tempo todo. Como explica Langridge (2006, p. 11) o fato de que a maioria das pessoas não percebe o quanto classifica é meramente um indício da natureza fundamental do processo de classificação.

O processo de classificação é uma formação metodológica e sistemática onde se estabelecem critérios para a divisão, isto é, a formulação de um esquema de categorias, classes e subclasses, baseado nas características e relações dos objetos considerados. É, também, um sistema logicamente estruturado onde os conceitos pré-determinados correspondem a um código identificador.

Os sistemas de classificação podem ser vistos sob três aspectos, de acordo com Phillips (1955): quanto à finalidade, quanto à característica e quanto à área do conhecimento. Segundo a finalidade, temos a classificação filosófica, das ciências ou do conhecimento voltadas para a definição e a hierarquização do conhecimento humano e a bibliográfica, destinada à organização dos documentos (livros, periódicos, etc.) para facilitar a localização. Quanto à característica a classificação pode ser natural, quando o fator de distinção é fundamental para a natureza do objeto e artificial, quando a base da divisão depende de uma qualidade acidental. Em relação à área do conhecimento, a classificação pode ser geral no caso da enciclopédica e especializada que podem ser científicas ou bibliográficas.

As classificações bibliográficas tiveram grande parte do seu desenvolvimento no final do século XIX e início do século XX e se originaram dos modelos de classificação filosófica, porém, com o objetivo de organizar e localizar os documentos. São sistemas predeterminados de conceitos logicamente estruturados e acompanhados de um código identificador. Este código é atribuído a conceitos ou a documentos em função da correspondência de assuntos.

Segundo Bliss (1952), uma classificação de material bibliográfico é uma classificação de assuntos, que, na verdade, constitui-se numa organização estrutural do conhecimento e do pensamento e que atende a objetivos funcionais de organização do conhecimento. Dos sistemas de classificação bibliográficos, os mais conhecidos são: Classificação Decimal de Dewey (CDD); Classificação Decimal Universal (CDU); Classificação da Biblioteca do Congresso dos Estados Unidos (LCC – Library of Congress Classification).

Neste capítulo, adotaremos a denominação classificação para nos referir aos sistemas de classificação utilizados para organização e recuperação da informação, que englobam os supracitados, assim como outros desenvolvidos para atender necessidades específicas.

Há interesse renovado no estudo das classificações e seu uso foi ampliado com a necessidade de organizar conteúdos na web. As classificações são utilizadas para a organização conceitual que auxilia a busca por palavras. Destacamos, ainda, a importância da aplicação de princípios classificatórios

desenvolvidos pela Biblioteconomia na construção de sistemas de organização do conhecimento.

4.2 Tesauros

O termo tesauro é originário do latim – thesaurus e do grego – thesaurós, que quer dizer tesouro, armazenamento, repositório. Cavalcanti (1978, p. 27) define tesauro como uma "lista estruturada de termos associados empregada por indexadores para descrever um documento com a desejada especificidade e para permitir aos usuários a recuperação da informação que procuram".

O tesauro é geralmente temático e atende uma área específica do conhecimento, mas existem tesauros multidisciplinares que atendem a sistemas de informação dessa natureza. Segundo Gomes (1990), os tesauros são classificados quanto à tipologia, de acordo com a(s) língua(s) que englobam, em monolíngues e multilíngues. Quanto ao tipo de conceitos, são macrotesauros (representam conceitos mais amplos) e microtesauros (representam conceitos específicos) e, de acordo com a abrangência temática, são multidisciplinares e de disciplina específica.

No contexto da organização e recuperação da informação, tesauros são instrumentos de controle terminológico, utilizados em sistemas de informação para traduzir a linguagem dos documentos, dos indexadores e dos pesquisadores numa linguagem controlada, usada na indexação e na recuperação de informações.

Guinchat e Menou (1994, p.146) apontam que as vantagens do tesauro são sua especificidade, maleabilidade e capacidade de descrever as informações de forma completa. Os tesauros auxiliam o indexador a localizar o conceito num campo semântico e o usuário a identificar assuntos relacionados que podem ser de interesse na busca. O emprego das novas tecnologias da informação possibilita diferentes formas de apresentação, mais amigáveis e compreensíveis para orientar os usuários na navegação, ampliando ou restringindo sua busca.

De acordo com o IBICT (1984, p. 1-2), as principais finalidades de um tesauro são: a) controlar os termos usados na indexação mediante um instrumento que traduza a linguagem natural dos autores, usuários e indexadores, para uma linguagem mais controlada; b) uniformizar, mediante esta linguagem documentária, os procedimentos de indexação de profissionais em uma instituição ou numa rede cooperativa; c) limitar o número de termos necessários à explicitação dos conceitos expostos pelos autores de uma área; d) auxiliar a tarefa de recuperação da informação, fornecendo termos adequados para a estratégia de busca.

Os tesauros são sistemas de uso consolidado em sistemas de informação e são, inclusive, padronizados por meio de normas, como a ISO 2788 (1986) e ISO 5964 (1985), no caso de tesauro multilíngue e ANSI / NISO Z39.19-2003, que define diretrizes de construção e formatação de tesauros. Há, ainda, diretrizes e manuais de elaboração de tesauros[5], disponíveis também em língua portuguesa.

4.3 Taxonomias

Etimologicamente, o termo taxonomia deriva do grego, com o significado de: taxis – ordenação, grupo e nomos - lei, norma, regra. A palavra taxonomia foi introduzida por A. de Candolle, em 1813, para designar as normas e leis utilizadas em sistemática (CURRÁS, 2005 p. 53-54).

Reitz (2004), no Online Dictionary for Library and Information Science[6] define taxonomia como a ciência de classificar que abrange os princípios gerais pelos quais objetos e fenômenos são divididos em classes, as quais estão subdivididas em subclasses, e essas em sub-subclasses e assim

[5] Alguns manuais de elaboração de tesauros: AITCHISON, J.; GILCHRIST, A. *Manual para construção de tesauros*. Rio de Janeiro RJ: BNG/Brasilart,1979, 142 p.; AUSTIN, D. *Diretrizes para o estabelecimento de tesauros monolíngües*. Brasília DF: IBICT/SENAI, 1993; GOMES, H.E. *Manual de elaboração de tesauros monolíngües* . Brasília DF: PNBU,1990, 78 p.; IBICT. *Diretrizes para elaboração de tesauros monolíngües* . Brasília DF: IBICT, 1984, 70 p.

[6] http://lu.com/odlis/search.cfm

sucessivamente. As taxonomias permitem que se estabeleçam padrões de classificação e ordenação de informações por meio de herança, ou seja, pelo relacionamento hierárquico (gênero/espécie) entre os objetos, em que as características das classes gerais são repassadas por herança às subclasses.

No contexto da Ciência da Informação, uma taxonomia é um método de classificar coisas reais, estabelecendo categorias de similaridades e diferenças (KENT; LANCOUR, 1968, p.187). No domínio das representações do conhecimento, as taxonomias são instrumentos que organizam logicamente os conteúdos informacionais. O conceito de organização remete para um procedimento classificatório e permite agrupamento categorizado, isto é, a partir de um assunto formam-se categorias que se dividem em classes e subclasses hierarquicamente, formando uma lista de categorias de assunto estruturada.

Graef (apud EDOLS, 2006) apresenta as taxonomias como esquemas que classificam coisas – organismos vivos, produtos, livros – em uma série de grupos hierárquicos para serem mais fáceis de identificar, estudar e localizar. E são compostas de duas partes – estruturas e aplicações. As estruturas consistem em categorias e os relacionamentos e as aplicações são as ferramentas de navegação disponíveis para ajudar os usuários a encontrarem as informações que desejam.

Na organização e recuperação de informação, as taxonomias são usadas para a criação de metadados ou termos para descrever um objeto e na categorização para definir classes e subclasses, como suporte à navegação no ambiente web. Diante do contexto dinâmico da informação, Conway e Sligar (2002) destacam que as taxonomias não são documentos estáticos que ficam guardados em um lugar seguro. Ao contrário, elas adaptam-se e se transformam conforme as alterações de conteúdo e de conhecimento de quem as utilizam.

As taxonomias ganharam importância na organização de informações no contexto empresarial. Alan Gilchrist e Peter Kibby citados por Edols (2001) destacam que as taxonomias:

- estabelecem correlação de diferentes linguagens funcionais usada pelas empresas;
- apoiam a navegação e permitem melhor acesso ao conteúdo intelectual da empresa;
- são ferramentas que permitem a etiquetagem de documentos e outros objetos informacionais úteis para as ferramentas de busca e construção de mapas de conhecimento.

Segundo Campos e Gomes (2007) as taxonomias caracterizam-se por:

- conter uma lista estruturada de conceitos/ termos de um domínio;
- incluir termos sem definição, somente com relações hierárquicas;
- possibilitar a organização e recuperação de informação através de navegação;
- permitir agregação de dados, diferentemente das taxonomias seminais, além de evidenciar um modelo conceitual do domínio;
- ser um instrumento de organização intelectual, atuando como um mapa conceitual dos tópicos explorados em um sistema de recuperação de informação;
- ser um novo mecanismo de consulta em portais institucionais, através de navegação.

Conway e Sligar (2002) classificam as taxonomias em três tipos. A taxonomia descritiva, cujo objetivo é dar suporte à recuperação de informações por meio de busca. Está diretamente ligada ao conteúdo dos documentos da instituição, a partir da qual é desenvolvida. A taxonomia navegacional, como o nome expressa, tem por objetivo a descoberta do conhecimento por meio da navegação e é construída com base no comportamento

do usuário e não no conteúdo dos documentos. Por último, a taxonomia para gerenciamento de dados, que contém pequeno conjunto de termos controlados rigidamente, é uma lista de termos autorizados, sem qualquer estrutura hierárquica, utilizada para apoiar as transações comerciais.

Observa-se que as taxonomias adquirem importância no contexto das organizações, como instrumentos auxiliares à gestão do conhecimento e à organização e recuperação da informação. Elas asseguram que todas as aplicações da instituição utilizem a mesma linguagem para organizar, armazenar e apresentar a informação.

4.4 Ontologias

Assim como os termos tesauro e taxonomia, ontologia deriva das línguas antigas, vem do grego - ontos e logoi, que significa conhecimento do ser. A área de informação toma o termo 'emprestado' da Filosofia, em cujo âmbito refere-se à teoria sobre a natureza da existência.

As ontologias são objetos de estudo de diferentes áreas do conhecimento (Filosofia, Ciências Cognitivas, Ciência da Computação, Linguística, Ciência da Informação, Medicina, Empresarial), é natural que haja multiplicidade de acepções e de definições variadas que refletem a visão de cada domínio.

Para a organização e recuperação da informação, o conceito empregado é proveniente da proposta de Berners Lee (2001) e está associado ao contexto da Web Semântica[7]. Para o autor, ontologia é "documento ou arquivo que define formalmente as relações entre termos." Outra definição comumente encontrada na literatura é a apresentada por Gruber (1996 p.1) - "ontologia é uma especificação formal e explícita de uma conceitualização, o que existe é aquilo que pode ser representado" [...]

As ontologias definem conceitos e relações de alguma área do conhecimento, de forma compartilhada e consensual e promovem e facilitam a interoperabilidade entre sistemas de informação, em um processo "inteligente" dos agentes (computadores). A conceitualização proposta por uma ontologia deve ser representada de maneira formal, legível e utilizável por computadores, de maneira a permitir o compartilhamento e o reuso do conhecimento entre os sistemas. Uma ontologia define os termos usados para descrever e representar uma área do conhecimento. Podem ser usados por pessoas, bancos de dados e aplicações que necessitam compartilhar informações em um domínio.

Durante os últimos anos, as atenções têm se voltado para a metodologia de construção de ontologias, as quais podem ser construídas a partir do reuso de ontologias disponíveis. Diferentes processos de construção de ontologias tendem a estar associados com diferentes tipos de ontologias. No entanto, como destacam Noy e McGuinness (2001), o desenvolvimento de uma ontologia inclui definição de classes, estruturação de classes numa hierarquia taxonômica, definição de atributos[8] e descrição dos valores permitidos para esses atributos.

Para cada domínio as abordagens para construção de ontologias, geralmente, são específicas e limitadas e dependem dos interesses para os quais são construídas. Mas, independente do domínio, a elaboração de uma ontologia requer um trabalho complexo e dispendioso.

Na Web Semântica, as ontologias especificam descrições para conceitos, com semântica expressiva, explícita e bem-definida, possivelmente interpretável por máquina, que processam automaticamente as informações encontradas na web. Para tanto, usam linguagens compatíveis (OWL, XML e

[7] Nas palavras de Tim Berners-Lee (2001) "web semântica não é uma web separada, mas uma extensão da atual". É a criação e implantação de padrões tecnológicos permitindo que dados sejam compartilhados e reusados, é um esforço colaborativo liderado pela W3C com participação de pesquisadores e empresas. Está baseado na RDF (R*esource Description Framework*), que integra uma variedade de aplicações usando XML (*eXtensible Markup Language)* para sintaxe e URI (U*niform Resource Identifier)* para identificação. O objetivo da web semântica é transformar o conteúdo atual da web num formato no qual não só humanos, mas também computadores entendam o significado das informações, de forma a permitir que agentes inteligentes recuperem e manipulem informações pertinentes.

[8] Além de atributos, encontramos também, na literatura sobre ontologias, os termos *slots,* propriedades, papéis ou características com a mesma significação

XML *Scheme*, RDF e RDF *Scheme*) com o ambiente digital.

Classificações, tesauros, taxonomias e ontologias estruturam, classificam, modelam e representam conceitos e seus relacionamentos pertinentes num domínio do conhecimento. São formados, basicamente, por vocabulários controlados acompanhados de relacionamentos semânticos entre os termos. Desempenham a função de organizar e recuperar informações. No entanto, como veremos a seguir, os quatro sistemas mantêm pontos em comum, mas diferenciam em alguns aspectos.

5. QUAIS SÃO AS SEMELHANÇAS E AS DIFERENÇAS ENTRE CLASSIFICAÇÕES, TESAUROS, TAXONOMIAS E ONTOLOGIAS?

A questão que colocamos é difícil de ser respondida, como alerta Uschold (2010), em comentário ao *paper* intitulado *What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model?*, de autoria de Pidcock (2010). Portanto, nesta seção não pretendemos apresentar respostas definitivas, procuramos, por meio de exemplos, contribuir para o esclarecimento acerca das semelhanças e diferenças entre classificações, tesauros, taxonomias e ontologias. Identificamos os pontos comuns entre os quatro sistemas e analisamos as peculiaridades de cada um deles. Para Soergel (1999), existem vários tipos de bases de conhecimento: esquemas de classificação e tesauro, dicionários e ontologias desenvolvidos para aplicações de Inteligência Artificial, sistemas linguísticos ou definições de elementos de dados. Apesar de se desenvolverem com diferentes objetivos, os diferentes tipos de bases de conhecimento sobrepõem-se muito e seguem princípios e métodos muito similares para sua construção.

Uschold (2010) destaca os seguintes aspectos comuns acerca de vocabulários, taxonomias, tesauros, ontologias e meta-modelos:

- são diferentes abordagens para ajudar a estruturar, classificar, modelar e/ou representar conceitos e relações de uma área de interesse de uma comunidade;
- estabelecem um conjunto de termos que a comunidade concorda em usar para referir-se a estes conceitos e relações;
- especificam o significado dos termos em algum nível;
- são noções não muito bem definidas e usadas de diferentes maneiras por diferentes indivíduos e comunidades.

Zeng (2006) observa que os diversos sistemas compartilham características comuns, mesmo sendo construídos segundo diferentes padrões e convenções. Em trabalho mais recente, a autora assinala que os diferentes tipos de SOC têm, ainda, como ponto comum as seguintes funções fundamentais: eliminar ambiguidades, controlar sinônimos, estabelecer relacionamentos (hierárquicos e associativos) e apresentar propriedades [dos conceitos]. (ZENG, 2008). De certa forma, os autores citados concordam que, de maneira geral, os SOCs possuem semelhanças. No entanto, como destaca Uschold (2010), há diferenças entre os SOCs quanto ao nível de especificação do significado de cada termo e quanto ao tipo de notação ou linguagem adotada para especificar o significado. Fato que observamos, também, em relação às classificações, tesauros, taxonomias e ontologias.

Com o objetivo de identificar as características comuns e as diferenciadoras dos quatro sistemas, passamos a comentar em seguida cada um dos aspectos considerados. Como ponto de partida para nossa análise, como já esclarecemos anteriormente, assumimos que classificações, tesauros, taxonomias e ontologias (CTTOs) são formas de representação do conhecimento aplicadas à organização e recuperação da informação. Esses sistemas possuem os aspectos comuns e diferenciadores elencados a seguir.

5.1 Princípios de classificação

CTTOs compartilham os princípios de classificação, que orientam a estrutura sistemática e possibilitam o agrupamento de conceitos que possuem características comuns.

Os conceitos agrupam-se em classes[9] ou categorias[10], como ilustram os quadros a seguir:

Quadro 1. Exemplo de classe da Classificação Decimal Universal (CDU)

6 Ciências Aplicadas. Medicina. Tecnologia. Podemos ver como se subdivide:

61 Ciências Médicas.
62 Engenharia. Tecnologia em Geral.
63 Agricultura. Silvicultura. Agronomia. Zootecnia.
64 Ciência Doméstica. Economia Doméstica.
65 Organização e Administração da Indústria, do Comércio e dos Transportes.
66 Tecnologia Química. Indústrias Químicas.
67 Indústrias e Ofícios Diversos.
68 Indústrias, Artes e Ofícios de Artigos Acabados.
69 Construção Civil. Materiais de Construção. Prática e Processos de Construção.

A subclasse **62 Engenharia.** Subdivide-se por sua vez em:

- **620** Engenharia em Geral. Testes dos Materiais. Energia.
- **621** Engenharia Mecânica.
- **622** Engenharia de Minas.
- **623** Engenharia Naval e Militar.
- **624 Engenharia Civil e Estruturas em Geral. Infra-estruturas. Fundações. Construção de Túneis e de Pontes. Superestruturas.**

624 Engenharia Civil divide-se em áreas diferentes que podem por sua vez ser divididas novamente em áreas ainda mais especializadas:

- **624.01 Estruturas e elementos estruturais segundo o material e o processo de construção.**
  - **624.011 Estruturas e materiais de origem orgânica.**
  - **624.012 Estruturas de alvenaria.**
    - **624.012.45 Estruturas de betão armado.**

[9] Uma classe compreende todas as coisas que são vistas como semelhantes, que correspondem ao conceito da classe, que são definidas pela classe e denominadas por seu termo. (Bliss, 1952)

[10] São propriedades gerais dos conceitos, utilizadas para agrupá-los segundo características comuns.

- **624.1** Infra-estruturas das construções. Fundações. Construção de túneis.
- **624.2/.8** Construção de pontes.....etc.

(Fonte: http://www.bib.ualg.pt/bibliotecas/cdu.htm)

Figura 2. – Categorias do Tesauro de Folclore e Cultura Popular Brasileira.
(Fonte: http://www.cnfcp.gov.br/tesauro/sistematica/sistematico.htm)

A classe 6 da CDU agrupa 9 subclasses, que, por sua vez, subdividem-se sucessivamente, como exemplifica a subclasse 62, que engloba as engenharias. As figuras 2 e 3, ilustram o uso de categorias em tesauro.

Na *Figura 2* observamos que o tesauro possui 15 categorias de assunto, já na *Figura 3* encontramos uma definição que indica a abrangência da categoria e, em seguida, alguns exemplos de termos que se encontram nela inseridos.

A taxonomia de navegação por diretórios, ilustrada na *Figura 4*, possui 13 categorias principais. A categoria Artes e Entretenimento, por exemplo, reúne páginas de música, televisão, rádios, entre outras.

Figura 3. – Categoria Alimento do Tesauro de Folclore e Cultura Popular Brasileira.
(Fonte: http://www.cnfcp.gov.br/tesauro/00001579.htm)

Figura 4: Taxonomia de navegação por diretório.
Fonte: http://www.google.com.br/dirhp?hl=pt-BR

Na *Fgura 5* encontramos a ilustração das 12 classes da ontologia utilizada no projeto *Friend of a Friend*. No detalhe da classe *Group* (grupos), conforme *Figura 6*, encontra-se a definição que restringe os membros dessa classe: *coleção de agentes individuais*. Por sua vez, a classe *Agent* (agente) abrange *aquilo que pode executar uma ação*.

Figura 5: Classes da Ontologia utilizada pelo projeto Friend of a Friend (Fonte: http://xmlns.com/foaf/0.1/)

Class: foaf:Group

Group - A class of Agents.
Status: stable
in-domain-of: foaf:member

The foaf:Group class represents a collection of individual agents (and may itself play the role of a foaf:Agent, ie. something that can perform actions).

This concept is intentionally quite broad, covering informal and ad-hoc groups, long-lived communities, organizational groups within a workplace, etc. Some such groups may have associated characteristics which could be captured in RDF (perhaps a foaf:homepage, foaf:name, mailing list etc.).

While a foaf:Group has the characteristics of a foaf:Agent, it is also associated with a number of other foaf:Agents (typically people) who constitute the foaf:Group. FOAF provides a mechanism, the foaf:membershipClass property, which relates a foaf:Group to a sub-class of the class foaf:Agent who are members of the group. This is a little complicated, but allows us to make group membership rules explicit.

The markup (shown below) for defining a group is both complex and powerful. It allows group membership rules to match against any RDF-describable characteristics of the potential group members. As FOAF and similar vocabularies become more expressive in their ability to describe individuals, the foaf:Group mechanism for categorising them into groups also becomes more powerful.

While the formal description of membership criteria for a foaf:Group may be complex, the basic

Figura 6. – Detalhamento da classe Group da Ontologia utilizada pelo projeto Friend of a Friend.
(Fonte: http://xmlns.com/foaf/0.1/#term_Agent)

Os exemplos ilustrados mostram o uso do princípio classificatório nos quatro sistemas. A maneira como é aplicado esse princípio e os tipos de entidades tratadas em cada sistema resultam em diferentes formas de tratamento das classes e categorias.

5.2 Rótulos ou etiquetas

Os CTTOs utilizam algum tipo de signo para representar os conceitos. Tomando ainda os exemplos ilustrados nas figuras da subseção 5.1, observamos diferentes tipos de signos. Na classificação, o uso concomitante do código numérico e do termo (622. Engenharia de Minas). No tesauro e na taxonomia, o uso de termos (azeite, farinha, queijo e revistas, televisão, rádio). Na ontologia, o uso de uma forma de expressão compatível com padrões como XML[11], XML Namespaces[12], RDF[13], e OWL[14]. Ex.: foaf:Group, foaf:Agent.

Os exemplos não esgotam as possibilidades. Tesauros e taxonomias podem utilizar códigos numéricos ou alfanuméricos para representar os conceitos, mas, o que comumente se observa é que esses dois sistemas tendem a utilizar termos. As classificações são as que mais empregam os códigos numéricos. Já as ontologias necessitam expressar o conceito para que o computador possa identificá-lo, daí a necessidade de formalismos definidos pelos padrões da web semântica.

5.3 Padronização terminológica

Os CTTOs cumprem a função de padronizar os termos empregados para representar os conceitos. A padronização contempla: aspectos de forma, como, por exemplo, uso do gênero masculino e feminino, plural e singular, sigla e nome por extenso; controle de sinonímias e controle de homografias. A diferença está, mais uma vez, na maneira como cada sistema efetua a padronização. Para elaboração de tesauros, há normas internacionais que definem os aspectos de controle terminológico (ISO 2788 (1986), ISO 5964 (1985) e ANSI / NISO Z39.19 (2005) e que determinam, entre outras convenções, o uso da relação de equivalência, representada pelos símbolos USE e UP para controle de sinônimos e o uso de qualificadores[15] para controlar homografias.

As ontologias contam com padrões definidos pela OWL, entre os quais destacamos, no caso dos sinônimos, as relações equivalentClass, utilizada para indicar 'classes sinônimas'; equivalentProperty, que indica a sinonímia entre propriedades de conceitos ou classes e a relação sameAs, empregada para criar diferentes nomes que referem ao mesmo indivíduo. (fonte: http://www.w3.org/TR/2004/REC-owl-features-20040210/#sameAs)

Para as classificações e taxonomias, no entanto, não há padrão definido acerca da padronização terminológica. Nesse caso, sobretudo em relação às taxonomias, observamos alguns problemas terminológicos. Ao analisar taxonomias navegacionais, observamos, por exemplo, a ausência de padrões relativos ao uso de termos em português e de abreviaturas e siglas, como na taxonomia de livrarias virtuais, em que encontramos alguns exemplos: *Pocket Books* e *Franchising*; Adm. Financeira / Mat. Finan. ; Dic. Francês/Port., (AQUINO; CARLAN; BRÄSCHER, 2009). Numa rápida consulta a taxonomias de sítios de comércio eletrônico o leitor encontrará diversos problemas relativos à forma dos termos.

---

[11] XML – eXtensible Markup Language – http://www.w3.org/XML/
[12] XML Namespaces - http://www.w3.org/TR/REC-xml-names/
[13] RDF – Resource Description Framework http://www.w3.org/RDF/
[14] OWL – Web Ontology Language http://www.w3.org/TR/owl-features/
[15] Qualificadores são termos agregados entre parênteses aos descritores para delimitar seu significado e eliminar a homografia. Ex.: tênis (esporte); tênis (calçado).

5.4 Especificação do conceito

CTTO apresentam recursos para delimitar o significado que compreendem, desde o simples agrupamento de termos em classes ou categorias adotado pelas classificações, até as complexas definições de propriedades/atributos e formalismos empregados pelas ontologias. As classificações e taxonomias, de maneira geral, não incluem definições precisas dos termos que arrolam, como nos exemplos do quadro 1 e da figura 4. No entanto, o posicionamento dos termos em determinadas classes ou cadeias hierárquicas delimita o significado. A inclusão de 'gol' na classe de empresas aéreas, por exemplo, restringe o seu significado e exclui outras possibilidades, tais como marca de automóvel e evento do futebol.

Figura 7. – Exemplo de termo do Vocabulário Controlado Básico do Senado Federal (VCB)
(Fonte:http://biblioteca.senado.gov.br:8991/F/KM99K6EPDEYGDIXAMVA92RSLKSELMRQSSEX9HPEF7JK841VI9C-00180?func=full-setset&set_number=161411&set_entry=000007&format=999)

Figura 8.. – Exemplo de ontologia de vinho
(Fonte: http://protege.cim3.net/download/old-releases/2.0/)

Nos tesauros os termos costumam ser acompanhados de definições expressas em linguagem natural. A norma ANSI/NISO Z39.19 determina que o escopo dos termos é restrito aos significados que possuem no domínio que o tesauro abrange, e, ainda, que os termos devem ser formulados de maneira a tornar claro, a qualquer usuário do tesauro, qual o escopo pretendido. Para tanto, o recurso utilizado em tesauros são as notas explicativas ou de escopo, que "devem exprimir o significado selecionado para o termo e, também, indicar outros significados reconhecidos na linguagem natural, mas que foram deliberadamente excluídos do vocabulário controlado" (ANSI/NISO Z39.19, 2005, p. 22). A inclusão dos termos em categorias e seus relacionamentos são outros recursos que auxiliam na delimitação dos conceitos em tesauros, como no exemplo da *Figura 2*.

O termo indexação é empregado em sentidos diferentes na Ciência da Informação e na Economia, portanto, é um termo homógrafo. O qualificador (economia) é utilizado para solucionar a homografia e já restringe o significado do termo. Mesmo que observássemos apenas a subclasse 330 – Economia – Finanças públicas ou as relações 'termo específico' e 'termo relacionado' e não atentássemos para o qualificador, já poderíamos inferir que se trata do conceito de indexação usado na economia. No entanto, a definição que acompanha o termo é o recurso que melhor cumpre a função de delimitar o significado.

Assim como nos tesauros, os recursos de categorização, relacionamentos conceituais e definições estão presentes nas ontologias. No entanto, as ontologias dispõem de mecanismos mais complexos que permitem elencar todos os atributos ou propriedades de determinada classe ou conceito, restringir propriedades e definir sua cardinalidade. A figura 8 exemplifica os recursos mencionados no desenvolvimento de uma ontologia na ferramenta *Protégé.*

No exemplo da figura 8 observa-se, à esquerda, a *class hierarchy* (hierarquia de classes), na qual se encontra a classe *Wine* (vinho), que é subclasse da classe *Drink* (bebidas), que, por sua vez, é subclasse da classe *Consumable thing* (coisas consumíveis). Ao analisar a classe *Wine* observa-se que suas subclasses são compostas por tipos de vinho: *White wine* (vinho branco), *Red wine* (vinho tinto), *Rose wine* (vinho rosé) e *Dessert wine* (vinho de sobremesa). Na parte do lado direito da figura há o *class editor* (editor de classe). O campo *documentation* (documentação) contém texto em linguagem natural que indica alguma restrição de significado. O item *Template slots*[16] contempla os atributos das classes, que, no caso de vinho são: *body* (corpo), *color* (cor), *flavour* (sabor), *grape* (uva), *maker* (produtor), *name* (denominação) e *sugar* (açúcar). Os atributos são herdados por todas as subclasses da classe vinho.

Para cada atributo é possível definir: *cardinality* (cardinalidade), *type* (tipo) e *other facets* (outras características), como no exemplo do atributo *sugar*, em que a cardinalidade é *single* (única), ou seja, esse atributo tem apenas um valor possível. O tipo é símbolo e indica que há um conjunto de símbolos pré-determinados para o atributo, o qual está indicado na característica *allowed-values* (valores permitidos) e essa, por sua vez, indica que o valor do atributo deve ser selecionado entre os indicados como válidos, no caso: *dry* (seco), *sweet* (doce) e *off-dry* (semi-seco).

Com base na descrição dos elementos acima, podemos observar que, entre os quatro sistemas, as ontologias são as que oferecem maior potencial de restrição conceitual. Devemos lembrar, no entanto, que ontologias são desenvolvidas com o objetivo de serem interpretadas por computador e de apoiarem motores de inferência.

5.5 Relacionamento conceitual

CTTO estabelecem relações entre conceitos. Os tipos de relacionamentos e a forma de representá-los variam segundo o sistema, sendo que os quatro contemplam a noção de hierarquia entre conceitos. Em tesauros e ontologias, as hierarquias seguem princípios rígidos, por outro lado, as classificações e taxonomias não possuem recursos para a diferenciação precisa dos tipos de relacionamentos hierárquicos.

As diretrizes e normas de elaboração de tesauros definem os seguintes relacionamentos hierárquicos:

a) relações gênero/espécie (TG – termo genérico / TE – termo específico). Ex.: instituição de ensino superior / universidade
b) relações todo/parte (TGP - termo genérico partitivo / TEP – termo específico partitivo). Ex.: universidade/faculdade/departamento
c) relação de instanciação (é um, ou *IsA*)[17]. Ex.: universidade /Universidade de Brasília.

[16] Nesse caso, não há ainda tradução para o português para o termo. Em geral, adota-se o termo *slot*.
[17] A relação de instanciação foi incluída mais recentemente em tesauros e encontra-se prevista na ANSI/NISO Z39.19.

Nas ontologias, a relação gênero/espécie orienta a hierarquia de classes, como vimos no exemplo da ontologia de vinho. Em tesauro teríamos que vinho é um termo específico de bebida, na ontologia temos que vinho é uma subclasse de bebida. Nas ontologias, a relação de instanciação ocorre pela ligação de cada indivíduo como membro de determinada classe. Assim, ainda no exemplo do domínio vinho, *Casillero del Diablo, Cabernet Sauvignon*, 2007 [18] é uma instância da classe vinho tinto. A relação todo/parte, assim como outros tipos de relacionamentos entre conceitos podem ser especificados nas ontologias por meio das propriedades (rdf:Property, na OWL), as quais são definidas segundo os objetivos da ontologia.

A identificação do relacionamento hierárquico em classificações e taxonomias é feita, de maneira geral, pelo posicionamento, na estrutura desses sistemas, de um conceito abaixo de outro, como nos exemplos a seguir, extraídos da taxonomia navegacional da loja virtual submarino (http://www.submarino.com.br/)

- Casa, Cozinha & Jardim
    - Cama, Mesa & Banho
    - Ferramentas
    - Automotivo
    - Eletrodomésticos
    - Eletroportáteis
    - Eletrônicos Audio & Vídeo
    - Utilidades Domésticas
    - Galeria de Arte

Ao selecionar a categoria Cama, Mesa & Banho, temos:

- Cama Solteiro
    - Jogo de Cama Solteiro
    - Edredom Solteiro
    - Colcha Solteiro
    - Kits de Solteiro
- Colchões
    - Solteiro
    - Casal
    - King
    - Queen

Podemos observar que não há indicação de tipo de relacionamento, há uma ideia de subordinação quando as categorias se expandem, mas os relacionamentos dentro delas são diversos. Na hierarquia da categoria Colchões, há relação gênero/espécie (colchões e tipos de colchões), já em Cama de Solteiro, a relação é de outra natureza (colcha que serve para cama de solteiro).

[18] Vinho tinto produzido pela vinícola chilena Concha y Toro.

Da mesma forma, quando analisamos o exemplo de classificação do *Quadro 1* (CDU), há relacionamentos que podem ser interpretados pela lógica todo-parte, como o que ocorre entre a grande área do conhecimento expressa pela classe 6 Ciências Aplicadas. Medicina. Tecnologia e as áreas que a compõem (Ciências Médicas, Engenharia, etc.). Já no caso de estruturas de alvenaria (624.012) e estruturas de betão armado (624.012.45), a relação é gênero/espécie, uma vez que as últimas são tipos da primeira.

Quanto aos relacionamentos, podemos afirmar que os quatro sistemas tratam das relações de equivalência, hierárquicas e associativas. A relação de equivalência está presente em todos eles, por meio de diferentes recursos: remissivas Ver, em classificações e taxonomias, relação USE/UP, em tesauros e *same as* (sameAs), nas ontologias. Os relacionamentos hierárquicos são tratados de forma clara em tesauros e ontologias, com as especificações empregadas para distingui-los. O mesmo não ocorre nas classificações e taxonomias, onde, além de não haver clara distinção de tipos de relações hierárquicas, essas não diferenciam também das associativas, como no exemplo da taxonomia da loja virtual Submarino, em que a relação 'serve para' (colcha para cama de solteiro) seria um tipo de relação associativa e é apresentada da mesma maneira que a relação hierárquica.

Sob as relações associativas, os tesauros agrupam todas as demais relações que não sejam de equivalência ou hierárquicas. O mesmo não ocorre nas ontologias, que definem precisamente o tipo de relacionamento associativo: causa/efeito, agente/processo, processo/instrumento, entre outros.

O tratamento dos relacionamentos depende, evidentemente, dos objetivos de cada um dos sistemas. Não podemos dizer que as ontologias, por apresentarem maior riqueza de relacionamentos, são melhores que taxonomias ou tesauros. Elas certamente serão melhores para determinadas finalidades que exigem maior grau de especificação dos tipos de relações entre conceitos. Mas, para cumprir com o objetivo de navegação, como se propõem as taxonomias navegacionais, talvez o detalhamento excessivo da tipologia relacional cause dificuldades para o usuário final.

6. CONSIDERAÇÕES FINAIS DO CAPÍITULO 8

O trabalho apresenta uma visão geral dos sistemas de organização do conhecimento, o conceito e suas funções na opinião dos principais autores. Nosso foco foi dar ênfase nas classificações, tesauros, taxonomias e ontologias apontando as características de cada om desses sistemas, por meio de exemplos, na tentativa de contribuir para o esclarecimento dos pontos comuns e os que diferenciam dos diversos tipos de sistemas.

Mesmo antes do surgimento da Ciência da Informação, a Biblioteconomia e a Documentação já se ocupavam com a busca de linguagens padronizadas para representação de assuntos, dando origen à época aos cabeçalhos de assuntos e sistemas de classificação. Posteriormente, com o uso de sistemas automáticos de sistemas de recuperação da infotmação, surgem os tesauros. Essas linguagens são conhecidas como linguagens documentárias. É natural que a evolução do conhecimento venha acompanhada de nova tecnologia, o que ocorre, em 1998, com o surgimento do termo Sistemas de Organizaçào do conhecimnto, o que nos causa estranheza é o fato de não se observar, na proposta de nova denominação, nenhuma correlação com as tradicionais linguagens documentárias.

É necessário resgatar o conhecimento acumulado em décadas de pesquisa acerca das linguagens documentárias, para a construção de referencial teórico e metodológico relativo aos SOCs, os quais podem utilizar bases teóricas e metodológicas comuns, oriundas da teoria da classificação e teoria geral da terminologia, entre outras.

Como vimos neste capítulo, alguns SOCs apresentam pontos comuns: compartilham princípios de classificação; utilizam algum tipo de signo para representar os conceitos; cumprem a função de padronizar os termos empregados para representar os conceitos; apresentam recursos para delimitar o significado e estabelecem relações entre conceitos. Observamos, no entanto, que cada sistema possui peculiaridades quanto à estrutura e funções.

Os tesauros, as ontologias e as taxonomias são esquemas de representação 'específicos' ou seja, apesar de terem alguns pontos em comum, não se pode afirmar que um deles 'incorpora' o conceito do outro. Tesauros e ontologias não são taxonomias, porque o princípio da taxonomia é a relação

gênero-espécie ou coisa-tipo e os tesauros e ontologias vão mais além, incorporam outros tipos de relacionamentos. A taxonomia está presente em ambos, mas isto não é suficiente para concluir que ela é o conceito mais abrangente e sim apenas uma característica comum aos três esquemas. Ambos têm uma estrutura muito parecida, composta de categorias e de relações entre os termos. No entanto, as relações entre os termos que constituem um tesauro são estabelecidas por normas, compreendendo a relação de equivalência, hierárquica e associativa. Já nas taxonomias, as relações são estabelecidas segundo critérios de uso.

Cabe ainda ressaltar que as ontologias agregam valor aos demais esquemas de representação, por meio de uma semântica mais profunda e também sob os aspectos conceitual, relacional e tecnológico.

Os quatro sistemas podem ser empregados em ambientes digitais para organizar e recuperar informações em sites, bibliotecas digitais e repositórios. Por outro lado, alguns sistemas não se aplicam ao ambiente tradicional, como é o caso das ontologias, que, por sua vez, são as que apresentam características próprias para uso no contexto da web semântica.

A decisão sobre o uso de um ou outro depende da aplicação pretendida e nem sempre o sistema mais recente ou mais atual é o mais adequado. Taxonomias resolvem bem a organização de informações em portais corporativos ou de comércio eletrônico, tesauros atendem à indexação e recuperação de informações em bases de dados bibliográficas. As ontologias oferecem maior capacidade de representação semântica, no entanto, seu uso dependerá da capacidade do sistema de informação interpretar e aplicar o conhecimento nelas contido.

Esse trabalho é apenas um ponto de partida para invfestigações mais apuradas (específicas), e para facilitr a integração dos diferentes tipos de sistemas de organização do conhecimento. Assim como melhorar o diálogo entre os especialistas das diversas áreas que operam com esses sistemas.

7. REFERÊNCIAS DO CAPÍTULO 8

AQUINO, I; CARLAN, E.; BRÄSCHER; M. Princípios Classificatórios para a Construção de Taxonomias. *PontodeAcesso,* Salvador, v. 3, n. 3, p. 196-215, 2009. Disponível em: http://www.portalseer.ufba.br/index.php/revistaici/article/view/3626. Acesso em: março 2010.

AMERICAN NATIONAL STANDARDS INSTITUTE / NATIONAL INFORMATION STANDARDS ORGANIZATION. Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. ANSI/NISO Z39.19-2005. NISO Press, Bethesda: Maryland, 2005. Disponível em:<www.iso.org/kst/reports/standards/>. Acesso em: jul. 2008.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*. 2001. Disponível em: http://www.cs.nyu.edu/rgrimm/teaching/reading/semantic-web.pdf. Acessado em: fev. 2010.

BLISS, H. E. *A bibliographic classification*. 2ª ed. New York: H.W. Wilson, 1952.

BRÄSCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9, 2008, São Paulo. *Anais.* São Paulo: ECA/USP. Programa de Pós-Graduação em Ciência da Informação, 2008. ENANCIB. 1835 pdf. Set./out. 2008. CD-ROM.

BROUGHTON, V. et. al. Knowledge Organization. In: *European Curriculum Reflections on Library and Information Science Education*. Copenhagen: Kajberg, L. and Lorring, L. 2004. chapter 7, pages pp. 133-148. Royal School of Library and Information Science, Copenhagen. Disponível em: http://dlist.sir.arizona.edu/1050/. Acesso em: mar 2009.

CAMPOS, M. L. de A.; GOMES, H. E. Taxonomia e Classificação: o princípio de categorização. *DataGramaZero – Revista de Ciência da Informação*, v. 9, n. 4, ago. 2008. Disponível em: <http://www.dgz.org.br/ago08/F_I_art.htm> Acesso em: abril 2009.

CAVALCANTI, C.R. *Indexação e tesauro*: metodologia e técnicas. Brasília, ABDF, 1978.
CONWAY, S.; SLIGAR, C. Building Taxonomy. In: UNLOCKING KNOWLEDGE ASSETS. Washington, Microsoft Press, 2002. 256 p.
CURRÁS, E. *Ontologías, Taxonomías y Tesauros*: manual de construcción y uso. 3a ed. Espanha: Ediciones Trea, S. L. 2005. 337 p.
DAHLBERG, I. Knowledge organization: its scope and possibilities. *Knowledge Organization, 20*(4), 211-222, 1993.
DAHLBERG, I. Knowledge organization: a new science? *Knowledge Organization,* v. 33, n. 4, p. 11-19, 2006.
DAHLBERG, I. Teoria do Conceito. *Ciência da Informação,* Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978.
EDOLS, L. *Taxonomies are what?* Free Pint, Reino Unido, n.97, out. 2001. Disponível em: <http://www.freepint.com/issues/041001.htm#eature>. Acesso em: out. 2008.
GOMES, H. E. *Manual de elaboração de tesauros monolíngues*. Brasília: o Programa, 1990. 78 p.
GREEN, R.; BEAN, C. A.; MYAENG, S. H. *The Semantic of Relationships:* An Interdisciplinary Perspective. Norwell-USA: Kluwer Academic Publishers, 2002. 244 p.
GRUBER, T. (1996). *What is an ontology?* [S. I.: s. n.], 1996. Disponível em: <http://ksl.stanford.edu/kst/what-is-an-ontology.html> Acessado em: abril 2007.
GUINCHAT, C.; MENOU, M. *Introdução geral às ciências e técnicas da informação e documentação*. 2a ed. Brasília: FBB e IBICT, 1994. 540 p.
HJORLAND, B. *Discussion of Heidelberg theory of concepts and knowledge organization (KO)*. Última edição em 26 fevereiro 2007. Disponível em: <http://www.db.dk/bh/lifeboat_ko/CONCEPTS/discussion_of_dahlberg.htm>. Acessado em: março 2010.
HJORLAND, B. *Knowledge Organization Systems (KOS)*. Última edição 2008. Disponível em: <http://www.db.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_systems.htm>Acesso em: fevereiro 2010.
HODGE, G. *Systems of Knowledge Organization for Digital Libraries:* beyond traditional authorities files. Washington, DC, the Council on Library and Information Resources. 2000. Disponível em: <http://www.clir.org/pubs/reports/pub91/contents.htm> Acesso em: 24 maio 2008.
IBICT. *Diretrizes para elaboração de tesauros monolíngues*. Brasília, 1984.
KENT, A.; LANCOUR, H. *Encyclopedia of library and information science*. New York: M. Dekker, 1968.
LANGRIDGE, D. *Classificação*: abordagem para estudantes de Biblioteconomia. 1ª reimpressão. Rio de Janeiro: Interciência, 2006.
MCCULLOCH, Emma; MACGREGOR, George. Analysis of Equivalence Mapping for Terminology Services. Journal of *Information Science,* v. 34, n. 1, p. 70-92, 2008.
MONTGOMERY, C. A. Automated language processing. *Annual Review of Information Science*, v. 4, p. 145-174, 1969.
NOY, N. F.; McGUINNESS, D. L. *Ontology Development 101:* A Guide to Creating Your First Ontology. Stanford University: Stanford, 2001. Disponível em:<http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101- noy-mcguinness.html> Acesso em: maio 2008.
PHILLIPS, W. H. *A primer of book classification*. London: Association of Assistant Librarians, 1955. 235 p.
PIDCOCK, W. *What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model?* Disponível em: <http://infogrid.org/wiki/Reference/PidcockArticle> Acesso em: fevereiro 2010.
REITZ, J. M. ODLIS Online Dictionary for Library and Information Science. Libraries Unlimited. 2004. Disponível em: <http://lu.com/odlis/odlis_t.cfm> Acesso em: 22 out. 2009.
SALTON , G.; McGILL, M.S. *Introduction to modern information retrieval*. New York: McGraw-Hill Book Company. C1990. 448p.

SHIRI, Ali; MOLBERG, Keri. Interfaces to knowledge organization systems in Canadian digital library collections. *Online Information Review,* v. 29, n. 6, p. 604-620, 2005.

SOERGEL, D. The rise of ontologies or the reinvention of classificatio n. *Journal of the American Society of Information Science*, v. 50, n.12, 1999, p. 1119 -1120.

SOWA, J. F. *Conceptual structures: information processing in mind and machine*. Massachussetts: Addison-Wesley, 1984. 435p.

USCHOLD, M. *What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model?* Disponível em: <http://infogrid.org/wiki/Reference/PidcockArticle> Acesso em: fevereiro 2010.

VICKERY, B. C. *Knowledge representation:* a brief review. Journal of Documentation. v. 24, n. 3, p.145-159, 1986.

VICKERY, B. *On 'knowledge organisation'*. Última edição 2008. Disponível em: <http://www.lucis.me.uk/knowlorg.htm#start>. Acesso: 30 mar. 2010.

W3C. *Simple Knowledge Organization System* - SKOS. 2004. Disponível em: <http://www.w3.org/2004/02/skos/intro>Acesso em: março 2010.

ZENG, M. L.; SALABA, A. Toward na International Sharing and Use of Subject Authority Data. IN: FRBR WORKSHOP, OCLC, 2005. Kent State University. Disponível em: <www.oclc.org/research/events/frbr.../zeng/Zeng_Salaba.ppt... >Acesso em: março 2010.

ZENG, Marcia Lei. Sharing and use of subject authority data. *International Cataloguing and Bibliographic Control*, v. 35, n. 3, p. 52-54, Jul-Sep., 2006.

ZENG, Marcia Lei. Knowledge Organization Systems (KOS). *Knowledge Organization,* v.35, n. 2-3, p. 160-183, 2008.

Capítulo 9

Categorização lingüística como esteio da organização do conhecimento

Fabio José Dantas de Melo[1]

Resumo do Capítulo 9

O processo de *categorização lingüística* permeia a vida de modo geral. Sua natureza processual descarta sua concepção como *produto* formado *a priori*, estático ao longo do tempo e do espaço; ao contrário, cabe ao indivíduo a realização de constantes "reformulações" em seu aparato cognitivo, a fim de ajustar uma "visão particular de mundo" a uma "visão comunal, socializada de mundo" o que torna a convivência possível e rica. Como categorização lingüística pode ser definida no binômio *cosmovisão* + *linguagem*, a seleção lingüística para representar conceitos e idéias de determinado domínio do saber se mostra, sim, como uma prática institucionalizada (na qual se lança mão de vocabulários controlados), mas fortemente influenciada por uma variável cultural, popular. Neste artigo, a preocupação básica foi a conceituação deste processo lingüístico e das contribuições que o seu estudo pode trazer a área da Ciência da Informação, preocupada em compreender as operações de classificação e organização que os vários segmentos sociais e profissionais fazem do conhecimento que lhes é atinente. A implicatura existente entre *categorias cognitivas* e *categorias lingüísticas* deixa entrever que estamos diante de um assunto que não pode ser bem entendido a não ser de uma abordagem multidisciplinar.

Palavras-chave: Categorização linguistica; Organização do conhecimento; Relatividade lingüística; Semântica cognitiva

Abstract

Linguistic categorization as support of the knowledge organization

The process of linguistic categorization permeates life. Its processual nature rejects its comprehension as a preconceived product, static throughout time and space; On the contrary, the individual is always undergoing constant reformulations in the cognitive apparatus in order to adjust a particular point of view to a communal and socialized one, which makes life in society possible and fruitful. Once linguistic categorization can be defined as cosmovision + language, the linguistic selection used to represent certain concepts and ideas from specific knowledge fields is an institutionalized practice (in which controlled vocabulary is present) but strongly influenced by culture. In this article, the main concern was the conceptualization of this linguistic process and its contributions to the Informational Science, whose goal is to understand classification and organization operations made by the panoply of social and professional segments regarding their knowledge. The relation between cognitive categories and linguistic categories elicits that this is a subject which cannot be comprehended but through a multidisciplinary approach.

Keywords: Linguistic categorization; Knowledge organization; Linguistic relativity; Cognitive semantics

Como citar o Capítulo 9

MELO, Fabio José Dantas de. Categorização lingüística como esteio da organização do conhecimento. *In:* Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento.* Brasília DF: IBICT, 2010, 335 p. ISBN: 978-85-7013-072-3. Capíitulo 9, p. 177-182. Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

[1] Doutor em Lingüística, Pós-doutor em Ciência da Informação pela Faculdade de Ciência da Informação– FCI da UnB, Pesquisador Associado Júnior do Programa de Pós-Graduação em Ciência da Informação - PPGCInf, Bolsista do CNPq de PDJ. CV Lattes: http://lattes.cnpq.br/1542927746260681. E-mail: melodantas@ig.com.br.

> **21** *Criou, pois, Deus os grandes animais marinhos e todos os seres viventes que rastejam, os quais povoavam as águas, segundo as suas espécies; e todas as aves, segundo as suas espécies.(...)* **24** *Disse também Deus: Produza a terra seres viventes, conforme a sua espécie: animais domésticos, répteis e animais selváticos, segundo a sua espécie.* **26** *Também disse Deus: Façamos o homem à nossa imagem, conforme a nossa semelhança; tenha ele domínio sobre os peixes do mar, sobre as aves dos céus, sobre os animais domésticos, sobre toda a terra e sobre todos os répteis que rastejam pela terra.*
>
> *Gênesis 1: 21;24;26*

> *A língua (...) se pode considerar um complexo de símbolos refletindo todo o quadro físico e social em que se acha situado um grupo humano*.
>
> *Edward Sapir*

Gastamos tempo de nossas vidas organizando papéis que se amontoam em nossa mesa de trabalho, ou o mobiliário de nossa casa a fim de torná-la mais funcional ou acolhedora, ou mesmo estabelecendo critérios de arrumação das peças de nosso vestuário, de modo que toda interferência externa no produto desta nossa atividade é perceptível e, na maioria das vezes, requer a pronta restauração da ordem. Processo recorrente na vida diária dos seres humanos, o ato de **nomear** e **organizar** objetos, seres e fenômenos ao nosso redor responde a um apelo de nossa mente por referenciais que dêem sentido a caleidoscópica realidade. Nesse sentido, é bem ilustrativa a imagem, proposta pelo escritor Jorge Luis Borges, do mundo como uma *biblioteca de simetrias sinistras*.

Este capítulo visa a refletir sobre o problema da representação do conhecimento no ser humano e os desdobramentos deste processo, a que damos o nome de **categorização**[1], na vida em sociedade. Categorização constitui um mecanismo de organização da informação [por generalização ou abstração e discriminação] obtida a partir da apreensão da realidade. Como explicam Cuenca & HIlferty (1999, p.32), "mediante a categorização agrupamos elementos diferentes em conjuntos, o qual nos permite pensar, perceber, agir e, inclusive, falar".

Tomarei a Linguística Cognitiva (LC), modelo teórico surgido no final da década de 80, como parâmetro por sua vinculação a um ponto de vista filosófico denominado *experiencialismo* (ou r*ealismo experiencial*) que difere do *objetivismo* praticado por outras linhas investigativas, para as quais "o pensamento é uma manipulação mecânica de símbolos abstratos, que adquirem seu significado por correspondência direta com o mundo exterior" (CUENCA; HILFERTY, 1999, p.15). Para os experiencialistas, a linguagem não é um sistema estanque na mente, formado de estruturas estáticas e estabelecidas *a priori*, mas "(...) como o resto das capacidades cognitivas humanas, está fundado na experiência do mundo" (*Idem*, p.17). E disto advém uma das assunções-chave da LC de que **a conceptualização encontra-se condicionada pela experiência de nosso corpo[2], do mundo externo e de nossa relação com o mundo.**

De fato, o homem não é uma ilha. E suas ponderações acerca do universo físico, sobretudo a respeito de seu nicho sociocultural, demandam legitimação coletiva, grupal. Como afirmara Saussure na primeira metade do século XX (SAUSSURE, 2001), a língua é uma instituição social que funciona como um conjunto de convenções determinado por uma espécie de *contrato social* firmado por indivíduos de um mesmo grupo. Para Lévi-Strauss

[1]Este conceito foi escolhido por ser inerente ao processo de organização e representação da informação, haja vista sua apropriação pela teoria da classificação ao longo dos tempos.

[2] Lembre-se que boa parte das medidas espaciais mais antigas foram estabelecidas pelas diversas culturas ao redor do mundo com base em partes do corpo. Assim, temos 24 **pés** de espessura, 300 **cúbitos** de altura, 35 **polegadas** e outras.

(1996), a língua é produto e, simultaneamente, condição de cultura, pois é, por meio dela, que se transmitem valores, rituais, ofícios e interdições... Significa que, ao longo da vida, toda "(re)modelagem do real" operado pelo e no indivíduo sofre a influência da coletividade. Exemplo explícito da força do social sobre as individualidades encontra-se no fenômeno de mudança lingüística, em que toda inovação no sistema-língua introduzida por um indivíduo ou grupo de indivíduos (denominados na teoria de *target group* segundo Thomason & Kaufmann (1988)) está condicionada – para sua efetivação – ao crivo da comunidade por meio de seu uso freqüente. Assim, poderíamos generalizar dizendo que **alterações na ordem do real e nas representações deste real exigem o compartilhamento pela maioria dos membros de uma organização, sociedade, cultura** etc.

Dahlberg (1978, p.102) compreende *conceito* como "a compilação de enunciados verdadeiros sobre determinado objeto". Donald Davidson (*apud* Marcuschi, 2004, p. 5), formulador da *Coherence Theory of Truth* (Teoria Coerencial da Verdade), recusa uma **linguagem** e um **conhecimento privados**. Para ele, a objetividade (em outras palavras, a realidade extralingüística) tem sua fonte na intersubjetividade o que corrobora a afirmação expressa no final do parágrafo anterior. A verdade de um enunciado equivale à sua coerência com as outras crenças, entendido *crenças* não como "simples fruto de uma subjetividade privada, mas de uma comunidade de mentes" (MARCUSCHI, 2004, p. 5). As pessoas concordam intersubjetivamente porque **classificam** e **organizam** o mundo de forma parecida quando vivem na mesma cultura, afirma Marcuschi ao tratar do trinômio Linguagem, Cultura e Categorização. Basta recordarmo-nos de um supermercado, onde a busca por um produto A exige de o cliente seguir as mesmas categorias mentais, que nortearam a segmentação do espaço físico do estabelecimento em grupos de alimentos ou produtos. Caso resolvesse guiar-se por uma classificação particular, que não incluísse "queijos" como pertencentes à seção de laticínios, a tarefa de encontrar o que procura seria demasiadamente onerosa senão impossível.

Outro aspecto desta questão é que **o real é composto de miríades de eventos e entes e todo esse volume informacional precisa ser *estruturado, sistematizado*,** por meio do agrupamento de seus elementos em **classes** e **subclasses** segundo caracteres comuns, de forma a facilitar o acesso a eles. É caricata a atitude do personagem Irineu Funes, o Memorioso, saído da pena de Borges (1944). Sua preocupação era projetar um idioma "em que cada coisa individual, cada pedra, cada ave e cada ramo tivesse um nome próprio". O problema de Funes, como se deduz da citação, era sua incapacidade de idéias gerais. "Não só lhe custava a compreender que o símbolo genérico *cão* abrangesse tantos indivíduos díspares de diferentes tamanhos e diferente forma; incomodava-o que o cão das três e catorze (visto de perfil) tivesse o mesmo nome que o cão das três e um quarto (visto de frente)". Pobre Funes, obstinado em nomear diferentemente "o mesmo ser" visto em horários diversos, esquece-se de que é vital para a memória humana trabalhar com moldes que reduzam a multiplicidade a um finito número de categorias básicas, sob as quais possam, aí sim, agrupar-se a infinidade de elementos.

Esta situação extraída de uma obra de ficção revela que o conceito lingüístico de "cachorro" não é independente de nosso conhecimento acerca de cachorros. Este conhecimento, explica-nos Taylor (2003), não apenas diz respeito a nossas crenças sobre o *status* de cachorros como uma espécie do reino animal, mas também nosso conhecimento sobre as diferentes raças de cachorro. "Este tipo de conhecimento é a **matriz de nossa representação mental** do protótipo[3] de cachorro, e fornece uma base racional para delimitar uma coleção de criaturas que podem ser referidas como cachorros" (TAYLOR 2003, p. 86).

[3] Segundo Rosch, os conceitos são estruturados de forma gradual, havendo um membro típico ou central das categorias (o chamado **protótipo**) e outros menos típicos ou mais periféricos. Por exemplo, na categoria *ave*, teremos alguns membros mais centrais, ou seja, prototípicos, como "periquitos" e "papagaios", que têm a maioria das propriedades associadas a *aves*; e outros mais periféricos como "pingüins" (*apud* Cançado, 2005).

Subsidiária ao exercício classificatório, a tarefa de atribuir um rótulo aos subconjuntos de fatos da realidade pressupõe um ponto anteriormente levantado (sobre **a convencionalidade das denominações lingüísticas)** e remete-nos a outro, o da **variedade lingüística.** Um dos princípios da comunicação eficaz é o emprego de um código comum entre o emissor e o destinatário. Mas língua alguma é um objeto monolítico, homogêneo. Ela é uma tapeçaria de variedades históricas, geográficas, sociais e de estilo. Segundo Oliveira (2001, p.110), em estudo acerca dos "regionalismos brasileiros", [...] "podemos verificar, mormente no âmbito do *léxico*, "marcas" regionais por vezes bastante afastadas entre si devido, principalmente, às grandes extensões geográficas de nosso território, ao isolamento em que se encontram algumas de nossas regiões e, em alguns casos, à influência de povos procedentes de outros pontos da Europa."

Além do mais, cada língua faz um "recorte" diferente da realidade. Isto foi apresentado por Edward Sapir e seu aluno, Benjamin Lee Whorf, sob a denominação de *relatividade lingüística*. Sapir, em seu célebre artigo *Língua e Ambiente*, sustentava que "o **léxico** da língua é que mais nitidamente reflete o ambiente físico e social dos falantes" (SAPIR, 1969, p.45). A relatividade lingüística, em certa medida, contempla a teoria da coerência, de Davidson (*apud* MARCUSCHI, 2004, p. 6), quando esta afirma: "A verdade é constituída pela *correspondência coerente entre crenças* e não pela correspondência de uma crença com algo externo" (grifo meu). Tenhamos consciência ou não, o mundo é percebido pelo indivíduo através da arquitetura de sua língua. Umberto Eco, com base no relativismo lingüístico, acredita que **cada língua organiza o seu mundo lexical de acordo com suas necessidades práticas e, conseqüentemente, considera como relevantes diferentes aspectos do mundo.** Note a riqueza vocabular, a especialização lexical, de determinadas línguas em certos domínios quando contrastadas com outras (o conjunto de termos para a cor branca dos *Inuits*, os vários nomes de ventos existentes na língua de povos do deserto, ou os termos precisos para muitas espécies de animais marinhos encontrados na fala dos índios *nutka*). A existência deste fato lingüístico corrobora o que, em metodologia científica, se afirma da realidade: de que *não é nem verdadeira, nem falsa*, mas sim o julgamento ou inferência que dela faz o observador.

Um desdobramento da idéia tratada acima é que um grupo humano passará ao longo do tempo por transformações em seu *modus vivendi*, muitas vezes como resposta a mudanças na área tecnológica, educacional, política e espiritual. O "conhecimento novo" produzido pelo grupo acarreta mudanças culturais de maior ou menor proporção. Nesse sentido, "somos forçados a admitir", com Sapir (1969, p.61) "que uma complexidade rapidamente crescente da cultura exige mudanças correspondentes, embora não tão rápidas, *na forma e conteúdo lingüístico*", ou seja, em seu Léxico. Isso aponta para alterações nos significados das palavras (estreitamento e alargamento semânticos, por exemplo), bem como nas fronteiras dos campos lexicais, das quais deve estar ciente o analista da informação.

Toda essa discussão insere-se numa das disciplinas da Ciência da Informação, essencial no processo de representação da informação, que é a **Organização do Conhecimento**. Em Ciência da Informação, muito se tem discutido sobre os termos "informação" e "conhecimento". A proposta conceitual de Bräscher & Café (2008) [4]quanto à diferenciação de *Organização do Conhecimento* e *Organização da Informação* tem por base as características identificadas por Fogl (1979) sobre o que sejam *informação* e *conhecimento*. Este autor considera a informação segundo o ponto de vista semântico ("[...] avalia a informação em termos de sua veracidade, confiabilidade, conhecimento, adequação dos juízos de valor e assim por diante" [5]) e o pragmático ("o valor da informação depende do significado particular atribuído a ela pelo receptor desta informação, [...] segundo um determinado propósito" [6]); com relação ao método de fixação do conhecimento e dos juízos de valor e, por fim, o suporte material utilizado. Relacionando esses crité-

[4] Comunicação oral apresentada ao GT-02 – Organização e Representação do Conhecimento no IX Enancib
[5] Bräscher & Café (2008, p. 4)
[6] *Ibidem*

rios, podemos compreender a *informação* como algo que existe fora da consciência individual e independente dela, ao passo que o *conhecimento* apresenta-se como resultado da cognição e, portanto, o conteúdo ideal da consciência humana. O suporte físico permite à *informação* ter existência e cumprir sua função social.

Organização da Informação	**Organização do Conhecimento**
Processo que envolve a descrição física e de conteúdo dos objetos informacionais	Processo que se aplica a unidades do pensamento (conceitos) e visa à construção de modelos de mundo que se constituem em abstrações da realidade.
Seu produto é a **representação da informação,** conjunto de elementos descritivos que representam os atributos de um objeto informacional específico	Seu resultado é a **representação do conhecimento** que reflete, assim, um modelo de abstração do mundo real, construído para determinada finalidade.
Mundo dos objetos físicos	*Mundo da cognição ou das idéias*

Tabela construída a partir das conclusões de Bräscher e Café (2008)

A Organização do Conhecimento, como demonstrada acima, por constituir-se em um processo de construção de modelos de mundo, está em estreita conexão com a **categorização lingüística**, na medida em que a estruturação do léxico de uma língua na mente do falante reflete e é condicionada pela sua leitura de mundo. Num léxico, as *categorias cognitivas* mantêm-se em correspondência com as *categorias lingüísticas*. E se "na formação de uma língua é preciso considerar a *influência exercida pelo ambiente através da experiência social,* [pois] o contato entre língua e realidade irá determinar a linguagem como reflexo da realidade e, sobretudo, como força geradora da imagem de mundo que o indivíduo possui" (grifo meu) (OLIVEIRA 2001, p. 109), seria indispensável – da parte do indexador – preocupação em traduzir o conteúdo temático de um objeto informacional sem descaracterizar o ambiente dentro do qual este objeto foi gerado sob pena de comprometer a recuperação da informação pelo usuário, também produto deste mesmo ambiente. Estar familiarizado com tais questões auxiliará o profissional de CI a realizar com mais apuro seu papel na sociedade.

REFERÊNCIAS BIBLIOGRÁFICAS DO CAPÍTULO 9

BORGES, Jorge Luis. *Ficções (1944)***.** São Paulo: Companhia das Letras, 2007.
CANÇADO, Márcia. *Manual de semântica: noções básicas e exercícios*. Belo Horizonte: UFMG, 2005.
CUENCA, Maria Josep & HILFERTY, Joseph. *Introducción a la Linguística Cognitiva*. Barcelona: Editorial Ariel, 1999.
DAHLBERG, Ingetraut. Teoria do Conceito*. Ciência da Informação,* v.7, n.2, 1978.
LÉVI-STRAUSS, Claude. *Antropologia estrutural***.** São Paulo: Cosac Naify, 2008.
MARCUSCHI, Luiz Antônio. A Construção do Mobiliário do Mundo e da Mente: Linguagem, Cultura e Categorização. *In*: *Anais do Congresso de Cognição*, Juiz de Fora MG, 2004

SUMÁRIO_GERAL

OLIVEIRA, Ana Maria Pinto Pires de. "Regionalismos Brasileiros: a questão da distribuição geográfica". *In*: ISQUERDO, Aparecida Negri *et al. As Ciências do Léxico*. Campo Grande, MS: Ed. UFMS, 2001.

SAPIR, Edward. *Lingüística como ciência*. Rio de Janeiro: Livraria Acadêmica, 1969.

SAUSSURE, Ferdinand de. *Curso de lingüística geral*. 23. ed. São Paulo: Cultrix, 2001.

TAYLOR, John R. *Linguistic Categorization*. New York: Oxford University Press, 2003.

THOMASON, Sarah Grey; KAUFMAN, Terrence. *Language contact, creolization, and genetic linguistics*. Berkeley: Univ California Press,1988.

IV. ESTUDOS MÉTRICOS DA INFORMAÇÃO

Capítulo 10 – Metrias da informação: História e tendências – Jaime ROBREDO e Jayme Leiro VILAN FILHO, *p. 184-258*

Capítulo 11 – Rede metodológica integrando epistemologia, organização do conhecimento, bibliometria e tesauros: Concepção e construção do Tesauro Brasileiro de Ciência da Informação – Lena Vania Ribeiro PINHEIRO e Helena Dodd FERREZ, *p. 259-276*

Capítulo 10
Metrias da Informação: História e Tendências

Jaime Robredo[1] **e Jayme Leiro Vilan Filho**[2]

SUMÁRIO DO CAPÍTULO 10

Resumo do Capítulo 10, *p. 185*
Como citar o Capítulo 10, *p. 185*

1. INTRODUÇÃO DO CAPÍTULO 10, *p. 186*
2. SOBREVÔO DA HISTÓRIA DAS METRIAS DA INFORMAÇÃO, *p. 186*
   - 2.1 A corrente européia ocidental, *p. 186*
   - 2.2 A corrente russa e européia oriental, *p. 198*
   - 2.3 A corrente indiana, *p.192*
   - 2.4 A corrente americana, *p. 192*
   - 2.5 A bibliometria no Brasil, *p. 194*
     - 2.5.1 *Um pouco de história para contextualizar*, *p. 195*
     - 2.5.2 *Um pouco de história sobre a bibliometria no Brasil*, *p. 197*
3. TENTATIVA PARA FIXAR A TERMINOLOGIA, *p. 205*
4. METRIAS DA INFORMAÇÃO E CIÊNCIA DA INFORMAÇÃO ANTES DA WEB, *p. 206*
   - 4.1 Lei de Bradford, *p. 207*
   - 4.2 Lei de Lotka, *p. 210*
   - 4.3 Lei de Zipf, *p. 210*
5. METRIAS DA INFORMAÇÃO E CIÊNCIA DA INFORMAÇÃO EM AMBIENTES WEB, *p. 212*
6. APLICAÇÕES RECENTES QUE SINALIZAM TENDÊNCIAS, *p. 213*
   - 6.1 Aplicações bibliométricas na gestão e desenvolvimento de coleções especializadas, *p. 213*
   - 6.2 Aplicações bibliométricas e infométricas usando estatística descritiva, *p. 214*
     - 6.2.1 *Produção e colaboração científica*, *p. 215*
     - 6.2.2 *O gênero das autorias científicas no Brasil*, *p. 216*
   - 6.3 Outras aplicações infométricas, *p. 219*
     - 6.3.1 *Identificação automática de grupos e linhas de pesquisa*, *p. 219*
     - 6.3.2 *Identificação automática de clusters temáticos, ontologias e mapas de tópicos*, *p. 224*
     - 6.3.3 *Análise de citações e outras cientometrias*, *p. 227*
     - 6.3.4 *Mineração de textos em bases de periódicos científicos especializados*, *p. 231*
     - 6.3.5 *Aplicações inferenciais*, *p. 233*
7. CONSIDERAÇÕES FINAIS DO CAPÍTULO 10, *p. 235*
8. BIBLIOGRAFIA DO CAPÍTULO 10, *p. 236*

ANEXO 1. **Aplicação da lei de Zipf à análise de um resumo de 3.100 palavras, procedente de um artigo de um dos autores (ROBREDO *et al.*, 1988)**, *p. 253*

[1] Doutor em Ciências, Pesquisador Associado Sênior e Líder do Grupo de Pesquisa "Estudos sobre Representação e Organização da Informação e do Conhecimento – EROIC". Programa de Pós-Graduação em Ciência da Informação (PPGCI) – Faculdade de Ciência da Informação da Universidade de Brasília (UnB/FCI). CV Lattes: http://lattes.cnpq.br/9669125022187444. E-mail: ssrrinfo@uol.com.br.

[2] Doutor em Ciência da Informação. Programa de Pós-Graduação em Ciência da Informação (PPGCI)Professor da Faculdade de Ciência da Informação, Universidade de Brasília (UnB/FCI). Brasília DF. CV Lattes: http://lattes.cnpq.br/3364096778603273. E-mail: jleiro@unb.br

Resumo do Capítulo 10

Numa releitura da história das metrias da informaçõ em todas suas variantes, o presente Capítulo resgata a contribuição de numerosos pesquisadores da Ïndia, bem como da Europa Oriental e da antiga União Soviética, estes últimos notadamente no domínio da cientometria. O Interesse pelos estudos infométricos no Brasil, e mais particularmente pela bibliometria, nos anos 70-80 do passado século, experimentou posteriormente um declínio significativo, para renascer com nova pujança nos últimos anos, emnumerosas aplicações. A intenção deste longo Capítulo é mostrar, com o auxílio de exemplos concretos, a variedade de aplicações das metrias da informação e, o que é mais importante, "como fazer". Sob uma variedade de nomes – bibliometria, infometria, cientometria, webmetria, etc. – as técnicas infométricas abrem à ciência da informação um brilhante leque de aplicações nos procesos informacionais de representação, organização, gestão, recuperação, planejamento, inferência, tomada de decisão, competitividade, inovação, e todos os desdobramentos políticos, sociais, econômicos, educativos e culturais.

Palavras-chave: Infometria; Cientometria; História; Btasil; Aplicações

Abstract

Infometrics: History ans trends

In a new reading of the history of infometrics in its whole variety, this Capter uncovers the contribution of a number of Indian, as well as East-European and Russian researchers, the last ones mainly in the domain of scientometrics. The interest, in Brazil, on infometrics, and more precisely in bibliometrics, in the decades of the s seventies and eighties of the last century suffered later on a significant decrease by a recent and strong revival in numerous issues. Special attention is paid in this lon Chapter to show, with the support of numerous examples, to the diversity of infometrics uses and, more important, to "how to do it".Under a variety of names – bibliometrics, infometrics, scientometrics, webmetrics, and so one – infometrics opens a wide and briklliant diversity of actual applications in information recording, organizining, managing, processing, retrieving, forecasting, innovating, decision-making, as well as founding social, economic, cuktural and educationa policies.l

Keywords: Infometrics; Scientometrics; History; Brazil; Applications

Como citar o Capítulo 10

ROBREDO, Jaime; VILAN FILHO, Jaime Leyro. Metrias da informação: História e tendências. *In:* Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento.* Brasília DF: IBICT, 2010, 335 p. ISBN: 978-85-7013-072-3. Capíitulo 10, p. 184-258. Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

1. INTRODUÇÃO DO CAPÍTULO 10

Pretende-se neste Capítulo, após um breve sobrevôo da história das metrias da informação e de tentar fixar a terminologia que com elas se relaciona – atualmente, uma verdadeira Babel –, examinar como eram essas metrias e como se relacionavam com a ciência da informação antes de Internet, e como se apresentam hoje em ambientes Web. São mostradas algumas aplicações reais que sinalizam tendências, tais como o monitoramento da produção científica brasileira em ciência da informação a partir das publicações em periódicos da área, mediante a aplicação de técnicas infométricas clássicas na construção e análise de séries históricas que evidenciam as mudanças comportamentais e de gênero entre os autores, ocorridas nos últimos 25 anos, bem como a identificação automática de grupos e linhas de pesquisa, a partir de grandes volumes de dados, e a identificação também automática de *clusters* temáticos, visando a construção de ontologias e mapas de tópicos, em domínios específicos.

Cabe destacar que a intenção dos autores, ao apresentar como exemplo alguns resultados de suas pesquisas é, simplesmente, mostrar o interesse redobrado pela aplicação de técnicas infométricas, tanto bibliométricas como cientométricas, quando se trata de identificar, medir e avaliar informações, identificar atores e ajudar a desenvolver novas pesquisas, que possam beneficiar a sociedade, com maiores facilidades e recursos tecnológicos, num ambiente que se amplia com o crescimento acelerado da Web.

Trata-se, assim, de colocar à disposição de professores e alunos um material que esperamos que possa contribuir para promover uma reflexão em sala de aula e a confirmar o interesse dos métodos quantitativos. Em outras palavras, a orientar, planejar e implementar o choque de gestão de que o País tanto precisa. Hoje, políticas e ações em todas as áreas – administração, economia, educação, infra-estrutura, saúde, inovação e bem-estar social – só dão certo quando fundamentadas em sólidas análises do que se tem e do que falta para ter o que se deseja. Em outras palavras, indicadores situacionais mais confiáveis, para alicerçar decisões. A abundante bibliografia consultada poderá ajudar a identificar pistas para maiores aprofundamentos.

2. SOBREVÔO DA HISTÓRIA DAS METRIAS DA INFORMAÇÃO

Nesta Seção, trataremos de reconstruir o percurso da bibliometria desde seu surgimento, com seus desdobramentos infométricos e cientométricos, até os dias atuais, em que assistimos a um crescimento quase que explosivo no mundo globalizado, aproveitando a abertura de novas trilhas informacionais, que já viraram infovias gigantescas que se entrecruzam e cobrem o espaço virtual da informação e da comunicação planetária para, com as webometrias e cibermetrias, levar o conhecimento de qualquer ponto para qualquer parte, em frações de segundo.

Começaremos esse percurso na Europa ocidental, destacando alguns fatos que nos parecem significativos, sem esquecer a Europa do Leste e a Rússia, passando pela Índia e pousando finalmente nos Estados Unidos. Esse sobrevôo permitirá perceber como, apesar das distâncias e das correntes culturais que as diferenciam, essas regiões, que partiram de pontos diversificados e seguiram caminhos que tanto as aproximaram como as separaram, encontram-se hoje – junto com outros parceiros – reunidos no mesmo movimento que privilegia a pesquisa colaborativa e o intercâmbio de idéias, em benefício de todos.

E, como não poderia deixar de ser, na última parte da Seção, voltaremos nosso olhar para o Brasil, seguindo o caminho aqui percorrido pela bibliometria/infometria e outras metrias, e tratando de ver onde estamos e para onde nos dirigimos.

2.1 A corrente européia ocidental

Como avançado acima, começaremos nosso sobrevôo pela França, continuando pelo Reino Unido, Bélgica, Holanda e Alemanha. Segundo Barts, em sua recente tese de doutoramento (BARTS 2008, p.17-18):

> *"No vasto domínio das ciências da informação e da comunicação (CIC), a infometria consiste na aplicação de modelos e métodos matemáticos e estatísticos à análise da informação. Dois de seus campos de aplicação são a 'vigília tecnológica' e a avaliação da pesquisa"* (GRIVEL 2000).

E, citando Devals, prossegue:

> "Nos dois campos, a **bibliometria** apresenta-se como uma das principais ferramentas: "É preciso saber delimitar as necessidades, organizar a coleta de informações, analisar estas e extrair a informação estratégica que será interpretada pelos especialistas e transmiti-la aos a quem cabe tomar as decisões" (DEVALS 1992). A **bibliometria** é a aplicação desses métodos a um corpus de informações elaboradas com a finalidade de extrair relações significativas entre seus diversos elementos. Por informações estruturadas entendemos as informações textuais estruturadas, como os registros bibliográficos ou de patentes.

A **cientometria** é a aplicação desses mesmos métodos à informação científica e técnica (ICT), e mais particularmente mediante a análise das publicações científicas. Esse termo designa, também, de forma mais geral a aplicação de métodos estatísticos a dados quantitativos (econômicos, sociais bibliográficos) que caracterizam a situação da ciência: segundo Van Raan, "é o campo de pesquisa onde se utilizam métodos e técnicas matemáticas, estatísticas bem como a análise de dados para reunir, processar, interpretar e prever diversas características tais com a performance, o desenvolvimento e a dinâmica da ciência e da tecnologia" (VAN RAAN, 1988). Nesse sentido, **cientometria** é a ciência da ciência (uma metaciência ou supraciência). Entretanto, o termo se usa com freqüência num sentido mais restrito como equivalente a bibliometria. O termo **bibliometria** foi introduzido pelo bibliógrafo e bibliólogo Paul Otlet (Otlet 1934), mas não se generalizou até 1969, com Estivals (ESTIVALS 1969) e Pritchard (PRITCHARD 1969) a quem é geralmente atribuído. O termo **cientometria** foi introduzido, também em 1969, por de Solla Price (PRICE 1969). O termo **infometria** foi adotado em 1987 pela Federação Internacional de Documentação (FID).[3] O termo engloba tanto a **bibliometria** como a **cientometria**".

Barts observa, ainda *"Dentre esses três termos, o vocabulário não foi ainda bem fixado, sendo usado com freqüência pelos autores indistintamente".* Barts se mostra favorável ao uso preferencial do termo *bibliometria* que, segundo ele, tem a vantagem de lembrar o trabalho com publicações e revistas (científicas).

Polanco (1993), situando os campos 'métricos' (*info-, biblio-, ciento-metria*) na esfera global dos estudos da informação, destaca suas relações com o aspecto científico, como mostra a *Figura 1,* que sugere que a infometria se insere num certo sistema científico e que a bibliometria, aplicada à informação científica e técnica, passa a ser o que se chama de cientometria. O mesmo autor situa os modelos empíricos matemático-estatísticos, de que falaremos mais adiante – Lei de Lotka, que descreve a produtividade dos autores científicos, Lei de Bradford, que descreve a distribuição da produtividade das publicações periódicas (número de artigos publicados sobre um tema), e Lei de Zipf, que descreve a ocorrência das palavras em um texto – na interface entre a teoria da bibliometria e a infometria.

Figura 1. – Relação das metrias da informação com a tecnologia da informação e da comunicação (POLANCO, 1993).

[3] Barts usa o termo *infométrie* quando se refere à adoção do termo *informetrics* pela FID. Mais adiante trataremos do assunto das diferentes grafias utilizadas em diferentes línguas, para designar os mesmos conceitos.

Uma representação das relações entre a ciência da informação – e as metrias da informação – e outras ciências tanto exatas e naturais como humanas e sociais, assim como com as tecnologias da informação e da comunicação foi apresentada por Le Coadic (1994). Essa representação, embora mais elaborada, conserva uma visão bastante coincidente com a visão de Polanco, no que diz respeito aos temas principais.

A informação bibliográfica – ou mais geralmente a informação textual – é o material de base da bibliometria. *"Dentre os tipos de informação mais acessíveis, as referências bibliográficas mais ou menos codificadas e/ou detalhadamente indexadas, são o material preferencial..."* (ROSTAING 1996, *apud* BARTS, *loc. cit*). O tratamento da informação textual utilizando métodos estatísticos e de análise de dados permite a classificação, estruturação e valorização da informação, assim como a elaboração de instrumentos de apoio à tomada de decisão.

> "A bibliometria foi utilizada inicialmente em bibliologia e biblioteconomia. Em um segundo tempo alguns sociólogos americanos aplicaram as técnicas bibliométricas ao estudo de fenômenos sociais relativos à comunidade científica. Os resultados desses trabalhos foram posteriormente desviados de seu objetivo inicial, passando a ser considerados pelas instâncias superiores como indicadores para a avaliação das pesquisas públicas. Os trabalhos bibliométricos orientaram-se então para a elaboração de macro-indicadores de avaliação da pesquisa científica e tecnológica em escala internacional" (ROSTAING 1996).

Isso acontece, segundo Barts, mediante a aplicação de técnicas bibliométricas em grandes bases de dados que reúnem diferentes publicações. Como todo estudo bibliométrico, a avaliação ou o auxílio ao monitoramento da pesquisa precisa seguir várias etapas; pode-se distinguir esquematicamente:

- a coleta de informações nas bases de dados bibliográficas representativas do domínio a estudar,
- a padronização dessa informação,
- os tratamentos bibliométricos,
- a análise, a interpretação e a apresentação dos resultados (BARTS 2008).

A vigília tecnológica (*veille technologique*) é uma atividade que aplica técnicas de aquisição, armazenagem e análise de informações relativas a umproduto ou processo, sobre o estado da arte e a evolução de seu conteúdo científico, técnico, industrial ou comercial com a finalidade de reunir, organizar e, postriormente, analisar e difundir as informações pertinentes que permitirão antecipar a evolução, facilitarão a inovação. Pode-se alimentar de dados procedentes de fontes e métodos de coleta diversos (por exenmplo: publicações, patentes). Algumas de suas características são:

- recupera informações abertas, acessíveis a todos, e não de uso interno da empresa,
- não e interessa exclusivamente pelas informações procedentes de uma empresa só.

A vigília estratégica (*veille stratégique*) engloba geralmente a vigília tecnológica. A vigília tecnológica é uma atividade preparatória da transferência de tecnologia (WIKIPEDIA, 2009; POLANCO, 2001; ROSTAING, 1993).

Duas grandes frentes de pesquisa sobre metrias da informação se destacam na França : o *Institut de l'Information Scientifique e Technique,* do *Centre National de la Recherche Scientifique* (CNRS/INIST) que muito deve ao impulso de Polanco, bem conhecido na América Latina, inclusive no Brasil[4], e a *Université Paul Cézanne de Aix-Marseille III, Faculté des Sciences et Techniques*, onde os professores Henri Dou e Luc Quoniam são, também, bem conhecidos no Brasil, assim como Michel Zitt, premiado, em 2009, juntamente com o húngaro Péter Vinkler (ver *Nota 12*), com a medalha Derek de Solla Price, da Sociedade Internacional para Cientometria e Infometria.[5] Deve-se também lembrar o trabalho pioneiro de J.P. Courtial e M. Callon sobre a probabilidade de co-ocorrência de termos e sobre indicadores de temas estratégicos em frentes de pesquisa.[6]

---

[4] Ver, por ex: o notável trabalho *"Aux sources de La scientométrie"* (POLANCO, 1995) onde é apresentada uma pormenorizada síntese da imfometria e dabibliometria até os primórdios do século XXI, numa visão revisitada do pensamento de Derek de Solla Price (1963). Ver t6ambém Polanco (2001) "*Text-Mining e intelligence économique: Aujourd'hui et demain".* Colloque Veille Technologiq, Intelligence Économique ET Bibloiométrie, Université Catholique de Louvain-la-Neuve, Belgique, 23-24 janvier, 9 p.; Polanco (2000) "*Hacia la construcción de una nueva base de datos bibliográficos y de citas para la producción de indicadores bibliométricos (resumen)*", II Taller Bibliometria, CINDOC-RICYT, Madrid. 1 p.; Polanco (1999) "*Transformación de la información en conocimiento y del conocimiento en decisiones estratégicas, IV Taller de Indicadores de Ciencia y Tecnologia*", RICYT, 12-14 de julio, México, D.F., 6 p.

[5] Ver, p. ex.: DOU (2008) e QUONIAM; TARAPANOFF; ARAÚJO JÚNIOR; ALVARES (2001). Michel Zitt é pesquisador sênior no *Institut National pour la Recherche Agronomique* (INRA), em Nantes, e assessor em cientometria e indicadores sobre pesquisa e desenvolvimento, no Observatório das Ciências e Técnicas, Paris (OST). O Observatório publica cada dos anos o *OST Report em Science e Technologie*, disponível em: http://www.obs-ost.fr/en/know-how/etudes-en-ligne/studies-2008/biennial-report-2008-edition.html#c716 (consultado em agosto de 2009).

[6] Ver, p. ex.: CALLON, M.; COURTIAL, J.P. ; LAVILLE, F. (1991), COURTIAL, J.P.; CALLON, M. (1991) e CALLON, M.; COURTIAL, J.P.; SIGOGNEAU, M. (1994).

Para não alongar demais a relação de nomes ilustres dos estudiosos de influíram significativamente no desenvolvimento das metrias da informação, nos limitaremos a citar, além do nome de Derek de Solla Price (Reino Unido) – considerado na Europa ocidental e nos Estados Unidos como o pai da cientometria – os nomes dos ganhadores da Europa ocidental do prêmio Derek de Solla Price, instituído pela *Sociedade Internacional para Cientometria e Informetria*: São eles:

- *No século passado*: Bertram C. Brookes (Reino Unido), em 1989; Anthony F.J. Van Raan (Holanda), em 1995; Ben Martin (Reino Unido) e Joh Irvine (Reino Unido) compartilharam o premio em 1997; Henk F. Moed (Holanda) e Wolfgang Glänzel (Alemanha/Hungria), em 1999 (ver fim da *Nota 7*). São indicados alguns endereços e referências, que permitem conhecer suas trajetórias e localizar algumas de suas publicações mais significativas.[7]
- *No novo milênio*: Leo Hegghe e Ronald Rousseau e (ambos belgas) em 2001; Loet Leydesdorff (Holanda), em 2003; Peter Ingwersen (Dinamarca), em 2005; Michel Zitt (França), citado anteriormente, indicando alguns endereços e referências que permitem conhecer suas trajetórias e localizar suas publicações mais significativas.[8]

---

[7] *Derek de Solla Price* (1922-1983) obteve um PhD em física experimental, na Universidade de Londres em 1946. Após três anos lecionando em Cingapura, retornou a Inglaterra, onde obteve um segundo doutorado em história da ciência. Trabalhou no *Smithsonian Institute*, e lecionou história da ciência na Universidade de Yale até sua morte. Considerado o "pai da cientometria". Em 1976, foi premiado por seu artigo "*A general theory of bibliometric and other cumulative advantage processes*", considerado o melhor artigo publicado nesse ano pelo *Journal of the American Society for Information Science* (PRICE, 1976). Mundialmente reconhecido como pioneiro dos estudos sobre sociologia da ciência, recebeu a título póstumo, em 1984, o *ASIS Research Award* (Prêmio de Pesquisa da *American Society for Information Science*). [Dados parcialmente baseados na breve biografia de Price inclusa na *Home Page* de Rousseau (2009)]

Bertram C. Brookes (1910-1991) lecionou matemática e estatística, no Departamento de Engenharia, na *University College* de Londres e mais tarde ciência da informação, na *City University*, também em Londres. Pioneiro da ciência da informação e um dos fundadores da infometria, foi, ao mesmo tempo um estudioso da filosofia da ciência e dos métodos quantitativos aplicados no campo científico (SHAW, 1990). Ver, p.ex.: BROOKES (1970, 1977 e 1980).

Henk F. Moed é Pesquisador Sênior, Mestre em Matemática, em 1978 na Universidade de Amsterdam; Doutor em Estudos da Ciência PE Universidade de Leiden, em 1989. É professor visitante da Universitdade of Granada. (MOED, 2008).

Anthony F.J. Van Raan é professor de estudos quantitativos da ciência na Universidade de Leiden. Diretor do Centro para Estudos em Ciência e Tecnologia (CWTS, sigla em holandês), da Universidade de Leiden desde 1985. Principais tópicos de pesquisa: bibliometria, medidas de performance da pesquisa, mapeamento da ciência e tecnologia e de sua interface. Editor do periódico *Research Evaluation*, que inclui pesquisas bibliométricas sobre a integração de estudos avaliativos e políticas. (Ver, p.ex.: http://www.issi-society.info/price.html).

Ben Martin é físico, professor de estudos sobre políticas em ciência e tecnologia, e diretor da *Science and Technology Policy Research Unit* (SPRU), na Universidade de Sussex. Criado em 1966, SPRU é hoje um leader global em pesquisa, consultoria e ensino em política e gestão de ciência, tecnologia e inovação, com forte influência sobre a academia, responsáveis pela tomada de decisão no governo e gestores da indústria, bem como pioneiro em abordar o estudo da governança da ciência, tecnologia e inovação. Martins foi premiado em 1997, juntamente com John Irvine, com a medalha Derek de Solla Price. (Para maiores informações, ver, por exemplo: http://www.issi-society.info/price.html e http://www.sussex.ac.uk/spru/).

John Irvine é sociólogo. Incorporou-se ao SPRU em 1972. Juntamente com Martin trabalhou num projeto para estudar a convergência dos indicadores em uso para avaliar o desempenho da pesquisa científica. Seu trabalho chamou a atenção pelo pioneirismo de suas idéias – embora no faltassem detratores entre os setores mais conservadores – sendo estimulado a aprofundar sua linha de pensamento pelo próprio Derek de Solla Price. Eles cunharam a expressão "*research foresight*", para indicar a possibilidade de se chegar a uma visão antecipada do que deveria ser a pesquisa (MARTIN; IRVINE, 1988 e IRVINE; MARTIN, 1989). Seus trabalhos são referência para o planejamento das políticas científicas, tecnológicas e de inovação. Ver, p.ex.: Pavitt (1997).

Wolfgang Glänzel trabalhou com Tibor Braun em Budapeste (ver *Nota 12*), na Biblioteca da Academia Húngara de Ciências. Membro da *Alexander von Humboldt Foundation*, é tesoureiro da *International Society for Scientometricas and Informetrics* (ISSI), editor chefe de *ISSI e-Newsletters* e co-editoir do periódico *Scientometrics*. Ver, por exemplo: http://74.125.155.132/search?q=cache:umzzMO9GuU4J:www.steunpuntoos.be/wg/index.html+Wolfgang+Gl%C3%A4nzel&hl=pt-BR&gl=br&strip=1.

[8] *Leo Egghe* é Professor e Bibliotecário-Chefe no Centro Universitário de Limburgo (LUC). Ensina, também na Universidade de Antuérpia, no Pograma de Biblioteconomia e Ciência da Informação. Em 1987, Egghe e Rousseau organizaram a primeira Conferência Internacional sobre Bibliometria no referido Centro Universitário de Limburgo, à qual deram prosseguimento as seguintes (com nomes que variavam ao gosto dos sucessivos organizadores): London (Ontario, Canadá), em 1989; Bangalore (India), em 1991; Berlin (Alemanha), em 1993 (quando, por ocasião da criação da *International Society for Scientometrics and Informetrics* (ISSI), passou a se denominar *International Conference on Bibliometrics, Informetrics and Scientometrics*, sendo que na Conferência seguinte, o termo *Bibliometrics* foi suprimido, passando a se chamar *International Conference of the International Society for Scientometris and Informetrics*); River Forest (USA), em 1995; Jerusalem (Israel), em 1997; Colima (Mexico), em 1999; Sydney (Austrália), em 2001; Beijing (China),em 2003; Estocolmo (Suécia), em 2005; Madri (Espanha), em 2007; e Rio de Janeiro (Brasil). Os trabalhos de Egghe, desde 1984, deram lugar a numerosos artigos publicados, entre outras, nas seguintes revistas: *Journal of the American Society for Information Science and Technology, Scientometrics, Information Processing and Management, Mathematical and Computer Modelling,* e *Journal of Information Science.* Não poderiamos deixar de mencionaras duas obras nucleares de Egghe, em colaboração com Rousseau: "*Introduction to informetrics"* e "*Elementary Statistics for effective Library and Information Management"* (EGGHE; ROUSSEAU, 1990 e 2001). Para maiores detalhes: http://lspserver.lsp.luc.ac.be:8088/bib/esource.nsf/bcb4207a18ae751ac1256c0e00264b08/600708b520815d8ac1256c0e002ea1a0?OpenDocument#The%20research%20project%20%22Informetric

Ronald Rousseau é Professor Associado Professor da *Catholic School for Higher Education*, Bruges-Ostend e Professor Convidado da *School for Library and Information Science,* da Universidade de Antuérpia. A quantidade de trabalhos de diversos tipos publicados por Rousseau desde 1985 até hoje, ultrapassa os 150, muitos dentre eles em co-autoria com Egghe (ver acima) e pesquisadores de diversos países. A título de curiosidade, vale destacar que existem mais de cinqüenta trabalhos de sua autoria, publicados na China, sejam eles em colaboração com autores chineses ou traduções de outros trabalhos publicados em outros países. Mais informações em http://www.issi-society.info/price.html. Ver curriculum vitae e lista completa de publicações em Rousseau (2009).

- Outros nomes europeus, como Brian Campbell Vickery, inglês, e Gernot Wersig, alemão, embora não fazendo parte do grupo dos europeus honrados com a medalha Derek de Solla Price, não poderiam deixar de ser lembrados.[9]

2.2 A corrente russa e européia oriental

Na Rússia, por razões históricas bem conhecidas, os estudos sobre metrias da informação têm se desenvolvido de forma bastante independente, por não dizer isolada, sendo os contatos entre os pesquisadores Americanos e da Europa ocidental, e seus colegas da Europa oriental e da Rússia, limitados a contatos esporádicos, principalmente por ocasião de alguns congressos internacionais, que raramente conduziam a intercâmbios e colaborações significativos.

Brookes, numa brilhante e oportuna apresentação na 1ª Conferência Internacional sobre Bibliometria e Aspectos Técnicos da Recuperação da Informação, realizada na Bélgica, em 1987, intitulada "*Biblio-, Sciento-, Infor-metrics??? What are we talking about*" (BROOKES, 1989) desenha o quadro em que se desenvolveram as pesquisas dos estudiosos russos das metrias da informação, entre as duas Grandes Guerras (1914-1918 e 1939-1941) e no subseqüente pós-guerra, destacando a originalidade de seu percurso. Ele observa:

> "Lenin, durante seu exílio na Europa ocidental, estudou atentamente as teorias socio-econômicas de Marx, nas bibliotecas de Bruxelas, Paris, Londres e Genebra e, com certeza, teve conhecimento das idéias de Otlet sobre Documentação. Desde a criação da Federação Internacional de Documentação (FID) em 1937, a URSS foi seu apoiador mais ativo. Assim, a partir de 1917, a URSS montou uma linha de desenvolvimento um pouco diferente da seguida pelo Ocidente na estruturação de suas bibliotecas e de seus serviços de informação. Isso serviu à URSS da melhor forma possível, na Segunda Guerra Mundial, quando surpreendidos pelos alemães, [...] os Russos conseguiram detê-los antes de Leningrado e Moscou, durante o inverno, e mandá-los de volta para Berlim. Foi uma vitória conseguida, em primeiro lugar, pela tenacidade de povo russo e, em último termo, pela tecnologia, desenvolvida às pressas, longe das bombas alemãs, por trás da 'muralha' dos montes Urais. A URSS apreendeu que uma base científica e tecnológica é crucial para fazer a guerra – ou para evitá-la."

Se referindo ao período do pós-guerra, Brookes acrescenta:

> "O Instituto da União [Soviética] para Informação Científica e Tecnológica foi criado em Moscou, em 1951. A missão a ele atribuída pela Academia nacional de Ciências da Rússia foi de organizar e coordenar os sistemas de informação científica da URSS. Com funções parcialmente operacionais, parcialmente educativas e parcialmente orientadas à pesquisa sobre os processos de 'cientifização'. Em russo, o termo **ciência** tem um sentido amplo, como termo alemão **Wissenschaft,** ou seja, engloba todas as formas de conhecimento organizado. [...] Os russos definiram **Informatics** como a 'disciplina científica que estuda a estrutura e propriedades da informação científica e as leis dos processos de comunicação científica' [(Mikhailov et al., 1984)]. A consideram como uma ciência social e, reconhecida pela Academia de Ciências, tem um status que a ciência da informação não alcançou ainda no Ocidente."

Brookes prossegue:

> "A primeira vez que tomei conhecimento do termo **scientometrics** foi por médio de uma nota do VINITI publicada pela FID, de autoria de G.M Dobrov e A.A. Karennoi [(Dobrov; Karennoi, 1969)]. Dobrov me visitou em Londres e definiu o novo termo como a 'medida dos processos informáticos'. Para exemplificar, me informou que tinha

---

Loet Leydesdorff ensina no *Department of Communication Studies*, da Universidade de Amsterdam. Seu maior feito é o desenvolvimento do "modelo da triple hélice" para representar as relações entre universidade-indústria e governo (LEYDESDORFF; MEYER, 2007). Mais informações em http://users.fmg.uva.nl/lleydesdorff/index.htm. Acesso em agosto 2009.

Peter Ingwersen é Professor de Pesquisa no Departamento de Estudos de Informação, na *Royal School of Library and Information Science*. Trabalha atualmente em: Fusão de dados com base nos princípios de poli-representação em IR; Poli-representação do espaço informacional; e Poli-representação do espaço cognitivo dos pesquisador. Autor de dois livros (INGWERSEN, 1992) e (INGWERSEN; JARVELIN, 2005) e mais de 70 artigos publicados desde 1982, com destaque para (ALMIND; INGWERSEN, 1997) e (BJÖRNERBORN; INGWERSEN, 2004). Além da medalha Derek de Solla Price (2005), recebeu prêmios da ASSIST, como pesquisador destacado em perspectivas cognitivas da informação (2003) e professor destacado em ciência da informação (2007), prêmio Thomsom, como o autor dinamarquês mais citado em ciência da informação (2005), e prêmio Jesse Farradane por suas contribuições para a ciência da informação (http://www.db.dk/ombiblioteksskolen/medarbejdere/default.asp?cid=684&tid=4#publikationer).

[9] *Vickery, Brian Campbell.* Nascido em 1918, na Austrália, estudou em Austrália, Cairo (Egito) e Canterbury (Inglaterra) obtendo um MA em Química, na Universidade de Oxford, em 1941. Trabalhou como químico na *Royal Ordnance Factory*, em *Bridgwater, Somerset*, Ingalaterra, de 1941 a 1945 e como editor assistente da revista *Industrial Chemist,* em Londres, iniciando suas atividades como bibliotecário em 1946, quando trabalhou, sucessivamente, nos *Akers Research Laboratories*, das *Imperial Chemical Industries*, em Welwyn, Hertfordshire, Inglaterra, na *UK National Lending Library for Science and Technology*, como *Principal Scientific Officer*, em Boston Spa, Yorkshire, Inglaterra, e como bibliotecário no *Institute of Science and Technology*, da Universidade de Manchester. *Research Director*, da ASLIB, em Londres, de 1966 a 1973 e, de 1973-1983, e Professor e diretor da *School of Library, Archive and Information Studies,* do *University College,* em Londres. Em 19883 foi nomeado professor emérito da Universidade de Londres, onde presta consultoria, em tempo parcial. (VICKERY, 1960; 1973; 2004 *apud* HORLAND, 2006).

Wersig, Gernot (1942-2006). É um dos poucos cientistas da informação que trabalhou sobre os fundamentos sociológicos da ciência da informação. Seu livro, publicado em 1973, é pioneiro sobre o assunto. A visão sociológica da teoria da informação se confirma em um trabalho de 2003, publicado na *International encyclopedia of library and information science* (WERSIG, 1973; 1993; 2003).

> medido a performance de cerca de 20 Institutos de Matemática em Ucrânia, [...] acrescentando que foram contados os totais de publicações, patentes, etc. de cada um deles e os resultados relacionados com o tamanho e Antigüidade de cada instituto, quadro de pessoal, número de doutores, etc. [...]. Os gráficos obtidos dividindo o numero de publicações e de patentes – desde a data de criação da instituição – pelo número de doutores indicam que a produtividade cresce regularmente até alcançar um total de, aproximadamente 15 cientistas mas, depois, a produtividade diminui à medida que a instituição aumenta de tamanho. [...] O termo **informetrics** foi sugerido por Otto Nacke, da Alemanha Ocidental em 1979. Em 1984, o VINITI criou um Comitê, na FID, com esse nome, sendo o próprio Nacke seu primeiro presidente. Seu sucessor, [T.N.] Rajan, do INSDOC da India, definiu os objetivos de 'informetrics' como sendo a provisão de dados confiáveis para a a pesquisa e o desenvolvimento, para o planejamento gestão e políticas e para a gestão de instituições, projetos, programas e atividades. O termo é também considerado como relacionado com as origens do desenvolvimento dos conceitos. Assim definido, o termo informetrics é mais amplo e profundo dos três termos métricos que nos interessam."

Dentre os pesquisadores russos deve-se citar Vasily V. Nalimov, premiado com a medalha Derek de Solla Price, em 1987, pesquisador visionário em todos os domínios da matemática. Outros nomes que merecem destaque são: Z.M. Mulchenko, Gennadii Mikhalovic Dobrov, Valentina Gorkova e, naturalmente, o Professor Alexander Ivanovich Mikhailov.[10] Especial destaque merece, também, a contribuição dos pesquisadores húngaros ao desenvolvimento da cientometria e da infometria. Entre outros lembraremos: Wolfgang Glänzel, Presidente da 4ª Conferência Internacional sobre Bibliometria, Informetria e Cientometria, em Alemanha em 1999. Premiado com a medalha Derek de Solla Price, concedida pela *International Society for Scientometrics and Informetrics* (ISSI), pela sua relevante contribuição ao estudos quantitativos sobre o desempenho e evolução da ciência. Atualmente trabalha, desde 2002, no Instituto para Organização da Pesquisa, também da Academia de Ciência. É autor de mais de cinqüenta publicações em língua inglesa sobre matemática e estatística, e sobre bibliometria, cientometria e infometria, totalizando mais de cem se somadas às publicadas em outras línguas.[11] Outros cientistas húngaros que mereceram a medalha Derek de Solla Price por suas destacadas contribuições sobre metrias da informação em 1986, 1993 e 2009 foram, respectivamente: Tibor Braun, András Schubert e Péter Vinkler. A esses nomes cabe acrescentar Jan Vlachý (antiga Checoslováquia), em 1989.[12]

---

[10] *Vasily V. Nalimov* (1910-1997) publicou, em 1969, juntamente com Z.M. Mulchenko (NALIMOV; MULCHENKO, 1969) o livro pioneiro intitulado "*Naukometriya*" (*Наукометрияя* em caracteres cirílicos; 'Nauka' = Ciência, e 'metriya' = medida, mensuração, metria, em russo) com o qual, segundo Hood e Wilson (2004) foi cunhado o termo *scientometrics,* que é, também, o título da publicação periódica *Scientometrics*, mencionada acima, publicação nuclear em nosso campo de estudo. Ver, por exemplo: http://www.issi-society.info/popup_nalimov.html. Consultado em ago 2009.
Z.M. Mulchenko. Co-autor, com Vasily V. Nalimov, da obra "*Naukometriyia*" (Ver acima).
Gennadii Mikhailovic Dobrov. Colaborador de Mikhailov, publicou na Rússia, em 1966, o livro intitulado "*Nauka o nauke"* (*Ciência da ciência*), considerado, juntamente com o livro "*Naukometriyia*" ("Cientometria") de autoria de Nalimov e Mulchenko (1969), como uma das obras precursoras do desenvolvimento dos estudos cientométricos. Aparentemente, a única tradução do original russo da obra de Dobrov, foi publicada em Berlim, em alemão, em 1969/1970 (DOBROV, 1969). A título de curiosidade, assinalamos que um – ou dois – exemplar(es) da versão alemã encontra-se na Biblioteca da *London School of Economic and Political Science* (LSE), na Coleção Lakatos, cuja referência no catálogo pode-se, por exemplo: http://library-2.lse.ac.uk/collections/lakatos/lakatos_d.htm. Consultado em agosto 2009.
Valentina Gorkova. Em 1988, o VINITI publicou um livro editado pela matemática Valentina Gorkova intitulado 'Informetrics' (GORKOVA, 1988 *apud* Brookes, 1989). É uma compilação atualizada dos resultados das pesquisas em *'informatics'.*
Alexander Ivanovich Mikhailov (1905-1988). Graduado em engenharia mecânica em 1931, no Instituto Mendeleev, foi um grande promotor dos estudos sobre informação e de sua relação com as políticas de ciência e tecnolgia, e um dos fundadores do Instituto da União [Soviética] para Informação Científica e Tecnológica (VINITI – *Vserosiisky Institut Nauchnoi I Tekhnicheskoy Informatsii*), em 1951, se tornando seu diretor em 1956 até o fim de seus dias. Simultaneamente, foi um entusiasta incentivador das atividades da Federação Internacional de Documentação (FID) – posteriormente Federação Internacional de Informação e Documentação – como vice-diretor duas vezes (1969-1976 e 1981-1988), como membro do Conselho e como coordenador do Comitê para a Pesquisa sobre os Fundamentos Teóricos da Informação (FID/RI – *Study Committee Research on Theoretical Basis of Information*), or FID/RI. Na década de 1960 desenvolveu o conceito de '*informatics*' ('*Informatika'* , em russo), relacionado com o estudo, organização, e disseminação da informação científica. Importantes estudos nessa área foram publicados entre 1971 e 1979. Ver, por exemplo: Mikhailov e Gilyarevskii (1971), e Mikhailov (1972, 1979).
Com mais de duzentos trabalhos na área, Mikhailov é um dos primeiros estudiosos da informação como objeto da pesquisa científica e, sem dúvida o mais influente estudioso da ciência da informação na União Soviética e na Europa Oriental. Ver, por exemplo: http://www.viswiki.com/en/Alexander_Ivanovich_Mikhailov. Consultado em ago 2009.

[11] Uma relação de seus trabalhos, em língua inglesa, encontra-se disponível em: http://www.steunpuntoos.be/wg.html. Consultado em agosto 2009.

[12] *Tibor Braun* é professor de química na Universidade L. Eötvös em Budapeste. Recebeu o prêmio internacional George Hevesy em 1975 e o prêmio de química da Academia de Ciências da Hungria em 1980. Continua trabalhando na área de química nuclear, com várias dezenas de artigos e vários livros. Fundador e editor-chefe do *Journal of Radioanalytical and Nuclear Chemistry.* Fundador, também, da Unidade de Pesquisa Ciência da Informação e Cientometria da Biblioteca da Academia de Ciências da Hungria, e fundador e editor-chefe da revista *Scientometrics* (1987). Disponível em: http://www.springerlink.com/content/y9450377p0hqr835/fulltext.pdf?page=1. Consultado em agosto 2009.
András Schubert é doutor em química pela Universidade Técnica de Budapeste. Mudou da físico-química para a biblioteconomia e a ciência da informação em 1979, quando se incorporou ao Grupo de Pesquisa em Ciência da Informação e Cientometria da Academia de Ciências de Hungria, onde chefia o Serviço Bibliométrico. Líder de dezenas de projetos de pesquisa, com mais de cem trabalhos publicados e numerosas palestras e cursos. Integra o corpo editorial da revista *Scientometrics*. Disponível em:

2.3 A corrente indiana

Na Índia, o nome maior, na história da bibliometria da Índia, é, sem dúvida, Shiyali Ramamrita Ranganathan (1892-1972). Matemático e estudioso das práticas bibliotecárias, idealizador do '*Colon Classification System*' é mundialmente conhecido como um teórico da classificação, pela sua classificação facetada, em cinco categorias (PMEST: *Personality* [personalidade], *Matter* [matéria], *Energy* [energia], *Space* [espaço] e *Time* [tempo]). Em 1947, segundo P.S. Kavatra (2000), Ranganathan criou, em sua obra "*Library Administration*", o termo *'librametry'* ('livrametria' ou 'livrometria'), no sentido de "*aplicação de métodos quantitativos na gestão de bibliotecas e de seus serviços.*" No ano seguinte, na *ASLIB Conference*, sugeriu que os bibliotecários começassem a aplicar a *'librametry'* de um modo semelhante ao aplicado em biometria, econometria e psicometria, pois "*muitos dos assuntos relativos às tarefas das bibliotecas implicam o uso de valores numéricos muito grandes.*"

Outros nomes que merecem ser citados, dentre outros, são: T.N. Rajan, I.K. Ravichandra Rao e A. Neelameghan.[13]

2.4 A corrente americana

Começaremos por destacar os nomes dos cientistas americanos premiados com a medalha Derek de Solla Price: Eugene Garfield (1984); Michael J. Moravcsik (1985); Henry Small (1987); Francis Narin (1988); Robert K. Merton (1995); Belver C. Griffith (1997); Howard D. White (2005); Katherine W. McCain (2007).[14]

---

http://www.kennismakers.be/ASchubert.html. Consultado em agosto 2009.

Péter Vinkler é autor de numerosas publicações, principalmente em periódicos húngaros, sobre indicadores infométricos e cientométricos, em diversas aplicações, tais como a medida do fator de impacto das publicações periódicas (VINKLER, 2007), metrias aplicadas ao estudo da ciência húngara (VINKLER, 2008), etc. Compartilhou o Prêmio De Solla Price com o francês Michel Zitt (ver Nota 5, acima).

Jan Vlachý é conhecido, desde 1960, na comunidade cientométrica por seus estudos quantitativos, no campo da física, sobre autores, produtividade, temas mais estudados, etc. Publica regularmente no *Czechoslovak Journal of Physics B,* e colabora regularmente com a revista *Scientometrics*, onde publica seus trabalhos em inglês (TODOROV, 1990).

[13] *T.N. Rajan.* Foi o segundo presidente do Comitê *'Informetrics'* da Federação Internacional de Documentação e Informação (FID), em 1986 (RAJAN,1986 *apud* VALDERRAMA ZURIÁN *et al*., 1996). Ver, também, Brookes (1989). Ganhou destaque por seus trabalhos no INSDOC – *Indian National Scientific Documentation Centre* (SEM *et al*., 2002) sobre o desenvolvimento dos currículos dos profissionais da informação em ambientes em mutação (RAJAN, 1983).

I.K. Ravichandra Rao. Graduado em matemática e estatística no *Mahatma Gandhi Memorial College*, em Udupi, mestre em estatística pelo Instituto Indiano de Estatística de Calcutá, em 1970, e PhD pela Universidade de Western Ontário, no Canadá, em 1981. Tem lecionado no Centro de Treinamento e Pesquisa em Documentação (*Documentation Research and Training Centre* – DRTC), em Bangalore, pertencente ao Instituto Indiano de Estatística. Com profunda experiência em estatística aplicada à bibliometria, é bem conhecido no Brasil pelo seu livro *Métodos Quantitativos em Biblioteconomia e Ciência da Informação,* tradução do original inglês de 1983, publicado pela Associação dos Bibliotecários do Distrito Federal (ABDF), em 1986, sob os auspícios do então Departamento de Biblioteconomia da Universidade de Brasília (UnB/BIB), com o apoio da Organização dos Estados Americanos (OEA).

Arashanipalai Neelameghan. É atualmente professor visitante honorário do DRTC (*Documentation Research and Training Centre*), em Bangalore, e Diretor executivo do Centro Ranganathan de Estudos da Informação, em Madras. Graduado em física e diplomado em *'LIbrary Sciences'*, pela Universidade de Madras possui cursos de especialização, na área, na *Vanderbilt University* e *Columbia University*. Após chefiar serviços de pesquisa em bibliotecas acadêmicas e corporativas, foi chefe e professor do DRCT. Professor visitante em diversas universidades (Western Ontário – Canadá; Pittsburgh, Rhode Island e Syracuse - Estados Unidos; Universidade Federal de Minas Gerais e Universidade de Brasília – Brasil; Simón Bolívar – Venezuela; CONACYT – México; Manila – Filipinas; etc.). Trabalhou vários anos com UNESCO como coordenador do Projeto UNESCO-UNDP "Cursos de Pós-Graduação para Profissionais em Ciência da Informação no Sudeste de Ásia" (1978-1982), como diretor do programa "*Institution Building and Networking*", e como assessor do PGI (Programa Geral de Informação) para a Região Ásia-Pacífico (1982-86). Presidente do Comitê Assessor do UNISIST (1974-78) e presidente do Comitê "*Informetrics*" (FID/CR), da Federação Internacional de Informação e Documentação – FID (1973–80). Publicou mais de 200 trabalhos científicos, relatórios técnicos e oito livros, e é editor da revista *Information Studies*, publicada em Bangalore. Recebeu o Prêmio ASIS (*American Society for Information Science*), em 1983, pelos esforços pra promover a cooperação internacional e, em 1992, o Prêmio Ranganathan, da FID. Ver, por exemplo: Neelameghan (1993, 1997).

[14] *Eugene Garfield* é considerado como um dos "pais" da bibliometria e da cientometria, e um pioneiro da informação científica. Fundador e presidente emérito do *Institute for Scientific Information* (ISI), atualmente *Thomson Scientific*. A partir de 1960, desenvolve novas ferramentas para facilitar o acesso às informações científicas, perdominantemente orientadas para as publicações anglosaxônicas: *Current Contents, Social Sciences*, *Citation Index* (SSCI), e *Arts and Humanities Citation Índex* (A&HCI). Em 2006 foi distinguido com o *Online Information Lifetime Achievement Award* em reconhecimento aos seus mais de cinqüenta anos de dedicação, liderança e inovação na indústria da informação (THOMSON REUTERS, 2006). Mais de 350 publicações – muitas acessíveis em texto completo – entre 1952 e 2008 (Ver: http://www.garfield.library.upenn.edu/pub.html. Acesso em agosto 2009).

Michael J. Moravcsik (1928-1989) foi qualificado de "*cientista multidimensional e herói da ciência do Terceiro Mundo: físico, cientometrista, crítico musical e embaixador da ciência.*" (GARFIELD; SMALL, 1991). Até 1989, foi professor de física teórica, na Universidade de Oregon.

Henry Small é diretor do Grupo de Serviços de Pesquisa da Thomson ISI. Especialista em análise de co-citações. Recebeu vários prêmios: 'Melhor trabalho publicado no *JASIS*' (1987) e 'Prêmio ao mérito' da *American Society for Information Science & Technology* (ASIST). É o sexto Presidente da ISSI, desde 2003 (Ver: http://www.issi-society.info/price.html. Acesso em agosto 2009).

Francis Narin especialista em desenvolvimento de indicadores para avaliação em ciência e tecnologia, avaliação de patentes e análise das relações entre ciência e tecnologia (NARIN, 1993).

Relacionamos a seguir os nomes de outros cientistas americanos que se destacaram pelas suas contribuições, em aspectos que direta ou indiretamente se relacionam com a gestão e organização dos acervos das bibliotecas, e que, com suas idéias sobre classificação, facilitaram o relacionamento de conteúdos e a descoberta de autores que mais se aprofundaram no estudo de determinados temas. Nomes de estudiosos tais como Melvil Dewey, Henry Evelin Bliss, e Charles A. Cutter[15], que se aproximam de outros nomes de outros países, já citados, como Paul Otlet, Shiyali Ramamrita Ranganathan, ou que serão lembrados no momento oportuno, como Ingetraut Dahlberg e – por que não? – Aristóteles.

Existem ainda nomes famosos dos primórdios da bibliometria, como os de George Kinsley Zipf (1902-1950), lingüista conhecido pela "lei" que leva seu nome e estudou as propriedades estatísticas da linguagem, e Alfred James Lotka (1980-1949), matemático, estatístico e físico-químico, autor da "lei do quadrado inverso", que serão lembrados, com especial detalhamento na Seção 4. "*Metrias da Informação antes da Web*", junto ao nome do matemático e bibliotecário inglês Samuel C. Bradford (1878-1948), que formulou a "Lei da dispersão".

E, finalmente, citaremos os nomes de Gerald Salton, Karen Spärck Jones, Frederick Wilfrid Lancaster e Tefko Saracevic, pelos seus trabalhos sobre avaliação da indexação e da qualidade da recuperação da informação, o que não quer dizer que não existam outros nomes merecedores de serem lembrados, mas, infelizmente, o espaço disponível para este sobrevôo, não é ilimitado. No decorrer da exposição que segue, os nomes de outros autores surgirão e, assim, o universo de nomes e referências irá se enriquecendo.[16]

Robert K. Merton (1910-2003). Um dos mais importantes sociólogos do século XX. Professor de sociologia na Universidade de Columbia (1941-1979)e, posteriormente, professor emérito. Membro honorário da Academia de Ciências de New York e da Academia Européia de Sociologia. Seu nome está estreitamente associado ao "Matthew-Effect" (efeito Mateus), sobre o sistema de comunicação e reconhecimento na ciência (MERTON, 1968).

Belver C. Griffith (1931-1999). Ensinou no *Drexel's College of Information Science and Technology*, de 1969 a 1991, quando se aposentou. Pioneiro da ciência da informação e cientometrista mundialmente reconhecido. Estudou em profundidade os padrões de comunicação entre cientistas. Premiado com o *Drexel University Research Achievement Award,* em 1980 e com o *Outstanding Information Teacher Award,* da ASIS, em 1982 (JOY MOYER, 2002).

Howard D. White. Após obter um PhD em biblioteconomia pela Universidade de California, Berkeley, em 1974, ingressou no *College of Information Science and Technology*, da *Drexel University*, onde é professor emérito. Coautor, com Marcia Bates e Patrik Wilson do, livro (1962) "*For Information Specialists: Interpretations of Reference and Bibliographic Work"* e autor de "*Brief Tests of Collection Strength"* (1995). Numerosas publicações dobre bibliometria, análise de co-citações, evaluação de serviços de referência. Em 1993, obteve o *Research Award* da *American Society for Information Science and Technology* (ASIST) e, em 1998, conjuntamente com Katherine McCain, ganhou o prêmio do melhor trabalho do *JASIS* (WHITE; McCAIN, 1998). Em 2004, foi agraciado com a maior honraria da ASIST, o Prêmio ao Mérito. Lista completa das publicações (algumas com *links*) encontram-se em: http://www.cis.drexel.edu/faculty/HUD.Web/HDWpubs.html. Acesso em agosto 2009.

Katherine W. McCain. Professora no *College of Information Science & Technology*, da *Drexel University*. Principal interesse nos últimos 25 anos foi o estudo com uma abordagem quantitativa e gráfica, do comportamento dos cientistas, como se organizam e se comunicam, como emergem e desaparecem os grupos (McCAIN, 1998; 2000), (McCAIN; SALVUCCI, 2006). Uma ista detalhada das publicações, encontra-se disponível em: http://www.ischool.drexel.edu/faculty/kmccain/. Acesso em agosto 2009.

[15] *Melvil Dewey* (1851-1931), nasceu numa família, numa pequena cidade ao norte do estado de Nova York. Foi um bibliotecário pioneiro, criador, aos 21 anos, quando trabalhava como auxiliar na biblioteca do *Amherst College* criou o sistema de classificação decimal que veio a se chamar chamada *Dewey Decimal Classification* (DCC). Seu trabalho pioneiro provocou uma revolução no trabalho bibliotecário, inaugurando uma nova era na biblioteconomia. Dewey bem merece o nome de "pai da biblioteconomia moderna". Participou da fundação da *American Library Association* (ALA), em 1876, da qual foi secretário de 1876 a 1890, e presidente de 1892 a 1893. Foi co-fundador e primeiro editor do *Library Journal* (1876), que muito contribuiria, como veículo da *American Library Association* para o desenvolvimento e normalização das práticas profissionais bibliotecárias. Ingressou como bibliotecário do *Columbia College* (hoje *Columbia University*) em Nova York, em 1883. Foi diretor da onde fundou a primeira escola de biblioteconomia em 1887. De 1889 a 1906, foi diretor da *New York State Library*, em Albany (OCLC, 2008). A DDC, cujos direitos foram adquiridos pela OCLC, é usada majoritariamente pelos países anglo-saxônicos. Ver também, Hjørland (2006).

Henry Evelyn Bliss (1870-1955). Responsável pela biblioteca do atual *City College* da Universidade de Nova York, na primeira década do século XX (BLISS, 1934, *apud* HJØRLAND, 2006). A primeira versão da Bibliographic Classification foi publicada em quatro volumes, entre 1940 e 1953 (BLISS, 1950-1953, *apud* HJØRLAND, 2006). A Classificação de Bliss foi adotada, rapidamente, nos países anglo-saxônico, mas restou praticamente dês conhecida no Brasil. Com a criação da *Bliss Classification Association* (BCA), nasceu um ambicioso projeto de revisão da primeira versão, denominada sucintamente BC2, com 16 volumes previstos, vocabulário ampliado e categorias rigorosamente ordenadas (ROBREDO, 2005, p. 138-139). Informações mais atualizadas sobre modalidades de aquisição, no portal da *Bliss Classification Association* (2007).

Charles A. Cutter. (1837-1903). Bibliotecário. As regras de Cutter (1876) são clássicas, embora negligenciadas com freqüência. Seu sistema de classificação, denominado *Expansive Classification* (Classificação Expansível – 1891–93) era alfabético, no lugar de numérico e era a base da classificação da Biblioteca do Congresso (CUTTER, 1976, *apud* HJØRLAND, 2005).

[16] *Gerald Salton* (1927-1995). Pioneiro na aplicação de métodos estatísticos na recuperação da informação, produção de resumos e indexação automáticas, e outros tratamentos computadorizados da linguagem natural. (DUBIN, 2004) e (HØRLAND, 2006)

Karen Spärck Jones (1935-2007). Professor emérito da Universidade de Cambridge. Maior interesse nos problemas lingüísticos na ciência da informação. Sua contribuição mais importante é, provavelmente, a "medida da especificidade estatística" (*statistical specificity* ou *term frequency–inverse document frequency* [tf–idf]) (HØRLAND, 2006).

2.5 A bibliometria no Brasil

O primeiro livro sobre bibliometria, organizado por Edson Nery da Fonseca e intitulado *Bibliometria:Teoria e prática; textos de Paul Otlet, Robert Estivais, Victor Zoltowski, Eugene Garfield*, foi publicado em 1986 (FONSECA, 1986). Destacamos algumas passagens da recensão de Gilda Maria Braga (1987):

> A aparição do primeiro livro de bibliometria no Brasil, em língua portuguesa, organizado por Edson Nery da Fonseca é um importante fato para os profissionais da informação. Os textos de Otlet (O livro e a medida. Bibliometria) e Estivals (Criação, consumo e produção intelectuais) são incluídos na "teoria"; os de Zoltowski (Os ciclos da criação intelectual e artística) e Garfield (Historiógrafos, Biblioteconomia e a História da Ciência) na "prática". Esta última está subdividida em Macrobibliometria - texto de Zoltowski - e Microbibliometria - texto de Garfield. A conceituação de bibliometria, para Otlet é "a parte definida da Bibliografia que se ocupa da medida ou da quantidade aplicada ao livro (Aritmética ou Matemática Bibliometria)". O contexto do termo é definido, a seguir: "A medida do livro consiste em reduzir todas as partes e elementos de um livro qualquer às partes e elementos de um livro tipo, padrão unidade. Esse tipo deveria constituir-se no melhor dos livros". Bibliometria, para Otlet, é a mensuração do objeto - livro, essencialmente: número de palavras por página, linhas por página; formatos; pontos tipográficos, peso do papel, preço. Estivais apresenta uma teoria bibliológica reunindo seus esforços de vinte anos; em sua opinião, a "Bibliometria bibliográfica pode concernir à economia do livro (papel, páginas, volumes, tiragens); à psicologia e à sociologia da inovação, pelas bibliografias de autores ou de movimentos célebres." Zoltowski ofereceu seu trabalho como uma "contribuição a uma teoria da atividade criadora." O caminho percorrido, do ponto de partida ao ponto de chegada, foi auxiliado por um índice que se baseia no número de unidades impressas ou gravadas, isto é, os livros, as partituras musicais e as estampas, de acordo com as bibliografias e sem considerar o seu valor intelectual ou artístico. "A conceituação de Garfield, do termo bibliometria é similar à de A. Pritchard. "Em anos ulteriores, Pritchard usou a palavra "bibliometria" a fim de descrever a análise quantitativa de citações (sic). Os historiadores russos de Ciência sugeriram o uso do termo "cientometria" para semelhante tipo de estudo.

É curioso verificar – conforme salienta o organizador da coletânea – a ignorância que os trabalhos de língua francesa têm dos de língua inglesa – e vice-versa. Datando de 1934, 1970 e 1955, os três primeiros trabalhos, em original francês, não mencionam os trabalhos de Bradford, Vickery, Brookes, Sandison, Line, Fairthorne, etc. – desenvolvidos na Inglaterra, e mais próximos da conceituação da bibliometria adotada por Pritchard – análise quantitativa da informação registrada – e aceita nos países de idiomas outros que não o francês.

> É pena que a coletânea não inclua os trabalhos de Lotka, Zipf, Bradford, Price, Goffman, e tantos outros que tentaram formular as leis da bibliometria, mas, o propósito do organizador foi o de disseminar, justamente, as idéias desses autores franceses. Os interessados na bibliometria sem dúvida apreciarão esse livro.

A origem dos estudos bibliométricos, no Brasil, é uma decorrência da fundação, em 1954, do Instituto Brasileiro de Bibliografia e Documentação (IBBD) como órgão do então Conselho Nacional de Pesquisas (CNPq), que, a partir de 1976, passou a se denominar Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), permanecendo vinculado ao CNPq, agora Conselho de Desenvolvimento Científico e Tecnológico, mantendo a sigla. O IBICT desempenhou um papel central na modernização das atividades bibliotecárias e informacionais, e promovendo o surgimento da ciência da informação no país. Os registros dos fatos e eventos mais marcantes dessa história encontram-se dispersos em um grande número de publicações, comunicações, etc., com profundidade e fidedignidade, infelizmente, variáveis, o que torna extremamente complicada a elaboração de uma síntese coerente do percurso – ou da ousada aventura – que elevou o Brasil do subdesenvolvimento a uma posição privilegiada entre as potências emergentes. E isso, num período de pouco mais de cinqüenta anos. Assim, nossa fonte principal, para elaborar essa breve síntese, será um conjunto de publicações da professora Lena Vania Ribeiro Pinheiro, testemunha e personagem dessa aventura, onde se

---

Frederick Wilfrid Lancaster. Cientista da informação trabalhou na *National Library of Medicine*, Bethesda, Md. (1965-68); professor na Universidade de Urbana, Ill. (1972-92), onde é professor emérito. Conhecido por seus trabalhos sobre avaliação de sistemas de recuperação da informação em linha (HØRLAND, 2006).

Tefko Saracevic. Estudou engenharia elétrica na Universidade de Zagreb, Croácia e completou o mestrado e doutorado em ciência da informação, respectivamente em 1962 e 1970, na *Case Western Reserve University*, Cleveland, Ohio, onde ensinou até 1985, quando se transferiu para a *Rutgers Univerity*. Promovido ao mais alto nível acadêmico em 1991, foi designado *Associate Dean* de 2003 a 2006. Com numerosas publicações sobre avaliação de sistemas de recuperação da informação, medidas de relevância da informação, interação homem-computador, valor da informação, análise da adequação das perguntas submetidas aos mecanismos de busca na Web, etc., trabalho também como consultor, pesquisador e/ou professor para numerosos organismos nacionais e internacionais (*National Science Foundation, Council for Library Resources, Rockefeller Foundation*, UNESCO, etc.). Presidente da *American Society for Information Science* (ASIS), em 1991; Prêmio Gerard Salton, da *Association for Computing Machinery* (SIGIR/ACM), em 1997; Prêmio ao mérito da ASIS, em 1995; Prêmio ao melhor trabalho do *Journal of the American Society for Information Science*, em 1999. Mais de 1193 citações, no *Social Sciences Citation Index* e no *Science Citation Index* – excluídas auto citações – até fevereiro de 2009. (SARACEVIC, 2009). Para algumas publicações selecionadas, ver Saracevic (1999, 2008), Spink *et al.* (2001), Jansen *et al.* (1998). Maiores detalhes em: http://comminfo.rutgers.edu/~tefko/articles.htm

reúnem dados e informações, fatos e eventos, alicerçados numa copiosa e confiável bibliografia[17], acrescida, quando necessário de outras referências. Dividiremos a narrativa dessa história, apresentada a seguir, em duas partes: "Um pouco de história para contextualizar" e "Um pouco de história sobre a bibliometria no Brasil".

2.5.1 Um pouco de história para contextualizar

O Catálogo Nacional de Publicações Seriadas, as bibliografias especializadas brasileiras, as buscas bibliográficas, o Cadastro de Pesquisas em Andamento, o Serviço de Intercâmbio de Catalogação a reprodução de documentos. Parte das ações do IBBD estavam voltadas à capacitação de recursos humanos, através de As principais atividades do IBBD foram, originalmente, o Catálogo Coletivo um Curso de Especialização em Pesquisa Bibliográfica, oferecido desde o ano seguinte à fundação do Instituto, posteriormente denominado Curso de Especialização em Documentação e Informação (CDC), descontinuado em 2002, e do Mestrado de Ciência da Informação, em funcionamento desde 1970. A informação científica e tecnológica aparece nas políticas públicas brasileiras na década de 1970, com o primeiro Plano Nacional de Desenvolvimento (PND), elaborado pelo governo militar, no poder desde 1964, para o triênio de 1972 a 1974, que previa, entre outras coisas, a implantação de um Sistema Nacional de Informação em Ciência e Tecnologia (SNICT), com o objetivo principal, expresso no Plano Básico de Desenvolvimento Científico e Tecnológico I, para esse período, de "captar, tratar e difundir, de forma sistemática e permanente, informações atualizadas na área de Ciência e Tecnologia, assim como os vários subsistemas que dele participariam, com suas respectivas áreas de atuação". No Plano Básico de Desenvolvimento Científico e Tecnológico II, 1975/79, a informação científica e tecnológica ganhou mais espaço, incluída entre as atividades de apoio ao desenvolvimento científico e tecnológico, é considerada como "elemento básico de apoio para a formulação de políticas e estratégia do governo". Curiosamente, o documento intitulado *"Diretrizes básicas para a implantação do Sistema Nacional de Informação Científica e Tecnológica (SNICT)"*, elaborado sob os auspícios do Conselho Nacional de Pesquisas (CNPq), "por razões ainda não esclarecidas, se volatilizou, no início da década de 80 [do século passado], na transmissão do poder, do penúltimo para o último presidente da ditadura." (ROBREDO, 2005).

O IBICT surge na perspectiva de órgão de fomento e coordenação, com a missão de promover a efetivação do Sistema Nacional de Informação em Ciência e Tecnologia, integrado por sistemas setoriais de informação, (Aguiar, 1980, *apud* Pinheiro e Loureiro, 2004), assumindo um perfil de entidade coordenadora e política. O terceiro Plano Brasileiro de Desenvolvimento Científico e Tecnológico, de 1980 a 1985, reconhece a importância do IBICT, atribuindo-lhe o papel de "órgão central com condições institucionais e materiais para desempenhar funções de coordenação descentralizada das atividades de informação em Ciência e Tecnologia no País" (Briquet de Lemos, 1986, *apud* Pinheiro e Loureiro, 2004). Este plano, diferentemente dos anteriores, não incluía programa, projetos e atividades, transferindo essa incumbência aos documentos específicos denominados "ações programadas", com destaque para três importantes marcos na política brasileira de informação:

- a Ação Programada de Informação Científica e Tecnológica;
- o Programa de Apoio ao Desenvolvimento Científico e Tecnológico (PADCT), com recursos do Banco Mundial; e
- o Plano Nacional de Bibliotecas Universitárias - PNBU.

A Ação Programada Brasil-Seplan-CNPq, elaborada em 1984, reflete o momento histórico brasileiro, de abertura democrática, dela tendo participado mais de cem profissionais de informação das mais diversas entidades brasileiras. A Ação Programada partiu das seguintes propostas: geração de documentos primários; formação e desenvolvimento de coleções; automação de bibliotecas; bases de dados bibliográficos; difusão e uso da informação; recursos humanos; e assuntos internacionais. Trata-se de um diagnóstico que aponta diretrizes, identifica instituições e suas respectivas atribuições para superar os problemas de informação mapeados. O IBICT manteve algumas experiências mais antigas como o Catálogo Coletivo Nacional de Publicações Seriadas (CCN), uma das primeiras redes automatizadas do país, e o Programa de Comutação Bibliográfica (COMUT) e a revista *Ciência da Informação*.[18] Dentre os programas brasileiros paralelos aos do IBICT, merece destaque o PNBU, surgido em 1986 e que marcou, decisivamente, as atividades de bibliotecas universitárias brasileiras, mantendo intensa atividade até meados de 1991, quando

[17] Sugerimos aos leitores que desejem aprofundar o assunto a consulta de Pinheiro (1997, 2000, 2005), Pinheiro e Loureiro (1995, 2004), Pinheiro; Bräscher; Burnier (2005), e Pinheiro e Silva (2008), e usufruir das citações reunidas. Ver, também, Oddone (2006).

[18] Outras inovações merecem ser lembradas, entre elas: o sistema de gerenciamento de bases de dados bibliográficos GERIR (BARCELOS et al, 198?); o sistema de automação de bibliotecas GIBI (SOUZA, 1991 e CARDOSO, 1988); a linguagem padronizada LINCE, para acesso a bases de dados (IBICT, 1988 e IBICT; CNEN-CIN, 1990); a distribuição e suporte ao sistema MicroISIS, da UNESCO (MIKI, 1989); o documento "*Diretrizes para Elaboração de Tesauros Monolíngües*" (IBICT, 1984); o sistema TECER, para elaboração de tesauros (IBICT, 1989); e o formato IBICT (IBICT, 1984).

começou a perder força gradativamente. O mestrado em Ciência da Informação nasceu da natural evolução da experiência do curso de especialização, implantado no antigo IBBD, em 1955, um ano após a fundação do Instituto, sob a denominação de Curso de Pesquisa Bibliográfica. Esse curso foi um reflexo do surgimento da Documentação e da Bibliografia e do novo campo de trabalho aberto por essas disciplinas, para profissionais de diferentes formações que trabalhassem com informação especializada. A partir de 1964 passou a ser chamado Curso de Documentação Científica (CDC), por força do convênio com a Universidade Federal de Rio de Janeiro (UFRJ), e a partir de 1984 transformou-se em Curso de Especialização em Documentação e Informação, mantendo a sigla CDC e o vínculo acadêmico com a UFRJ, na Escola de Comunicação (ECO).

Em 1972 foi criado, no então IBBD, o mestrado em Ciência da Informação e lançada a revista *Ciência da Informação*. O mestrado teve, no início, presença maciça de professores estrangeiros, principalmente dos Estados Unidos. Dentre os nomes reconhecidos internacionalmente, que ministraram aulas e orientaram dissertações, até o ano de 1981, destacam-se: Frederick Wilfrid Lancaster (34 orientações), Tefko Saracevic (13 orientações), LaVahn Marie Overmyer, Bert Roy Boyce e Jack Mills (2 orientações cada um), além de John Joseph Eyre, Ingetraut Dahlberg e Suman Datta. Também deve-se lembrar a presença em seminários e conferências de renomados cientistas, entre os quais Derek John de Solla Price, da Yale University e Simão Mathias, da Universidade de São Paulo (Pinheiro e Loureiro, 1995).[19]

Parafraseando Pinheiro (PINHEIRO, 2005): *"Nunca é demais destacar o significado do IBBD para o Brasil, se pensarmos que o VINITI, na antiga União Soviética, foi fundado na mesma década, apenas dois anos antes, em 1952."* É da década de 50 o início da formação de recursos humanos, pelo IBBD, com a criação do Curso de Pesquisas Bibliográficas, depois denominado Curso de Documentação e Informação Cientifica (CDC), como mencionado antes. Voltado para profissionais de outras áreas que fossem atuar em documentação ou informação científica e tecnológica, segue as idéias que permearam a documentação e atividades de documentalistas, portanto, em harmonia com o pensamento de Otlet e de europeus, em geral. O CDC funcionou regularmente até 1995 e foi interrompido em 2000, quando foi oferecido o último curso. Ao longo de mais de 40 anos formou mais de 700 especialistas brasileiros e deixou uma lacuna.

> É uma fase de mudança de paradigmas: do armazenamento e preservação para a disseminação da informação através de produtos e serviços de informação; das bibliotecas gerais para os Centros de Informação / Documentação, que atenderiam áreas especializadas, pela fragmentação da Ciência; e da diversificação de acervo, anteriormente espaço hegemônico dos livros, passando a incorporar diversificada tipologia documental, inclusive documentos iconográficos (ATHERTON, 1977, apud Pinheiro, 2005). No Brasil, a década de 1970 marca a institucionalização da área [...], tanto pela criação do primeiro mestrado em Ciência da Informação, no IBICT, [...] com mandato acadêmico da UFRJ, quanto de outros cursos de pós-graduação que se seguiram [... a partir de 1990] (PINHEIRO, 2000). Estes mestrados foram [...]: os seguintes: o da USP, em 1972, como área dentro da Comunicação; da UFMG, em 1976; os da PUCCAMP e da UFPB, em 1977; e o de Brasília, na UNB, em 1978. [...] Os anos 70 correspondem ao período [...] da implantação de sistemas, redes e centros de informação (REBAM, CIN, BINAGRI, e Bibliodata/CALCO, por exemplo), além do acesso às bases de dados estrangeiras.[20], pelo IBICT. É, ainda, o momento da mudança de nome de IBBD para IBICT, em 1976, seguindo as transformações terminológicas internacionais, vistas anteriormente, e passando de órgão prestador de serviços para político e coordenador. A descontinuidade de serviços, sobretudo de bibliografias brasileiras, em diversas áreas, privou o Brasil do registro contínuo e atualizado da produção científica nacional, que permitiria a geração de indicadores de Ciência e Tecnologia, que dependem de bases de dados completas e correntes, e poderiam ser valiosos instrumentos para políticas de C&T, com recursos de métodos bibliométricos, hoje com softwares especializados para tal.

---

[19] O livro *"Cronologia do desenvolvimento científico e tecnológico brasileiro, 1950-2000"* é uma importante fonte para dados de ciência e tecnologia, no Brasil, a partir do qual podem ser sistematizados os conceitos específicos de Informação em ciência e tecnologia (MDIC, 2002, apud Pinheiro 2005).

[20] Tendo sido um dos autores deste Capítulo diretor do projeto internacional "Sistema Nacional de Informação e Documentação Agrícola (SNIDA)", de 1974 a 1980, implantado numa parceria entre o Ministério da Agricultura do Brasil e as Nações Unidas, por intermédio de seus organismos PNUD e FAO, parece oportuno lembrar alguns dos desdobramentos e realizações que acompanharam a criação da Biblioteca Nacional de Agricultura (BINAGRI), acima mencionada por Pinheiro (2008). Dentre eles: *a)* a incorporação da literatura agrícola brasileira ao Sistema Internacional de Informação Agrícola (AGRIS), patrocinado pela FAO, *b)* a livre disponibilização do acesso à literatura agrícola mundial a partir das fitas magnéticas geradas mensalmente pelo Sistema AGRIS e o acesso às cópias ou microfichas dos documentos originais (sistema BRAGRIS); *c)* a criação e operação do primeiro serviço, na América Latina, de disseminação seletiva da informação (DSI), totalmente automatizado, que enviava mensal e gratuitamente uma seleção de referências extraídas das fitas magnéticas AGRIS, conforme ao perfil de interesse dos pesquisadores, docentes, etc. cadastrados, que chegou a atingir picos de mais de mil usuários (sistema BipAGRI); *d)* um sistema automatizado de acompanhamento das pesquisas agrícolas brasileiras em andamento, cujas sinopses eram incorporadas a base de dados internacional *Current Agricultural Research Information System* (CARIS), acessível a partir da BINAGRI (sistema BRACARIS); *e)* a produção automatizada e divulgação de bibliografias monográficas de produtos agrícolas brasileiros (banana, mandioca, milho, cana de açúcar, etc.); *f)* estabelecimento, por métodos bibliométricos, de listas básicas de títulos de periódicos agrícolas, para promover o aprimoramento dos acervos das bibliotecas estaduais de agricultura (Maranhão, Paraná, etc.) e outras bibliotecas cooperantes com a BINAGRI ou integrantes da rede do SNIDA; *g)* ações contínuas de treinamento e aperfeiçoamento de profissionais bibliotecários, documentalistas e informáticos, vinculados à pesquisa agropecuária (ROBREDO, 2005, p. 88, 192). Uma breve história da BINAGRI encontra-se no portal da BINAGRI, atualmente vinculada ao Ministério da Agricultura, Pecuária e Abastecimento (MAPA, 2009). Uma detalhada relação dos documentos técnicos produzidos até 1980 encontra-se nos *FAO Online Catalogues* (FAO, 2009).

Podemos afirmar, se considerarmos a influência da Documentação e das idéias de Otlet na criação do IBBD e a presença norte-americana na implantação do Mestrado, que o IBICT absorveu tanto o pensamento europeu quanto dos Estados Unidos, numa confluência que certamente contribuiu para fortalecê-lo e torná-lo o órgão nacional de ICT.

Sobre a década de 80, Pinheiro e Loureiro (2004) levantam e sistematizam diferentes eventos, merecendo destaque, como componente da formação científica da Ciência da Informação, a fundação da ANCIB- Associação Nacional de Pesquisa e Pós-Graduação em Ciência da informação, em 1989 e, em política de ICT, a Ação Programada de Informação Científica e Tecnológica, em 1984, além da inclusão de ICT no PADCT- Programa de Apoio ao Desenvolvimento Cientifico e Tecnológico, com recursos significativos do Banco Mundial.

2.5.2 Um pouco de história sobre a bibliometria no Brasil

Dentro do quadro referencial anterior, num trabalho considerado clássico, como um dos primeiros estudos bibliométricos/cientométricos a apresentar uma análise detalhada da produção dos estudiosos da ciência da informação, no Brasil, desde seus primórdios, nos primeiros anos da década de 1970, abrangendo um período de doze anos (1972-1983), Urbizagástegui Alvarado (1984) faz um levantamento utilizando como fontes: *i)* o *Catálogo de Dissertações e Teses* do IBICT (IBICT,1982), *ii) ABCD: Resumos e sumários*; *iii)* o levantamento sobre a lei de Bradford, de Quemel *et alii* (1980); *iv)* um levantamento sobre bibliometria brasileira, feito pelo Centro de Informação em Ciência e Tecnologia, do IBICT; e *v)* um levantamento de todos os autores que publicaram sobre bibliometria (lei de Bradford, lei de Lotka, lei de Zipf, lei de Goffman, frentes de pesquisa e obsolescência). O resultado obtido para a produção dos autores no período considerado, está representado no *Gráfico 1*.

Gráfico 1.- Publicações no período 1972-1983 (Urbizagástegui Alvarado, 1984).

A maior quantidade de trabalhos foi produzida nos anos 1978 e 1979, com doze e nove trabalhos respectivamente, e o ano de menor produção foi 1972, com dois trabalhos somente, que representam o início dos estodos bibliométricos no Brasil.[21]

Dos 78 trabalhos produzidos até 1983, 39 correspondem à aplicação da Lei de Bradford; 11 a estudos de produtividade (Lei de Lotka); 8 às análises de citações, e 6 às aplicações da Lei de Zipf. Portanto, nesses anos houve preferência pela aplicação da lei de Bradford.[22]

[21] Ver Figueiredo (1972) e Braga (1972)

[22] Essas três leis, que se constituem na pedra angular da bibliometria, serão tratadas com detalhe na Seção 4.

No entanto, não houve continuação nas aplicações da lei de Goffmam, frente de pesquisa e obsolescência da literatura. As aplicações destas leis têm sido pouco significativas, variando entre 2 e 3 publicações, como mostra a *Tabela 1.*

Tabela 1.- Tipos de bibliometrias em teses e dissertações no período de 1972-1983

ANOS	LEI DE BRADFORD	LEI DE ZIPF	LEI DE LOTKA	LEI DE GOFFMAN	FRENTE PESQUISA	OBSOLES-CÊNCIA	CITAÇÕES
1972	1	-	-	-	1	-	-
1973	2	1	-	-	-	-	-
1974	1	2	1	1	-	-	-
1975	-	-	-	1	-	-	1
1976	2	1	-	-	-	-	1
1977	2	-	1	-	-	-	-
1978	1	-	-	-	1	1	-
1979	-	-	-	-	-	1	-
1980	2	-	1	-	-	-	-
1981	2	-	-	-	-	-	2
1982	3	-	-	-	-	-	-
1983	-	-	1	-	-	1	1

Uma análise realizada sobre as 36 teses e dissertações aprovadas no mesmo período, mostra que a produção varia entre 1 e 5, com o maior valor em 1974 e o menor em 1979, como mostra o *Gráfico 2.*

Gráfico 2.- Teses e dissertações publicadas em 1970-1983 (Urbizagástegui Alvarado, 1984).

Na *Figura 2*, pode-se observar a influência de Saracevic, que, durante sua permanência no Brasil, orientou 7 dissertações da 'primeira geração' de mestres, seguido por Lancaster (3 orientações) e Boyce (1 orientação).

Figura 2.- Rede de orientadores/orientandos na bibliometria brasileira (Ibidem).

Os resultados de outro levantamento sobre as instituições em que as teses e dissertações foram aprovadas, no mesmo período, encontram-se na *Tabela 2.* A ênfase dada ao tema bibliometria na fase inicial do processo de implantação da pós-graduação, no IBICT/UFRJ, período em que contou com o apoio decisivo de professores estrangeiros fica evidente nessa Tabela.

Tabela 2.- Instituições, em que foram aprovadas as teses e dissertações, no período 1972-1983, e quantidades aprovadas (Urbizagástegui Alvarado, 1984).

INSTITUIÇÕES	APROVAÇÕES
IBICT/UFRJ	28
ECA/USP	2
UFMG	1
UnB	1
UFRJ	1
Denver University	1
Case West Reserve University	1
Drexel University	1
TOTAL	36

É interessante observar que o principal orientador, na fase de decolagem do mestrado do IBCT/UFRJ, dentre os orientadores – que o referido autor chama de 'introdutores' – da primeira geração de mestres, cede a estes a condição de novos orientadores, os quais. por sua vez, orientam os mestres da segunda geração, etc.

O trabalho de Urbizagástegui Alvarado (1984) é, sem dúvida, um trabalho que deve ser levado em conta, não somente pelos estudiosos da história da bibliometria no Brasil, antes do advento da Internet, mas também por todos aqueles que desejem conhecer as raízes das tendências e possibilidades atuais das metrias da informação em ambiente Web.[23]

Num trabalho de Pinheiro, já citado anteriormente (PINHEIRO, 2005), em que estuda a evolução e tendências da ciência da informação, no Brasil e no exterior, são apresentados os resultados de levantamentos realizados nos períodos de 1966-1995 e 1996-2004 pela *Anual Review for Information Science and Technology* (*ARIST*, 1996; 2005) que apontam para as disciplinas que deveriam constituir o núcleo básico dos cursos de ciência da informação, nos Estados-Unidos, onde *bibliometrics* aparece na 13ª posição (Ver *Tabela 3*).

Tabela 3.- Núcleo básico de disciplinas da Ciência da Informação e tendências atuais, por resultado de estudo de freqüência de artigos de revisão do ARIST.

DISCIPLINAS	FREQÜÊNCIA
1. Sistemas de informação	49
2. Tecnologia da informação	36
3. Sistemas de recuperação da informação	35
4. Políticas de informação	28
5. Necessidades e usos de informação	25
6. Representação da informação	25
7. Teoria da Ciência da Informação	16
8. Formação e aspectos profissionais	16
9. Gestão da informação *	14
10. Bases de dados	14
11. Processamento automático da linguagem	11
12. Economia da informação	10
13.Bibliometria	6
14. Inteligência competitiva e Gestão do conhecimento	5
15. Mineração de dados	5
16. Comunicação científica eletrônica	3
17. Bibliotecas digitais/virtuais	2

No mesmo trabalho, a autora fez um levantamento das freqüências com que aparecem os termos significativos nos artigos publicados na revista *Ciência da Informação*, entre 1972 e 2004, ou seja, num período equivalente ao considerado pela *ARIST*, levando em conta a defasagem com que se iniciam no Brasil os estudos bibliométricos quando comparados aos Estados Unidos. No caso da *Ciência da Informação*, o resultado para a bibliometria é bem mais favorável (Ver *Tabela 4*).

Embora os dados da *ARIST* e da *Ciência da Informação* não sejam totalmente comparáveis, pois estes últimos se referem só a essa revista, o resultado é significativo pela importância do IBICT e pela forte representatividade da sua revista no Brasil.

[23] Urbizagástegui Alvarado (1984), indica uma completa bibliografia sobre os autores citados, além de cerca de 80 referências complementares, num anexo que reúne a bibliografia sobre a bibliometria brasileira, dividida em seus aspectos teóricos, as quatro 'leis' (Bradford, Zipf, Lotka, Goffman), obsolescência e análise de citações).

Tabela 4.- Freqüência de temas nos artigos publicados na revista Ciência da Informação (adaptado de Pinheiro (2005)

TEMAS	1972-79	1980-89	1990-99	2000-04	TOTAL
1. Teoria da Ciência da Informação	4	8	23	14	49
2. Bibliometria	18	11	14	5	48
3. Representação da informação	8	9	15	14	46
4. Políticas de informação	3	11	12	17	43
5.Necessidades e usos de informação	6	9	16	11	42
6. Gestão da informação	2	7	25	7	41
7. Comunicação científica	7	10	12	10	39
8.Tecnologias da informação	-	1	18	18	37
9. Formação e aspectos profissionais	6	9	15	4	34
10. Sistemas e redes de informação	6	6	17	3	32
11. Disseminação da informação	4	7	16	4	31
12. Sistema de recuperação da informação	4	2	6	9	21
13. Bibliotecas virtuais/digitais	-	-	6	11	17
14. Inteligência competitiva	-	-	8	5	13
15. Política de Ciência e Tecnologia	-	3	7	2	12
16. Bases de dados	3	4	2	2	11
17. Organização/processamento da inform.	6	1	1	2	10
18. Economia da informação	-	2	5	2	9
19. iblioteconomia/Bibliotecas/Livros	2	3	-	2	7
20. Arquivologia	-	-	6	-	6
21. Processamento automático da linguagem	1	1	4	-	6
22. Automação de bibliotecas	2	1	2	1	6
23. Divulgação científica	-	1	2	1	4
24. Ciência e Tecnologia	-	-	3	-	3
25. Gestão do conhecimento	-	-	-	3	3
26. Política editorial	2	1	-	-	3
27. Sistemas especialistas	-	-	2	-	2
28. Comunicação social	-	-	2	-	2
29. Imprensa	-	-	2	-	2
30. Lingüística	-	-	2	-	2
31. Tecnologias	-	-	2	-	2
Temas com freqüência 1	2	2	5	1	10
Total	86	109	250	148	293

Cotejando esses dados com os apresentados por Urbizagástegui Alvarado (1984), se confirma um pico de produtividade na década de 1970, que se mantém, aproximadamente, até o fim do século XX, com uma queda significativa nos primeiros anos do presente século.

O quadro das relações orientador/orientando, registrado por Urbizagástegui Alvarado (1984) no que diz respeito às dissertações e teses aprovadas no curso de pós-graduação do IBICT, no período 1972-1998 (ver *Figura 2*, acima), foi atualizado por Pinheiro e Silva (2008) até 2008, num trabalho mais recente (ver *Quadro 5*).

Essas análises têm um foco predominantemente orientado para o IBICT. Por isso, buscando obter uma visão mais abrangente da produção bibliométrica no Brasil, destacamos alguns autores, de outras instituições, lembrados por Pinheiro e Silva, agrupados em dois conjuntos: *i)* docentes e/ou estudiosos que não exercem atividades em ciência da informação, mas desenvolvem pesquisas em metrias da comunicação científica ou indicadores de ciência e tecnologia, e *ii)* aqueles que atuam na área, com graduações diversas.[24]

[24] São eles (na ordem em que aparecem nesse trabalho):

i) Docentes e/ou estudiosos que desenvolvem pesquisas em metrias da comunicação científica ou indicadores de ciência e tecnologia:

Professora Lea Velho (UNICAMP), doutorado em Política Científica e Tecnológica, pela *Science Policy Research Unit* (SPRU), *University of Sussex*, em 1985, com produção intensa, no Brasil e exterior, forte atuação em política científica e tecnológica, inclusive indicadores de C&T. A autora pesquisa na área desde o início da década de 80 (VELHO; KRIGE, 1984) e (VELHO, 1985, 1986). Segue uma linha de estudos sociais da ciência, algumas mais quantitativas. Forte influência de Solla Price, bem como de Ziman, Kuhn, Merton, Bronowski e Bunge, entre outros (LIMA; VELHO, 2008).

Pesquisador Rogério Meneghini (USP), cientista de renome nacional e internacional em Bioquímica, a partir do final da década de 1980 se dedica à comunicação científica, política de C&T e indicadores, questões sobre as quais publicou trabalhos, no exterior e no Brasil. Uma de suas primeiras publicações nessa linha trata de indicadores alternativos de avaliação da produção científica no Brasil (MENEGHINI, 1988). Foi um dos criadores do SciELo. Em alguns de seus trabalhos tem contado com a colaboração de especialistas em ICT, como Abel Packer, da BIREME (MUGNAINI; PACKER; MENEGHINI, 2008) e Jacqueline Leta, (MENEGHINI; LETA, 2002), incluída entre os autores estudados.

Professor Pierre Ohayon, da Faculdade de Administração e Ciências Contábeis (FACC), da UFRJ. Formado e pós-graduado em administração, especializado em gestão e avaliação de C&T (MARCOVITCH; OHAYON, 1984), (MORAES; OHAYON; LEITÃO,1989). Mais recentemente, propôs um modelo para a avaliação de programas de pesquisa, desenvolvimento tecnológico e inovação (OHAYON, ROSENBERG, 2007).

Professora Jacqueline Leta (UFRJ), a mais jovem, mas com expressiva e consistente produção científica, no Brasil e exterior. Oriunda da área de Ciências Biológicas, tem mestrado e doutorado em Gestão, Educação e Difusão de Ciências, também pela UFRJ. Especialista em produção científica nacional, é co-autora do livro *"O perfil da ciência brasileira"* (MEIS, LETA, 1996) e autora de trabalhos sobre indicadores de C&T, o que inclui a pós-graduação brasileira, na sua relação com a produtividade científica (LETA, LANNES, MEIS, 1997).

i) Pesquisadores/docentes vinculados à ciência da informação:

Professor Jaime Robredo, um dos precursores da área no Brasil. Com formação em Química, inclusive pós-doutorado na Holanda, exerceu atividades no Serviço de Documentação do Instituto do Vidro, em Paris, por mais de dez anos. No Brasil, a partir de 1974 continuou nessa linha de estudos tendo devotado a sua vida acadêmica na UnB à ciência da informação. Um dos seus primeiros trabalhos, no Brasil, foi elaborado na década de 1970 (ROBREDO; CHASTINET; LOBO, 1975), e o foco de suas pesquisas são as técnicas bibliométricas nos processos de indexação e recuperação da informação, bem como listas básicas de periódicos, inclusive em Agricultura. Atualmente, é líder do Grupo de pesquisa "Representação e Organização da Informação e do Conhecimento", que inclui entre as linhas de pesquisa *e*studos métricos da informação (ROBREDO; CANTANHEDE, 2005). Participa de seu Grupo de Pesquisa a *professora Marisa Brascher*, também da UnB, que já apresenta produção em bibliometria (ver, Costa *et al*.. (2006), e Café e Brascher (2008)).

Professora Suzana Mueller, também da UnB, há longo tempo dedica-se à comunicação científica. É líder de Grupo de pesquisa "Comunicação Científica", e foi coordenadora do grupo de trabalho GT7 – "Produção e Comunicação da Informação Científica", da ANCIB (2005-2006). A temática de seus estudos tem sido o periódico científico, incluindo o estabelecimento de listas básicas, e produção científica. Sobre essas questões tem publicado inúmeros trabalhos, adotando metodologias quantitativas, desde o início dos anos 1990 (MUELLER, 1991), continuando nessa abordagem até os dias atuais, conforme comprova um de seus mais recentes trabalhos (MUELLER, 2008).

Professora Sely Costa, também da UnB é também líder do Grupo de pesquisa "Comunicação Científica", com estudos voltados ao acesso aberto, que incluem indicadores de sustentabilidade de periódicos eletrônicos brasileiros (COSTA *et al*.. 2006). Atualmente, Diretora da Biblioteca Central da UnB.

Professora Dinah Población (USP), que há muitos anos desenvolve pesquisas sobre produção científica em diferentes campos do conhecimento como saúde (POBLACIÓN, 1986), e especialmente na ciência da informação, numa contribuição relevante para o mapeamento e desenvolvimento da área. Fundadora, em 1992, do Núcleo de Pesquisa de Produção Científica, na USP/ECA, que congrega inúmeros pesquisadores e cuja produção tem sido divulgada em diversas publicações (POBLACIÓN, 2005). Entre os trabalhos mais recentes merece destaque um artigo sobre produção científica em Ciência da Informação (POBLACIÓN, 2005).

Professora Ida Stumpf (UFRGS). É líder do Grupo de pesquisa "Comunicação Científica", abrangendo as metrias da informação. Coordena atualmente o grupo de trabalho GT7 – "Produção e Comunicação da Informação Científica", da ANCIB Na sua atuação acadêmica, principalmente a partir dos anos 2000, tem produzido e orientado nessa linha, na pós-graduação, na qual conta com o reforço da professora *Sonia Caregnato* (STUMPF *et al*., 2006), (VANZ; CAREGNATO, 2003). Nas suas pesquisas, os métodos bibliométricos são aplicados sobretudo à produção cientifica da Comunicação e da Ciência da Informação, o que inclui análises de citações (STUMPF; MESQUITA, 2004).

Professora Maria Cristina Hayashi (UFSC), com produção em metrias da informação, especialmente dos anos 2000 em diante. Formada em ciências sociais, com pós-graduação em educação (HAYASHI, *et al*., 2008).

Professor Raimundo Nonato Macedo dos Santos (UFSCAR, anteriormente PUCCAMP). Graduado em engenharia civil, passou a se dedicar às metrias da comunicação científica nos anos 1990, o que pode ter sido decorrência de seu mestrado e doutorado na *Université Aix Marseille III*, conhecida pelos seus avançados estudos bibliométricos e outras metrias da informação, com aplicativos próprios (ver acima, Seção 2.1 *A corrente européia ocidental*). Dentre os últimos trabalhos do professor Santos, que analisam principalmente a produção da ciência da informação, um dos mais recentes trata das dissertações e teses (SANTOS, *et al*., 2007).

Para encerrar este demorado sobrevôo, e procurando ter uma idéia mais ampla do atual e real interesse pelos indicadores bibliométricos, cientométricos e infométricos, em geral, especialmente em domínios em expansão tais como as vigílias e observatórios estratégicos (inteligência competitiva, propriedade industrial, apoio à decisão, análise de tendências e estudos prospectivos, serviços de recomendação, etc.), foi feita uma pesquisa na Web, e mais concretamente na base dos currículos Lattes, do CNPq. Os resultados estão reunidos na *Tabela 5.*

Sem nenhuma pretensão de julgamento sobre o valor científico desses valores, nem sobre a maior ou menor confiabilidade dos motores de busca utilizados, poder-se-ia avançar que existe – ou existiu – de uma parte, uma preferência pelo tema bibliometria, usado freqüentemente como sinônimo ou conjuntamente com infometria/informetria e, de outra parte, um crescente interesse pelo tema cientometria/cienciometria que ultrapassa os limites – se é que existem – da ciência da informação.

Como conseqüência da maior facilidade de acesso a dados e informações em ambientes Web, e da necessidade de se descobrir novos nichos de inovação e desenvolvimento, para sobreviver e vencer num mundo cada dia mais globalizado e mais competitivo, embora os valores acima apresentados não o comprovem de forma contundente, e levando em conta que se trata de uma prática recente, tudo parece indicar que o uso da Web como meio de acessar dados, informações e documentos tende a crescer e, assim, os termos webmetria, webometria, cibermetria, e outras 'netmetrias', serão cada dia mais freqüentes, nas pesquisas brasileiras.

Quadro 5.- Relação orientador/orientando organizada por ano de titulação (adaptado de Pinheiro (2008).

ORIENTADOR	ORIENTANDO-ANO	TOTAL
BRAGA, Gilda Maria[25], IBICT	(Carvalho, 1975); (Sá, 1976); (Braga, H., 1979); (Christóvão,1979); (Ferrez, 1981); (Pinheiro, 1982); (Costa, 1982); (Nascimento, 1983); (Neri, 1984); (Peixoto, 1985); (Cavalcanti, 1989); (Souza, 1989); (Barreto, 1991); (Coelho, 1991)*; (Russo, 1992);(Santos, 1992); (Maranhão, 1996).	17
PINHEIRO, Lena Vania R., IBICT	(Rodrigues Garcia, 1988); (Valois, 1990); (Coutinho, 1991); (Brasil, 1992)*; (Silva, J. 1995)*; (Souza,1995)*; (Martins, 2000); (Silva, M. 2000); (Nicolau, 2008).	9
SARACEVIC, Tefko, (professor visitante)	(Braga, G. 1972); (Figueiredo[26], 1972); (Fernandez, 1973); (Ippolito, 1973); (Maia, 1973);(Dussilek, 1974); (Di Giorgi, 1974); (Paranhos, 1975); (Oliveira, 1976).	9
FIGUEIREDO; Laura Maia, IBICT	(Gomes, 1977); (Gusmão, 1977); (Nosetti Menendez, 1978); (Rodrigues, 1981).	4
LANCASTER, Frederick Wilfrid, (prof. visitante)	(Barbosa, 1979); (Queiroz, 1979); (Ribeiro, 1979).	3
SOUZA, Rosali F. de, IBICT	(Weltman, 1992)*; (Decourt, B, 1993); Pedrini (2000).	3
	TOTAL	**45**

* *Orientações nas quais participaram como co-orientadores.*

[25] Gilda Maria Braga defendeu sua dissertação de mestrado, orientada por Tefko Saracevic, em 1972 (Ver *Seção 8 – Bibliografia Consultada*), e a primeira orientadora brasileira do Mestrado em Ciência da Informação do IBBD-UFRJ, com o maior número de dissertações orientadas, no período de 1975 a 2000.

[26] Laura Maia Figueiredo foi autora da primeira dissertação sobre bibliometria defendida no Mestrado em Ciência da Informação do IBBD-UFRJ. (Ver *Seção 8 – Bibliografia Consultada*)

Tabela 5. – Ocorrências encontradas na base Lattes de termos relacionados com as metrias de informação, usando diferentes motores de busca.

TERMOS DE BUSCA	GOOGLE	YAHOO/ALTAVISTA	ASK
Lattes infometria Brasil	379	77	91
Lattes informetria Brasil	571	107	164
Lattes bibliometria Brasil	5830	2760	1710
Lattes cientometria Brasil	2240	680	1010
Lattes cienciometria Brasil	1460	581	331
Lattes webmetria Brasil	279	59	91
Lattes webometria Brasil	262	46	84
Lattes cibermetria Brasil	134	16	10

3. TENTATIVA DE FIXAR A TERMINOLOGIA

Não pretendemos, nesta Seção, chegar a uma solução do problema da crescente babelização da terminologia geralmente usada quando se fala de metrias da informação. Mas pelo menos, esperamos contribuir de alguma forma para a abertura de uma corrente de reflexão e diálogo, que nos aproxime de um consenso suscetível de facilitar a localização de um tema, ou de evitar 'silêncios' nas respostas, como conseqüência do uso de um determinado termo e da omissão de um ou mais sinônimos possíveis.

A situação se agrava quando se indica aos motores de busca que a pesquisa deve-se realizar 'na Web' e não em tal ou qual país específico pois, nesse caso, as perdas de boas respostas são multiplicadas pelo número de sinônimos omitidos nas outras possíveis línguas!!!!

Na *Tabela 6*, encontram-se os resultados correspondentes ao número de páginas recuperadas nas buscas com diversos termos e seus sinônimos, em português e em outras linhas.

Antes de indicar quais seriam nossas sugestões de escolha dos termos, na língua portuguesa, pode ser útil fazer uma breve reflexão sobre a crescente frequência de uso, em trabalhos brasileiros, dos termos cienciometria, webometria e informetria.

Quanto a primeiro, reproduzimos o texto da Nota 5, referente ao uso dos termos cientometria e cienciometria, do trabalho já citado de Pinheiro e Silva (2008):

> *Pela etimologia, Cientometria deriva de 'scientia', do latim, daí cientistas e científico, em português, ou Scientometrics, em inglês, portanto, a denominação cienciometria não seria correta em nossa língua, embora comumente utilizada em espanhol.*

Quanto aos termos webmetria e webometria, reproduzimos a Nota 6 do referido trabalho, que diz: *"Na literatura aparece o termo webometria e, em maior escala, webmetria, adotado no presente trabalho."* Vistos os resultados apresentados na *Tabela 5*, é claro que só podemos concordar com as escolha das autoras. Já no que diz respeito ao uso de 'informetria', em português, as autoras usam o termo, mas não apresentam as razões que justificariam sua escolha, limitando-se a informar:

> *Outro termo surge, em 1987, informetria, cunhado e adoptado pela FID – Federação Internacional de Documentação para designar 'o conjunto de atividades métricas, relativas à informação, cobrindo também a Bibliometria e a Cientometria" (EGGHE; ROUSSEAU, 1990).* [27]

Sobre a escolha de informetria, frente ao termo infometria, parece difícil concordar, e isso por uma razão que parece possuir um certo peso: porque até essa data de 1987, todo o mundo usava o termo infometria, inclusive no Brasil. Pode-se acrescentar que, como mostram as *Tabelas 5* e *6*, o número de ocorrências de uso do termo infometria continua sendo superior nos próprios países que publicam em língua inglesa, sendo exceção, a partir de uma data recente, Brasil, Portugal, Espanha e França, sem que até agora, ninguém tenha apresentado alguma razão plausível.

No intuito de apresentar uma visão da freqüência de uso, no Brasil, dos termos que foram utilizados como critérios de busca, na pesquisa que serviu para construir a *Tabela 6*, foi feita outra pesquisa com os mesmos termos, em português, na base ABCDM[28], na área da ciência da informação (*Gráfico 3*).

---

[27] Ver, a esse respeito, as observações de Brookes recolhidas na *Seção 2.2 – A corrente russa e européia oriental.*

[28] Ver Nota 40, na página 54

Tabela 6. – Número de páginas recuperadas nas pesquisas com os termos indicados, em diversas línguas, utilizando o Google como motor de busca.

Termo de busca	Idioma	Páginas	Termo de busca	Idioma	Páginas
Bibliometrics	en	1.510.000	Scienciometrics	en	1.020
Bibliometría	es	30.600	Cienciometría	es	6.470
Bibliométrie	fr	56.300	Scienciométrie	fr	0
Bibliometria	it	1.290	Scienciometria	it	0
Bibliometria	pt	16.700	Cienciometria	pt	5.270
Infometrics	en	131.000	Scientometrics	en	123.000
Infometría	es	10.650	Cientometría	es	129
Infométrie	fr	16.200	Scientométrie	fr	44.400
Infometria	it	109	Scientometria	it	429
Infometria	pt	7.060	Cientometria	pt	7.140
Informetrics	en	52.400	Webmetrics	en	6.080.000
Informetría	es	28.600	Webmetría	es	3.310
Informétrie	fr	88.000	Webmétrie	fr	42
Informetria	it	6	Webmetria	it	0
Informetria	pt	1.960	Webmetria	pt	1.470
Cybermetrics	en	39.000	Webometrics	en	344.000
Cibermetría	es	52.300	Webometría	es	378
Cybermétrie	fr	14.700	Webométrie	fr	1.820
Cibermetria	it	46	Webometria	it	3
Cibermetria	pt	255	Webometria	pt	815

Gráfico 3. – Número de artigos científicos publicados no Brasil, em que aparecem os termos que figuram na Tabela 6, por ano.

(Elaboração J. Vilan Filho)

Pode-se observar o crescimento do interesse em pesquisas sobre "metrias" a partir de 1998, bem como o aumento da variedade de termos utilizados, com destaque para bibliometria, cienciometria/cientometria e infometria/informetria. O termo 'webmetria´ não apresentou nenhuma ocorrência até o ano de 2007, última data considerada. Os termos foram extraídos dos títulos, subtítulos e palavras-chave dos periódicos das áreas de informação.

Como elemento para reflexão, os três pontos de interrogação ´???´, no título da comunicação de Brookes, publicada em 1989, nos *Proceedings of the 1st International Conference on Bibliometrics and Technical Aspects of Information Retrieval.* [29]

Assim, nos permitimos sugerir uma nova, ampla e profunda discussão em favor da adoção, em nossas publicações em português e/ou em inglês, do esquema terminológico mostrado no *Quadro 2.*

Isso não impede – bem pelo contrário – desenvolver classificações, taxonomias, tesauros, ontologias, mapas temáticos, léxicos e vocabulários, que levem em conta as indicações de sinonímia, equivalência, relação, entre termos, que se fizerem necessárias.

Dessa forma, os motores de busca poderão encontrar sua tarefa de nos fornecer boas respostas, consideravelmente facilitada, e nós agradeceremos.

Quadro 2. – Proposta de esquema terminológico para as metrias da informação

Cibermetria → **Webmetria**,
Cienciometria → **Cientometria**
Informetria → **Infometria**
Webometria → **Webmetria**,

Bibliometria *(pt)* e Bibliometrics *(en)*,
Cientometria *(pt)* e Scientometrics *(en)*,
Infometria *(pt)* e Infometrics *(en)*,
Webmetria *(pt)* e Webmetrics *(en)*.

(Elaboração J. Robredo)

As setas apontam para os termos (em negrito) que nos parece que deveriam ser usados preferentemente em português. No bloco inferior se apresenta um pequeno glossário português-inglês que pode ajudar a escolher os melhores termos ingleses, na hora de converter os resumos de nossos trabalhos em *abstracts.*

4. METRIAS DA INFORMAÇÃO E CIÊNCIA DA INFORMAÇÃO ANTES DE WEB

No que segue, será utilizada a terminologia apresentada na Seção precedente. Cabe esclarecer que os termos 'infometria' e 'metrias da informação' serão considerados sinônimos, e qualquer outro termo que não se enquadre nesse esquema somente será utilizado – se for o caso – em citações que reproduzem *ipsis litteris* textos de outros autores.

Quanto às relações hierárquicas entre os termos relativos às metrias da informação – ou infometria – seguiremos o esquema apresentado anteriormente na *Figura 1*, na Seção 2.1, nos distanciando, assim, do esquema apresentado por Tina Jayroe, da Universidade de Denver, em sua bem conhecida apostila "*Bibliometrics for Dummies*" (JAYROE, 2008),. Segundo essa autora, com base em Bates (1999, p. 1048), "[...] a bibliometria contribui para melhor entender o universo informacional, pois [ela] é o estudo de como as pessoas se relacionam com a informação." E, parafraseando Rousseau (1990, p. 197), prossegue: "[...] a pesquisa bibliométrica (ou informétrica [*sic*]) desenvolveu um rico corpus de conhecimento teórico que, pela sua vez, busca encontrar novas aplicações de todos os aspectos práticos do trabalho com informação."

[29] Ver Brookes (1989), na ´Bibliografia consultada', no fim do Capítulo.

A autora acrescenta:

> Em biblioteconomia [‘library science’, no original], bibliometria se refere, geralmente, à análise das características de dados e informações relativos à literatura – especialmente de autoria acadêmica – se utilizando de certos métodos e “leis”. Isso tem um impacto significativo na área de recuperação da informação e os pesquisadores estão utilizando atualmente a bibliometria para estudar estruturas relacionais na Internet, o que se denomina, também, ‘webometrics’ [sic]. Para concluir, todas as aplicações da bibliometria contribuem para tornar explícita a meta-informação no paradigma da ciência da informação (JAYROE, 2008).

A continuação, serão apresentadas alguns aspectos relativos ao que poder-se-ia denominar “leis” empíricas e aplicações tradicionais da infometria. Segundo Chen e Leimkuhler (2006, p. 308, *apud* Jayroe, 2008), “[...] as principais leis empíricas da bibliometria são a lei de Bradford a lei de Lotka e a Lei de Zipf [...] aquelas que foram usadas com maior freqüência na literatura para modelar a distribuição dos fenômenos informacionais.” Embora as três leis apontem para distribuições probabilísticas similares, suas diferenças decorrem do tipo de dados analisados. Dados esses que Diodato (1994, p. 93; *apud* Jayroe, 2008) denomina itens e fontes (‘*items*’ e ‘*sources*’).

4.1 Lei de Bradford[30]

Nas aplicações das analises baseadas na lei de Bradford, os itens são, geralmente, artigos de publicações periódicas, e as fontes os periódicos que publicam os artigos. A lei de Bradford é a lei bibliométrica por excelência. É uma excelente orientação para o desenvolvimento de coleções em campos de aplicação específicos.

Para quem não está familiarizado com a linguagem matemático-estatística, pode-se dizer que a lei mostra que um certo número de periódicos, num campo determinado pode ser dividido em três parcelas ou grupos, cada um contendo o mesmo número de artigos:

i) O primeiro grupo de periódicos constitui o núcleo básico de títulos de periódicos que contêm artigos de interesse sobre o tema e é formado por um número relativamente pequeno de títulos de periódicos que reúnem um terço, aproximadamente, do total de artigos pertinentes contidos no total de periódicos reunidos;

ii) O segundo grupo contém o mesmo número de artigos distribuído em um número maior de periódicos;

iii) O terceiro grupo contém o mesmo número de artigos interessantes que a que os precedentes, mas um número bem maior de títulos de periódicos.

A relação do número de títulos de periódicos no grupo nuclear (que contêm um terço dos artigos interessantes) para o número de títulos de periódicos necessário para reunir outro terço de artigos interessantes é, digamos, n vezes maior que no grupo nuclear, a relação dos títulos de jornais entre o primeiro e o segundo grupo é $1/n$, e a relação de periódicos entre o grupo nuclear e o terceiro grupo, para que este último reúna o último terço de artigos interessantes, deverá ser $1/n^2$. Em outras palavras, a relação de periódicos nos três grupos, se cada cada grupo contém um terço (1/3) do total de artigos de interesse é:

$$1 : n : n^2. \qquad [1]$$

Bradford formulou a lei após estudo detalhado de uma bibliografia sobre geofísica que reunia 326 títulos de periódicos da área, e encontrou que os 9 periódicos mais produtivos em artigos interessantes continham 429 artigos interessantes, 59 títulos continham 499 artigos interessantes, e 258 periódicos continham 404 artigos interessantes. Tratemos de reconstruir a experiência:

1) Os periódicos são ordenados em ordem decrescente de produtividade de artigos de interesse, totalizando 326 títulos diferentes;

2) É contado o total de artigos de interesse, no total de títulos de periódicos, totalizando 1.332 artigos;

3) Deseja-se que o total dos periódicos ordenados de forma decrescente seja dividido em três grupos, de forma que cada grupo venha a conter a mesma quantidade de artigos interessantes. Para tanto, o total de artigos registrado acima divide-se por 3, resultando 443 artigos para cada grupo;

4) Resta recontar os artigos em cada periódico, respeitando a ordem decrescente e separando um primeiro grupo de periódicos quando se atingir um total de, aproximadamente, 443 artigos de interesse, e repetir a contagem para delimitar o número de periódicos no segundo grupo. Os periódicos restantes, integrantes do terceiro grupo, deverão conter, também aproximadamente, 443 artigos interessantes.

Observe-se que os resultados obtidos por Bradford, com 429 artigos no primeiro grupo de 9 periódicos mais produtivos, 499 artigos nos 59 periódicos seguintes, e 404

[30] Samuel C. Bradford (1878-1948), matemático e bibliotecário inglês. Diretor da biblioteca do *Science Museum* de Londres, de 1925 a 1937. Em 1934 formulou sua lei, conhecida inicialmente como “lei da dispersão”, publicada em *Engineering,* no mesmo ano (BRADFORD, 1934; *apud* Hjørland, 2006). Fundador da *British Society for International Bibliography*, foi presidente da FID em 1945. Grande incentivador do uso da CDU e de resumos.

no terceiro, com 258 periódicos, não estão nada mal. Concordamos com Potter, quando afirma "[...] a Lei de Bradford não é, falando com rigor, estatisticamente exata. Mas, ainda é uma boa regra prática." (POTTER, 1988; *apud* JAYROE, 2008).[31] Senão, vejamos: a média dos valores experimentais é 444 artigos em cada grupo (429+499+404/3), e a calculada no passo 3, como já indicado, 443 artigos.

A relação [1] acima, no caso da experiência de Bradford, se converte, aproximadamente, em:

$$1 : 5 : 5^2 \qquad [2]$$

ou:

$$1 : 5 : 25. \qquad [2']$$

Pensando um pouco, poder-se-ia chegar à conclusão que com 9 assinaturas bem escolhidas se cobririam, naquela época, 66,67 por cento dos conhecimentos da área de geofísica, com economia do preço da assinatura de 317 periódicos de baixa – ou baixíssima – produtividade. A história da biblioteconomia não registra se, após a formulação de sua lei, Bradford aplicou a descoberta à fundamentação de uma nova política de aquisição e de desenvolvimento de coleções, que poderia ter levado a economizar rios de dinheiro ao Museu de Ciências de Londres. Cabe imaginar que sim, vista sua fama mundial e brilhante carreira, após sua saída da biblioteca do Museu, que culminou na sua nominação como presidente da então Federação Internacional de Documentação (FID), em 1945.

O que a história da biblioteconomia brasileira pode dizer, até o momento de escrever estas linhas – se não estamos mal informados – é que a aplicação da lei de Bradford e outras metrias da informação não parecem ser muito aplicadas nas bibliotecas universitárias e outras bibliotecas, nem ensinadas nos cursos que formam os futuros profissionais.

Para colocar os dados resultantes das contagens e de sua distribuição nas três zonas de igual conteúdo de artigos de interesse, não e necessário fazer grandes cálculos. Basta, de fato, colocar em uma folha de papel semi-logarítmico – que pode ser adquirido em uma boa papelaria – os dados acumulados do número de periódicos de cada zona (três zonas, ou mais se se deseja, cuidando que o número de artigos interessantes em cada zona se mantenha aproximadamente constante) no eixo X do gráfico (em escala semi-logarítmica), e o valor acumulado de artigos interessantes, no eixo Y do gráfico (em escala linear). O resultado será semelhante ao mostrado na *Figura 3*, que é uma representação gráfica genérica da lei de Bradford.

O traçado da curva tem a forma de um "J" inclinado ou de um "S" inclinado e esticado. As três zonas referidas, que contêm, aproximadamente, o mesmo número de artigos, distribuídos em um número de periódicos, que cresce na medida em que diminui sua produtividade em artigos de interesse, foram demarcadas pelos três quadrados pontilhados (ROBREDO, 2000; p. 88-93). [1]

Figura 3. – Representação gráfica da Lei de Bradford. (No eixo X, a escalaé logarítmica, e linear no eixo Y)

[31] *"Law of thumb"*, no original.

Outra possibilidade, ainda mais simples, consiste em montar uma tabela com os dados necessários utilizando uma planilha, e pedir a um aplicativo adequado[32] que monte o gráfico.

A *Tabela 7* reúne todos os dados resultantes do experimento de Bradford, organizados para facilitar a geração do gráfico. Os valores da última coluna foram calculados pelo aplicativo, a partir dos totais acumulados dos títulos dos periódicos, que figuram na terceira coluna.

Tabela 7. – Organização dos dados para facilitar a apresentação do gráfico da lei de Bradford com o Excel.

ZONAS DO GRÁFICO	NÚMERO DE PERIÓDICOS EM CADA ZONA	TOTAIS ACUMULADOS DOS PERIÓDICOS	NÚMERO DE ARTIGOS EM CADA ZONA	TOTAIS ACUMULADO DOS ARTIGOS	LOG DECIMAL DOS VALORES DA TERCEIRA COLUNA
1	9	9	429	429	2,632457
2	59	68	499	928	2,967548
3	258	326	404	1.332	3,124504

O *Gráfico 4* mostra o resultado final. Observe-se que o traçado do gráfico se reduz a dois segmentos, já que foram consideradas somente as três zonas de Bradford. Para obter uma curva, com o aspecto da representada na figura anterior, seria necessário repartir os títulos em um número de zonas consideravelmente maior. O importante, entretanto, é observar que ponto de inflexão do gráfico permite ver claramente, como observado anteriormente, que um terço dos títulos dos permite reunir dois terços dos artigos interessantes.

Gráfico 4. – Gráfico do experimento de Bradford.

[32] Por exemplo, o Excel.

A lei de Bradford faz parte de uma multidão de outras "leis", aplicadas em outras áreas para medição de informações, conhecidas com diversos nomes: lei de Pareto, lei do menor esforço, regra do 80/20, etc. Ainda no campo da biblioteconomia/bibliometria, é bem conhecido que 80% dos volumes usados numa biblioteca correspondem somente 20% do acervo. (RABER, 2003, p. 73; *apud* JAYROE, 2008). A aplicação da 'curva ABC', uma aplicação gerencial ao controle de estoques, baseada na lei de Pareto deveria merecer maior atenção por parte dos gestores de nossas bibliotecas acadêmicas, para melhorar a gestão dos estoques – perdão!! acervos – bibliográficos.[33] Bem que, para isso, também ajudaria bastante um controle mais rigoroso do material consultado em sala de leitura, especialmente em sistemas de estante aberto, onde os itens usados, deixados sobre as mesas, são repostos nas prateleiras sem nenhum controle de uso. Ver também Archuby (2006).

4.2 Lei de Lotka

Nas análises baseadas na lei de Lotka[34], as fontes são os autores e os itens os documentos produzidos por eles. A lei utilizada para identificar os autores mais prolíficos dentro de uma área, mostra que, geralmente, uns poucos autores contribuem com o maior número de publicações (HERTZEL, 2003, p.303; *apud* JAYROE, 2008). Segundo Lotka *"... o número de autores com n contribuições é, aproximadamente, 1/n² o número de aqueles com uma única contribuição; e a proporção de todos os autores que contribuem com um trabalho só é, aproximadamente 69 por cento."* (LOTKA, 1926; *apud* POTTER, 1988). Isso significa que, de todos os autores que publicam num determinado campo, 60 por cento terão uma publicação só, e 15 por cento terão duas publicações (1/2² vezes 60); 7 por cento dos autores terão três publicações (1/3² vezes 60), e assim por diante. De acordo com a lei de Lotka, também chamada 'lei da produtividade científica', somente seis por cento dos autores que publicam em um campo, publicam mais de dez artículos. A lei, quando aplicada a grandes volumes de publicações, durante um longo período de tempo, pode ser suficientemente correta, mas não estatisticamente exata (POTTER, 1988, *apud* UTEXAS [s.d.]). A lei pode ser representada pela fórmula:

$$x^n \cdot y = C \qquad [3]$$

onde: "[...] x é o número de publicações; y é o número de autores que produzem x trabalhos cada um; n é um parâmetro que, em áreas científica tem, geralmente, o valor 2; e C é uma constante..." (DIODATO, 1994, p. 105, 106; *apud* JAYROE, 2008). Em outras palavras, a equação acima significa que o número total de autores y, que trabalham numa área determinada, e produzem x publicações cada um, é inversamente proporcional à função exponencial x^n. Por isso, a lei de Lotka denomina-se, também, "lei dos inversos dos quadrados da produtividade científica". Tem sido discutido se a lei de Lotka é uma lei bibliométrica, com aplicações em biblioteconomia (HERTZEL, 2003, p. 304, 305; *apud* JAYROE, 2008). A opinião de Raber (2003, p. 72; *apud* JAYROE, 2008) é que pode ser utilizada com proveito em biblioteconomia, no desenvolvimento de coleções, bem como no aprimoramento da recuperação da informação. Em todo caso, vista a data em que foi estabelecida, a lei de Lotka é uma aplicação pioneira no campo da cientometria.

4.3 Lei de Zipf

Na lei de Zipf[35], as fontes são textos ou, mais precisamente, a posição que as palavras do texto ocupam quando ordenadas de acordo com sua freqüência, e os itens são o número de ocorrências de cada palavra no texto estudado. A lei de Zipf é chamada também a 'lei das palavras de frequência elevada' ou 'princípio do menor esforço', lembrando assim sua relação com a lei de Pareto.

---

[33] Vilfrido Pareto (1848-1923), economista e sociólogo, nasceu em Paris, mas viveu a maior parte de sua vida na Itália. Universalmente conhecido pela lei que leva seu nome, também denominada regra do 80/20. Formalizou o enunciado da lei em 1875, quando descobriu, num estudo sobre a renda e riqueza, observou que uma pequena parcela da população, 20%, que concentrava a maior parte da riqueza, 80%. (Ver por exemplo: Wikipedia, 2008). A titulo de curiosidade, o gráfico que figura na referida página da Wikipedia, apresenta um traçado em todo semelhante ao de nosso *Gráfico 4.*

[34] Alfred James Lotka (1880-1949), matemático e físico americano, quando trabalhava no *Statistical Bureau* da *Metropolitan Life Insurance Company*, realizou um estudo em que relacionava os nomes dos autores citados no *Chemical Abstracts* entre 1907 e 1916, com o número de publicações e, em 1926, formulou a lei que leva seu nome, conhecida também como 'lei dos quadrados inversos' ou 'lei exponencial inversa'.

[35] George Kingsley Zipf (1902-1950) nasceu na Alemanha, filho de uma rica família. Estudou lingüística e filologia em Harvard, onde lecionou a maior parte de sua vida acadêmica. Autor da lei que leva seu nome, que analisa a freqüência/distribuição das palavras em um texto para determinar sua posição, em função de sua freqüência de ocorrência no texto. Por ocasião do centenário de seu nascimento o periódico eletrônico *Glottometrics,* dedicou, vários volumes a Zipf e a sua lei , especialmente os volumes 3, 4 e 5 (GLOTTOMETRICS, 2000). Para maiores, detalhes consultar Ronald Rousseau (2002) e Karin Hinkle (2006 - [?]), da *School of Information Science* (Portsmouth Polytechnic School, UK). Duas obras nucleares de Zipf: *Selective Studies and the Principle of Relative Frequency of Language* (1982) e *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology* (1949) são frequentemente citadas nas publicações sobre bibliometria. Ver, por exemplo Hinkle (*op.cit.*, 2006 [?])

A fórmula mais comum para a lei de Zipf é:

$$rf = C \qquad [4]$$

onde: "[...] *r* é a posição de uma palavra, *f* é a freqüência de ocorrência dessa palavra, e *C* uma constante que depende do texto analisado" (WYLLYS, 1981, p. 1; *apud* JAYROE, 2008). Em outras palavras, o número de ocorrências de uma palavra, em um determinado texto, é –aproximadamente – inversamente proporcional à posição que ocupa na relação de todas as palavras ordenadas conforme sua freqüência de ocorrência (POTTER, 1988; *apud* UTEXAS, [s.d.]). Para deixar bem claro como proceder em uma aplicação concreta da lei de Zipf, escolhemos um resumo de 3.100 palavras, reproduzido no *Anexo 1*, no final do presente Capítulo. Com a ajuda de um editor de texto e um gerenciador de planilhas, são realizadas as seguintes etapas:

Etapa 1. – Escolha do texto a analisar;
Etapa 2. – Extração e ordenação das palavras do texto;
Etapa 3. – Reunião das palavras por radicais, desinências e flexões, contagem das respectivas freqüências de ocorrência;
Etapa 4. – Ordenação dos dados para montagem do gráfico;
Etapa 5. – Montagem do gráfico.

As três primeiras etapas são auto-explicativas e os resultados de cada uma são mostrados no referido Anexo. Na Etapa 4, com os dados resultantes da Etapa 3, monta-se uma tabela que relaciona a posição das palavras ordenadas conforme sua freqüência de ocorrência (*'ranking´ r*) com a sua freqüência *f* (*Tabela 8*).

Tabela 8. – Relação entre as palavras ordenadas e acordo com sua freqüência e a freqüência das mesmas.

EIXO X = *'RANKING´* **r**	EIXO Y = FREQÜÊNCIA **f**
1	142
2	62
3	33
4	18
5	15
6	13
7	11
8	9
9	8
10	7
11	6
12	1

Observe-se que, em que pese o pequeno número de palavras do texto analisado, o produto dos valores de cada linha se mantém relativamente próximo: 1x142=142; 2x62=124; 3x33=99; 4x18=**72**; 5x15=**75**; 6x13=**78**; 7x11=**77**; 8x9=**72**; 9X8=**72**; 10x7=**70**; 11x6= 66; 12x1=12, especialmente na zona central – com valores da constante *C* entre 70 e 78 – o que confirma as observações de outros autores sobre a maior dispersão dos valores nos extremos.

Finalmente, a Etapa 5, com a ajuda de um processador de planilhas (em nosso experimento foi utilizado o *Excel*), se obtém o gráfico abaixo, totalmente conforma com as previsões (ver *Gráfico 5*).

A lei de Zipf funciona bem com as palavras de freqüência elevada, mas não tão bem com as de baixa freqüência, de onde resultam numerosas modificações propostas por diversos autores. Uma variante da fórmula proposta pelo próprio Zipf sugere que a freqüência de ocorrência de um determinado evento (P), considerada como uma função da posição – ou *'rank´* – (i), quando essa posição vem determinada pela freqüência de ocorrência acima, é uma função da lei exponencial inversa

$$P_i \sim 1/i^a \qquad [5]$$

com o expoente próximo da unidade (1).

Gráfico 5 – Gráfico do experimento com a lei de Zipf.

Alguns autores preferem a representação gráfica da Lei de Zipf utilizando escala logarítmica nos dos eixos (em nosso caso foi utilizada uma escala linear em ambos os eixos). Nesse caso, a curva se transforma numa linha reta, como pode-se ver no *Gráfico 6,* abaixo, o que não muda nada a significação dos resultados.

Gráfico 6. – Representação gráfica da lei de Zipf em escala logaritmica.

Lembremos, para encerrar nossas considerações sobre as três grandes leis da bibliometria, que a lei de Zipf, embora sua aplicação original na análise de palavras em textos, não é, de modo algum, um 'jogo de palavras'. Ela é de extrema utilidade na construção de tesauros e léxicos, processos de indexação e recuperação da informação, e ganha a partir do início do século um novo interesse em aplicações em ambientes Web, e em processos de mineração de textos.[36]

5. METRIAS DA INFORMAÇÃO E CIÊNCIA DA INFORMAÇÃO EM AMBIENTES WEB

Nesta Seção, limitar-nos-emos a citar alguns trabalhos que orientarão os(as) leitores(as) no estudo das mudanças geradas pelo advento da Internet. Uma abordagem com maiores detalhes precisaria de um espaço ainda maior do que o resto do Capítulo.

[36] Para ampliar e aprofundar os fundamentos das idéias expostas, é recomendada a consulta e estudo da obra de Ravichandra Rao (op. cit., 1986, p. 74, 182, 183, 221, para Lotka, e p, 74, 180, 228, para Zipf).

A partir do início da década de oitenta do século passado, se observa, no mundo inteiro, um renovado e crescente interesse pelas técnicas bibliométricas/infométricas em ambiente Web, agora com novas denominações, tais como *Webmetrics, Webometrics, Cybermetrics*, etc., que se incorporam ao vernáculo como Webmetria, Webometria, Cibermetria, e por aí vai. Tendência essa que, no Brasil, se observa também a partir dos registros da base ABCDM (Arquivologia, Biblioteconomia, Ciência da informação, Documentação e Museologia) à qual foi feita referência anteriormente (ver *Gráfico 3* e *Tabela 6*, na Seção 3) e sobre a qual voltaremos mais adiante (ver Nota 41, na Seção 3.2)[37].

A Webmetria pode-se definir como "o uso das técnicas bibliométricas para estudar as relações entre diferentes 'sites' em ambiente Web" (UTEXAS, [s.d.]). Então, nada mudou? Aparentemente, nada; mas, de fato, muita coisa.

Assim, todas as técnicas infométricas/bibliométricas/cientométricas, tratadas acima, para analisar as relações entre novas fontes e novos itens ('sites', portais, repositórios e/ou páginas, documentos digitalizados, etc.), ampliam cada vez mais seu campo de aplicação.

Os links e hiperlinks tornam-se equivalentes às citações entre publicações, artigos e revistas, e novas possibilidades de mapeamento surgem para descobrir os 'sites'/portais/fontes mais produtivos e/ou mais influentes. Muito têm contribuído, também, os motores de busca atualmente disponíveis, cada dia mais potentes, para alargar o universo de nossas fontes de informação. A revolução nos hábitos dos governos, das corporações, das universidades, dos centros de pesquisa, para orientar suas políticas em prol da inovação e do desenvolvimento, não tem feito senão começar.

Convém alertar sobre a necessidade de completar a leitura das fontes, que sugerimos a continuação, com novas pesquisas e novas leituras críticas, pois as publicações sobre temas que hoje são 'tendência' podem, em pouco tempo, se tornar obsoletas, varridas por novos nichos de inovação ainda impensados. A intenção de nosso convite à leitura dessas fontes é, simplesmente, ajudar a dar os primeiros passos – ou a reforçar a intenção de continuar trilhando o caminho, na direção do bom uso dos métodos quantitativos/estatísticos em que se baseiam as metrias da informação.

Sem o auxílio desses métodos, torna-se muito difícil – por não dizer impossível – fundamentar, planejar, gerenciar e tornar possíveis nossas escolhas e decisões, alicerçadas em dados e informações cofiáveis, num mundo cada dia mais complexo e mais competitivo.

Eis uma pequena lista de leituras sugeridas, a completar pelos leitores: B.C. Brookes (1980); Paul Wouters (1988); I.F. Aguillo (2001, 2003, 2004); I.F. Aguillo *et al.*(2005a, 2005b, 2005c) ; Nadia A.P. Vanti (2002); L. Björneborn e P. Ingwersen (2004), já citado anteriormente; Mike Thelwall (2008); M. Thelwall e L. Vaughan (2004); M. Thelwall e I.F. Aguillo (2003); M. Thelwall *et. al*. (2005); Maricela López (2007); Prashant Goswani *et. al*. (2008); Judit Bar-Ilan (2009); V.K. Vaishnavi *et al*. (2009); F. Janssens *et al*. (2009).

6. APLICAÇÕES RECENTES QUE SINALIZAM TENDÊNCIAS

Nesta Seção serão abordados alguns exemplos recentes de aplicações infométricas, resultado das pesquisas concluídas ou em curso, dos autores do presente Capítulo, ou que se coadunam com as tendências observadas nos estudos publicados por autores de outros países, bem como por alguns autores brasileiros.

A Seção será dividida em três sub-seções que presentam alguns resultados recentes de pesquisas dos autores do presente Capitulo. São elas: a primeira, *Aplicações bibliométricas na gestão e desenvolvimento de coleções em áreas especializadas* ; a segunda, *Aplicações bibliométricas / infométricas usando estatística descritiva,* trata da produção *versus* colaboração científica, e do gênero das autorias científicas no Brasil; a terceira, *Outras aplicações infométricas,* trata, sucessivamente, da identificação automática de grupos e linhas de pesquisa*,* da identificação automática de clusters temáticos, ontologias e mapas de tópicos, da análise de citações, da mineração de textos em bases de periódicos científicos especializados, e de algumas aplicações inferenciais.

6.1 Aplicações bibliométricas na gestão e desenvolvimento de coleções em áreas especializadas

Os primeiros trabalhos de Robredo sobre aplicações bibliométricas e seus desdobramentos foram publicados na França, na década de 70 do passado século, quando era responsável pelo Serviço de Documentação do *Institut du Verre*, órgão de apoio da Federação das Indústrias Francesas de Vidro.

[37] A aparente contradição entre os dados apresentados no *Gráfico* 3, na Seção 3, com os mostrados por Urbizagástegui Alvarado (1984), para o período 1972-1983, reproduzidos no *Gráfico 1*, da Seção 2.5.2, se devem, provavelmente, à conjunção de dois fatores: i) os dados do Gráfico 1 cobrem teses e dissertações até 1983, orientadas pelo IBICT, enquanto o Gráfico 3 cobre artigos de periódicos da década de 1980 completa, e ii) as teses e dissertações aprovadas até 1983, começaram a ser elaboradas, respectivamente, pelo menos, dois ou um ano antes, ou seja num período (1980 - 1982), em que o impacto da Internet no Brasil não era muito significativo, em quanto os periódicos da base ABCDM cobrem a década de 80 completa.

Essas pesquisas abrangem os seguintes focos temáticos:

1) Estudo da dispersão da literatura (ROBREDO, 1973a);

2) Estudos sobre a influência da indexação sobre a qualidade da recuperação da informação (ROBREDO, 1973b), (ROBREDO; BRISNER, 1973c), e aplicações inferenciais (ROBREDO, 1970).[38]

O autor teve a oportunidade de dar prosseguimento às suas pesquisas bibliométricas/infométricas, no Brasil, nos anos seguintes, em outra área de aplicação, no quadro de suas atividades no processo de implantação do Sistema Nacional de Informação e Documentação Agrícola (SNIDA), no Ministério da Agricultura, Pecuária e Abastecimento (MAPA) e da fundação da Biblioteca Nacional de Agricultura (BINAGRI), aos quais foi feita menção anteriormente (MAPA, 2009), em parceria com o Programa das Nações Unidas para o Desenvolvimento (PNUD) e da Organização das Nações para Agricultura e Alimentação (FAO).[39]

Algumas linhas de estudo e aplicação mais destacadas são: *1)* Elaboração da lista básica de periódicos agrícolas (CHASTINET, 1975), e avaliação da produtividade e do nível técnico-científico dos periódicos brasileiros (ROBREDO *et al.*, 1974), (CHASTINET *et al*. 1975); *2)* Elaboração do thesaurus agrícola nacional compatível com o thesaurus do Sistema AGRIS *Agrovoc* (ROBREDO *et al*. 1975a; 1975b); *3)* Implementação de um serviço de disseminação seletiva da informação usando a base AGRIS e a base nacional, para pesquisadores brasileiros (ROBREDO *et al*., 1976), (CHASTINET *et al*. 1978).

Complementando a citação de Pinheiro (2005) sobre a criação da BINAGRI, detalhada na Nota 19 da Seção 2.5 deste Capítulo, pode-se sugerir a consulta a alguns documentos relacionados com as técnicas bibliométricas, publicados no período 1974-1979.

6.2 Aplicações bibliométricas e infométricas usando estatística descritiva

São apresentadas a seguir duas aplicações bibliométricas, usando estatística descritiva, relacionadas a temas cujo interesse está crescendo no Brasil nas áreas de informação: produção científica, colaboração científica e a participação das mulheres na comunidade científica[40]. Tais aplicações têm baixa complexidade estatística e computacional, usando uma base de dados bibliográfica, como fonte principal, e um programa de planilha eletrônica para elaboração de tabelas e gráficos.

Antes de apresentar as informações faremos uma breve revisão do contexto científico, especialmente da comunidade científica e dos processos de comunicação científica, não só para relevar a importância do tema, como para facilitar a compreensão da tabelas e gráficos.

No processo de pesquisa científica a comunidade científica não trabalha de forma homogênea e uniforme e seu padrão de comportamento está relacionado com uma divisão em grupos de especialistas por área (KUHN, 1979, p.312-313). A especificidade de cada área leva à adoção de maneiras diferentes de fazer e provavelmente de comunicar pesquisa (MEADOWS, 1999, p.39). Além disso, a estrutura socioeconômica de cada país influencia os indicadores científicos (PRICE, 1976, p.xi).

A comunicação científica, através dos seus vários canais "*situa-se no próprio coração da ciência*", sendo "*tão vital quanto a própria pesquisa*", pois para que os resultados de uma pesquisa sejam analisados e aceitos pelos pares, exige que seja necessariamente comunicada. Assim, "*a comunicação eficiente e eficaz constitui parte essencial do processo de investigação científica*" (MEADOWS, 1999, p.vii).

Para Ziman (1968), "*a literatura de uma área é tão importante para essa área quanto os próprios trabalhos de pesquisa*" (MUELLER, CAMPELLO; DIAS, 1996, p.1), consistindo de um dos três indícios de maturidade de uma área do saber e de sua institucionalização.

---

[38] Esses trabalhos contaram com o apoio da *Union Scientifique Continentale du Verre* – USCV (VERRE Online, 2006), que agrupava a maioria dos países da Europa ocidental, e da *International Commission on Glass* – ICG (ICG, 2009), a qual reunia as federações nacionais e associações científicas de mais de trinta países, e de cujo Sub-Comitê A12 "*Documentation and Information Retrieval*" coube ao primeiro autor do presente Capítulo a honra de ser o primeiro presidente. Merecem ser citados, pela sua repercussão no desenvolvimento da então denominada "*documentation scientifique e technique*" alguns trabalhos colaborativos sob os auspícios dessas entidades, tais como: *i)* o *Pool d'Abstracts Verriers* | *Pool of Glass Abstracts* (IPGA, s.d.); *ii)* a Extensão da Classificação Decimal Universal (CDU) no domínio do Vidro (FID, 1984); e *iii)* o *International Dictionary of Glass-Making*, em onze línguas (ICG, 1965, 1967, 1968, 1974, 1983, 1992); ver. também, por exemplo, Murray (1968) e Soergel (1997).

[39] Trata-se do Projeto PNUD/FAO/BRA/72/020, cujos registros em microficha dos principais documentos técnicos e publicações gerados durante a primeira fase de implantação do Projeto, no período 1974-1979, encontram-se disponíveis na Biblioteca da FAO, nos *FAO Online Catalogues* [Entrada: Robredo, J] (FAO, 1972 -->).

[40] O crescimento do interesse nestes temas pode ser observado pela freqüência de trabalhos apresentados em eventos, como ENANCIB, e/ou publicados em periódicos científicos brasileiros, nas áreas de informação, nos últimos anos, ou ainda por meio de buscas em bases de dados bibliográficos (como a base ABCDM do Departamento de Ciência da Informação e Documentação da Universidade de Brasília, à qual será feita referência mais adiante).

Um dos fenômenos relacionados com a comunidade científica que tem despertado interesse nos últimos anos é a pesquisa em colaboração que cresce em todas as áreas do conhecimento. A pesquisa em colaboração é incentivada atualmente por governos e agências de fomento, instituições de pesquisa e universidades, e prestigiada por editores, avaliadores e grupos científicos. Katz e Martin (1997, p. 1) atribuem esse prestígio à crença difundida entre comunidades científicas e responsáveis por políticas científicas de que "*colaboração em pesquisa é uma coisa 'boa' para a ciência e que deve ser encorajada*". Acredita-se que a colaboração reduz custos e aumenta os benefícios da pesquisa, inclusive a qualidade dos resultados, embora sejam levantadas questões sobre a veracidade dessa crença, especialmente em relação aos custos e benefícios. Dentre os diversos tipos de colaboração, a autoria múltipla de artigos apresenta maior facilidade de mensuração e tem sido o indicador mais freqüentemente usado, a ponto de, muitas vezes, o termo colaboração ser usado como sinônimo de autoria múltipla.

Este é um estudo bibliométrico e, como tal, pode-se considerar como "*um poderoso conjunto de métodos e medidas para o estudo da estrutura e do processo de comunicação acadêmica*" (BORGMAN; FURNER, 2002). Ainda, associado a bases de dados, torna-se um importante instrumento de análise para "*rastrear as tendências e o desenvolvimento da sociedade*", das disciplinas científicas e das áreas de produção e consumo" (WORMELL, 1998, p.210). O objetivo deste estudo é mostrar algumas das características específicas de parte da literatura científica das áreas de informação – Arquivologia, Biblioteconomia, Ciência da Informação, Documentação e Museologia (ABCDM)[41] – no Brasil, especialmente questões relacionadas com: *a)* a produção de artigos de periódicos, um importante indicador do nível de atividade científica de uma área, bem como dissertações e teses, *b)* a colaboração entre pesquisadores, um dos fatores que contribui para o aumento contínuo do número de artigos científicos publicados (PRICE, 1976, p.55); e *c)* o gênero das autorias, considerada uma variável sociológica importante e com impacto equivalente ao de classe social (SILVA, 1995, p.34), tendo sido lembrada por Engels, Marx, Freud, Darwin (FERREIRA, 2003, p.192) e Bourdieu (SILVA, 1995, p.33).

A fonte principal dos dados deste estudo é a base ABCDM, armazenada em CDS/ISIS for Windows (WinISIS), que contém cerca de cinco mil referências bibliográficas completas de artigos publicados em mais de vinte títulos de periódicos científicos brasileiros e portugueses das áreas de informação. Foram usadas neste trabalho 3.843 referências de artigos publicados entre 1980 e 2007 em periódicos científicos brasileiros das áreas de informação. Os dados foram extraídos da base, em formato padrão WinISIS, para arquivos do tipo texto (.txt), por meio do comando de impressão, e foram posteriormente inseridos no MS-Excel onde foram produzidas as tabelas e gráficos para análise. Os registros são em sua maior parte de artigos científicos, embora existam registros de outra natureza como relatos de experiências, trabalhos de congressos entre outros, que também são publicados em periódicos científicos[42].

É importante observar que os dados relativos a um determinado ano, utilizados neste estudo como referência da publicação dos fascículos, foram obtidos da data de referência da publicação de cada fascículo, e não correspondem, necessariamente, aos anos de sua efetiva publicação, em razão de eventuais atrasos na publicação dos fascículos. Uma forma de minimizar tais distorções é considerar a média dos dois últimos anos, representadas nos gráficos por linhas de tendência. Além disso, quando essas datas se apresentam como intervalo, considera-se apenas o último ano, i.e. em 1999-2000 considera-se apenas 2000.

Os dados relacionados com dissertações e teses foram obtidos por meio de consulta às listas de autores e aos próprios documentos do acervo do Programa de Pós-Graduação em Ciência da Informação (PPGCInf) do Departamento de Ciência da Informação e Documentação, da Universidade de Brasília.

6.2.1 Produção versus colaboração científica

Em relação à produção de artigos de periódicos nas áreas de informação, podemos observar no *Gráfico 9* que os valores variam entre 68 em 1980 e 273 em 2007. Podemos observar ainda que a produção não cresce de maneira constante, tendo passado por uma queda na primeira metade dos anos 1990, seguida de um crescimento acentuado a partir da segunda metade da mesma década.

---

[41] Foram considerados no estudo os artigos de fascículos publicados até o final de 2008 com data de referência entre 1980 e 2007, entretanto, nem todos os fascículos de 2007 tinham sido publicados. Os periódicos da ABCDM são: "*Acervo: Revista do Arquivo Nacional*"; "*Arquivística.net*"; "*Arquivo & Administração*"; "*Biblos: Revista do Departamento de Biblioteconomia e História*"; *"Cadernos de Biblioteconomia*"; *"Ciência da Informação*"; *"Ciências em Museus*"; *"DatagramaZero*"; "*Em Questão: Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS*"; "*Encontros Bibli*"; "*Estudos Históricos*"; *"Informação & Informação*"; *"Informação & Sociedade: Estudos*"; "*Informare: Cadernos do Programa de Pós-Graduação em Ciência da Informação (UFRJ/IBICT)* "; *"Perspectivas em Ciência da Informação*"; "*Ponto de Acesso: Revista do Instituto de Ciência da Informação da UFBA*"; "*Revista ACB: Biblioteconomia em Santa Catarina*"; "*Revista Brasileira de Biblioteconomia e Documentação*"; *"Revista Brasileira de Museus e Museologia (MUSAS)*"; "*Revista da Escola de Biblioteconomia da UFMG*"; "*Revista de Biblioteconomia & Comunicação*"; "*Revista de Biblioteconomia de Brasília*"; "*Revista Digital de Biblioteconomia e Ciência da Informação*"; "*Revista do Patrimônio Histórico e Artístico Nacional*"; "*Revista Eletrônica Jovem Museologia*"; "*Revista Museu*"; "*Transinformação".*

[42] Estima-se, com base na amostra de pesquisa de doutoramento em andamento, que cerca de 30% dos registros sejam de comunicações de congressos, relatos de experiências, entre outros tipos de literatura.

Gráfico 9. – Produção de artigos de periódicos brasileiros das áreas de informação (1980-2007)

Observação: A linha contínua representa o valor médio a cada dois anos.

Gráfico 10. – Produção de artigos de periódicos brasileiros das áreas de informação por tipo de autoria (1980-2007)

Observação: As linhas contínuas representam o valor médio a cada dois anos.

Uma das explicações para o crescimento acentuado no número de artigos científicos a partir de meados dos anos 1990 é a publicação em colaboração, como observado também por Meneghini (1996) em relação à produção científica brasileira visível no ISI entre 1981 e 1993.

A explicação faz sentido quando observamos o *Gráfico 10*, que mostra a produção dos artigos de periódicos das áreas de informação no Brasil por tipo de autoria entre 1980 e 2007. Podemos observar que até 1995 cerca de 80% dos artigos eram escritos em autoria única (linha superior), enquanto apenas cerca de 20% eram elaborados por mais de um autor (linha inferior). A partir de 1996 a proporção de artigos em autoria múltipla passa a aumentar de forma sistemática e se aproxima da metade dos artigos em 2007.

Ao observarmos as características da produção de dissertações e teses do PPGCInf/UnB, documentos científicos tão importantes quanto os artigos, entre 1980 e 2007, podemos notar alguns pontos em comum com a produção de artigos. Primeiramente, a produção de dissertações e teses no CID sofre uma grande variação na produção com queda em meados dos anos 1990 seguida de forte crescimento até o final da mesma década (ver *Gráfico 11*), quando então há nova queda, permanecendo a produção em valores médios próximos a 15 documentos.

A partir de 2006 a produção volta a subir chegando a 30 documentos em 2006 e 25 em 2007. No caso das teses, a explicação para esse aumento não pode ser a mesma que no caso dos artigos, sendo mais provável que o incremento da produção seja resultado das políticas de fomento, e da criação e consolidação dos cursos de doutoramento nas áreas de informação no Brasil na década de 1990, fatores que também devem interferir na produção de artigos.

6.2.2 O gênero das autorias científicas no Brasil

Antes de mostrar a evolução da produção científica brasileira nas áreas de informação por gênero, cuja importância já foi explicada anteriormente, convém esclarecer que o conceito de autoria (relação autor/artigo) é diferente dos conceitos de autor (pessoa) e artigo (obra), ou seja, uma autoria pode ser entendida como 'esforço' de um autor (pessoa) para escrever um artigo (obra) sozinho ou em conjunto com outro autor (pessoa). O entendimento do conceito de autoria como (relação) entre o autor (pessoa) e o artigo (obra) é importante e pode ser sintetizado da seguinte maneira: quando um artigo (Art1) é escrito por um autor (Aut1) em conjunto com outro autor (Aut2), e outro artigo (Art2) é escrito pelos mesmos autores, temos então dois artigos (Art1 e Art2), dois autores (Aut1 e Aut2), e quatro autorias: as relações Aut1-Art1, Aut2-Art1, Aut1-Art2 e Aut2-Art2.

Os gráficos anteriores usaram como unidade os artigos, dissertações e teses (obras), enquanto que os próximos gráficos usarão como unidade as autorias (esforços de um autor para produzir artigos).

Gráfico 11. – Produção de dissertações e teses do CID/UnB (1980-2007)

Observação: As linhas contínuas representam o valor médio a cada dois anos.

Gráfico 12.– Percentual de autorias por gênero nas áreas de informação no Brasil (1980-2007)

Observação: As linhas contínuas representam a média de dois anos.

O *Gráfico 12* mostra a participação de cada gênero nas autorias de artigos. Podemos notar que até o início da década de 2000, o percentual médio de homens (linha inferior) permanece próximo a 30% enquanto o de mulheres (linha superior) fica próximo dos 70%. Na década de 2000 há uma pequena tendência de aproximação entre os percentuais que se aproximam dos índices de 60% de autorias femininas e 40% de masculinas. Tais percentuais não surpreendem considerando que tradicionalmente nas áreas de informação as mulheres são em maior número, especialmente no setor profissional, e também no acadêmico e de pesquisa, como relatam Hayashi *et alii* (2007, p.178) na pós-graduação em Ciência da Informação da UFSCar (77%) em 2006, embora seja bem menor no Programa de Pós-Graduação do CID/UnB (43%) em 2009.

Gráfico 13 – Percentual de autorias únicas por gênero nas áreas de informação no Brasil (1980-2007)

Observação: As linhas contínuas representam a média de dois anos.

Gráfico 14. – Percentual de autorias múltiplas por gênero nas áreas de informação no Brasil (1980-2007)

Observação: As linhas contínuas representam a média de dois anos.

Ao separarmos os dados de gênero por tipo de autoria, únicas e múltiplas, obtemos uma evolução diferente nos percentuais de cada gênero (ver *Gráficos 13* e *14*).O *Gráfico 13* mostra a participação de cada gênero nos artigos em autoria única. Podemos notar que até 1996/1997 os valores médios de autorias femininas ficam na faixa de 60% a 70% e a partir daí passam a assumir percentuais médios menores do que 60%, chegando a 54% em 2007 contra 46% de autorias únicas masculinas.

Gráfico 15.– Percentual de autorias múltiplas por gênero nas áreas de informação no Brasil (1980-2007): autor principal

Observação: As linhas contínuas representam a média de dois anos.

Gráfico 16. – Percentual de autorias de dissertações e teses do CID/UnB por gênero (1980-2007)

Observação: As linhas contínuas representam a média de dois anos.

Já o *Gráfico 14* mostra a participação de cada gênero nos artigos em autoria múltipla, onde podemos notar que até 2000 os valores médios de autorias femininas ficam na faixa de 70% a 80% e a partir daí passam a assumir percentuais médios menores do que 70%, chegando a valores próximos a 60%, contra 40% de autorias múltiplas masculinas, em 2007.

Ao separarmos os dados de autorias múltiplas considerando apenas autores principais (campo autor principal), podemos observar no *Gráfico 15* que na década de 1980 os percentuais médios masculinos eram próximos de 20%, até o início da década de 2000, quando percentuais superiores 30% passaram a ser observados, chegando até 40% em 2007.

O comportamento das autorias múltiplas principais não se mostra, portanto, muito diferente do conjunto das autorias múltiplas principais e secundárias (*Gráfico 14*), exceto em 2006-2007 que tem índices ligeiramente superiores. Isto parece indicar que nas áreas de informação as mulheres não são discriminadas em relação à posição da autoria (principal ou secundária).

Podemos concluir que nas autorias únicas há um maior equilíbrio entre os gêneros, enquanto nas autorias múltiplas a proporção de autorias de mulheres é mais acentuada. Observa-se em todos os gráficos de gênero que a partir de 2003 os percentuais de autorias masculinas e femininas convergem de forma constante até o final da série (2007). Uma explicação para este fenômeno poderia ser o aumento do número de autores homens nas áreas de informação. Caso esteja havendo um aumento de homens nas áreas de informação, este fenômeno parece ter maior impacto nos índices de autorias múltiplas do que nos índices de autorias únicas.

Podemos notar indícios deste aumento ao analisarmos o gênero dos autores das teses e dissertações do PPGCInf/CID/UNB entre 1980 e 2007. O *Gráfico 16* mostra que houve um aumento do percentual de autores masculinos nas teses e dissertações a partir de 1996. Até 1995 o percentual médio de autoras é superior a 80%, a partir daí os índices vão se aproximando do equilíbrio. Embora os dados sejam apenas de um curso, é um indicador de que a presença de homens nos níveis de pós-graduação nas áreas de informação pode estar aumentando.

As aplicações mostradas no item 6.1 usam técnicas bem simples de conversão de dados bibliográficos para tabelas e gráficos em programas de planilhas eletrônicas, mostrando tendências em relação à produção de artigos, dissertações e teses nas áreas de informação no Brasil, bem como características relacionadas ao tipo de autoria e gênero. Tais aplicações são apenas alguns exemplos de inúmeras possibilidades de associações como afiliações das autorias (por instituição, estado, país, região, continente etc.), temas dos artigos, relações entre autores (orientações, membros de grupos de pesquisa e comissões, professor-aluno etc.). Cabe ainda salientar que além da coleta e organização dos dados, e elaboração de tabelas e gráficos, uma importante fase no uso da bibliometria é análise dos elementos e sua correta interpretação. Tal análise exige não só conhecimento do tipo de documento analisado e dos processos a eles relacionados, mas também conhecimento de todo o contexto científico e de suas especificidades relacionadas com a área temática e o país.

Tal abordagem é defendida por Araújo (2006), que aponta o uso cada vez maior do "instrumental das ciências sociais" em conjunto com a bibliometria. Dentro dessa linha, é oportuno enfatizar a importância de sistematizar a produção de indicadores quantitativos, cujas características estão relacioanadas em Trzesniak (1998).

6.3 Outras aplicações infométricas

É conveniente esclarecer que se trata de aplicações que de muito extrapolam os limites das bibliotecas tradicionais, e mesmo das virtuais, as quais se projetam cada vez mais em escala planetária. Isso não quer dizer que o que segue não interessaria aos profissionais bibliotecários, arquivistas, documentalistas, e da ciência da informação, em geral. Muito pelo contrário, e isso por duas razões, pelo menos: *1)* em que pese as deficiências observadas em alguns currículos atuais, no que se refere ao uso de métodos que permitem quantificar e relativizar valores e características, são esses os profissionais que formam ainda um dos grupos mais preparados para orientar e fundamentar escolhas que levem a vencer na concorrência, e a abrir o caminho da qualidade e da inovação; e *2)* porque são, também, esses profissionais os que se encontram, provavelmente, mais capacitados a representar e organizar dados, informações e conhecimentos de forma a, com o auxílio das tecnologias mais avançadas da informação e da comunicação, transformar essas informações e esses conhecimentos em conhecimentos mais amplos e profundos, e em novos saberes.

Fala-se muito, por exemplo, de 'valor agregado' a algo, a algum produto, a alguma coisa tangível ou intangível, mas quanto, que valor? Ou, como medir uma 'vantagem competitiva'? Como saber (ou medir) o que é maior, o que é melhor. Sem esse conhecimento, em que se baseiam nossas escolhas, nossas decisões? Como definir políticas, como planejar, como gerenciar? Qual é a massa crítica de recursos humanos, financeiros, técnicos, etc., para que nossos projetos (ou os dos nossos políticos) tenham alguma chance de dar certo?

Trata-se, de fato, de estabelecer ou encontrar indicadores quantitativos, baseados em propriedades ou características mensuráveis, sem os quais mal poderemos dar esse salto qualitativo e quantitativo que tanto desejamos para nós, para nossa sociedade, para nosso país.

Os métodos e técnicas infométricos – ou, melhor, as metrias da informação, como se fala cada vez mais, de uma forma mais genérica – muito podem contribuir para esses fins. Os exemplos a seguir mostram alguns caminhos 'que funcionam', facilmente generalizáveis e extensíveis a um sem-fim de aplicações nas mais diversas áreas.

6.3.1 Identificação automática de grupos e linhas de pesquisa

A realização das pesquisas, cujos resultados serão apresentados na presente seção e na seguinte, requereu a disponibilidade de uma base de dados suficientemente volumosa para permitir a aplicação dos cálculos estatísticos que nos levariam a identificar os grupos de pesquisa atuantes na Universidade de Brasília e as linhas em que trabalham. Assim, dedicaremos umas linhas a descrever a estrutura e conteúdo da base antes de abordar a discussão dos fundamentos e do *modus operandi* que nos permitiriam alcançar os objetivos propostos. Entretanto, antes de abordar o tema, convém esclarecer que nossa intenção não é apresentar em detalhe quais são os grupos mais atuantes, em que áreas trabalham, ou quanto e onde publicam, pois isso não é senão uma decorrência dos resultados obtidos num certo momento, que pouco ou nada significariam alguns meses mais tarde, fora do contexto espaço-temporal em que foram obtidos.

O que se pretende é mostrar como usar ferramentas simples, de comprovada eficiência e confiabilidade, para obter um retrato de uma determinada situação, num instante dado, bem como para monitorar sua evolução ao longo de um certo período, ou para obter indicadores que ajudem a planejar e implementar políticas e privilegiar o desenvolvimento ou o reforço de novos nichos de inovação. Mas, antes de chegarmos lá, vejamos como organizar dados e informações para montar a base.

É importante, também, esclarecer que, na síntese dos resultados obtidos no desenvolvimento dos projetos de pesquisa aos quais os resultados ora apresentados se referem foram introduzidos alguns aprimoramentos na ferramenta desenvolvida progressivamente no decorrer dessa pesquisa[43], no intuito de corrigir algumas inconsistências observadas nos trabalhos publicados e/ou nas comunicações apresentadas. Assim, os leitores que porventura venham a consultar algumas das fontes citadas poderão observar algumas diferenças, entre os exemplos aqui expostos – especialmente no visual das telas reproduzidas – se comparadas às respectivas telas anteriormente publicadas.

Para implementar a base, os dados da produção científica, técnica e artística da Universidade de Brasília, de 1999 a 2004, recebidos da Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES) e do Decanato de Pesquisa e Pós-Graduação da UnB, foram transformados em registros de bases de dados através da aplicação *MX*[44], como mostram a *Tabela 9* e a *Figura 4.*

Tabela 9. – Estrutura dos registros da base.

ETIQUETAS	*NOME DO CAMPO*
035	Identificação do registro
100	Autor(es)
245	Título / Fonte / Notas

```
0040900000000073000450003500160000099002000016100014400036245015500180-CDS    0
2(07)045-Produção Científica-SILVA, MARINA; NOVAES, WASHINGTON; BUARQUE, CRISTOV
AM RICARDO CAVALCANTI; CAMARGO, ASPASIA; MORHY, LAURO; CASTRO, EDNA RAMOS; BURSZ
TYN, MARCEL.- A questão amazônica: em busca de um projeto. In: MORHY, LAURO (ed.
). Brasile em questão: a universidade e a eleição presidencial. Brasília DF: UnB
, 2002.-"
0034600000000073000450003500160000099002000016100010100036245013500137-IB CEL9
5(04)013-Produção Científica-NITTI, GIAMPAOLO; ORRÙ, STEFANIA; BLOCH JÚNIOR, CAR
LOS; MORHY, LAURO; MARINO, GENNARO; PUCCI, PIERO.-Amino acido sequence and disul
phide-brigde pattern of three P-Thionins from sorghum bicolor.. Eur. J. Biochem,
v.288, 1995, p.250-256.-"
```

Figura 4. Estrutura de dois registros no formato ISO 2709 (Padrão UNISIST-CDS/ ISIS).

O sistema é composto por um banco de dados alimentado por uma aplicação escrita em linguagem Java. Os registros aceitos pela aplicação estão em formato ISO 2709 (ISO, 2008), mostrado acima. A base, com cerca de 50.000 registros, referentes à produção científica e técnica da Universidade de Brasília, no período 1994-2003, foi descrita anteriormente (ROBREDO; OLIVEIRA, 2003).

[43] Os trabalhos aos quais será feita referência decorrem, em grande parte, do apoio do CNPq através dos seguintes projetos de pesquisa: *1) Aplicação infométrica informatizada para acompanhamento da evolução da pesquisa científica e disponibilização via internet das bases de dados demonstrativas da produção científica e técnica da UnB* (1984-2002); *2) Prorrogado com o mesmo nome* (2003-2006); *3) Técnicas infométricas como ponto de partida da organização e gestão do conhecimento com base em ontologias* (2006-2010) – Em andamento.

[44] O *MX* é um utilitário em linguagem C, que trabalha com bases de dados padrão CDS/ISIS (UNESCO, 1993). Desenvolvido e distribuído gratuitamente pela BIREME – Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde e membro da Organização Pan-Americana da Saúde (OPS) e da Organização Mundial da Saúde (OMS). O MX opera em todas as plataformas realizando a maioria das funções do CDS/ISIS Windows (UNESCO, 1997), incluindo a geração de bases de dados a partir de arquivos de texto, conversão de padrões, importação e exportação de arquivos ISO 2709 (ISO, 2008), e controle de qualidade de dados. A interface de busca foi desenvolvida em WWWISIS – ou WXIS –, igualmente distribuído pela BIREME (2002). Esse utilitário age como um servidor de bases de dados CDS/ISIS em ambiente cliente/servidor Web ou Intranet. Utiliza uma linguagem de *script* baseada em XML, a 'IsisScript'. Os registros de uma base de dados CDS/ISIS podem ser convertidos em documentos HTML dinamicamente, ou seja, os resultados de uma pesquisa ou parte de um arquivo mestre podem ser recuperados e enviados ao cliente WWW. A saída em HTML é gerada via linguagem de formato, o que permite desenvolver uma interface de pesquisa totalmente gráfica em ambiente Web ou Intranet. O WWWISIS trabalha também em várias plataformas, como LINUX e UNIX, em diferentes CPUs, e MS-DOS para Windows (95, 98, 2000) e Windows NT/XP.

Figura 5 .- Arquitetura da aplicação InfoCluster atual.

Figura 6.– Nova página de abertura do InfoCluster.

Figura 7. – Página de abertura do InfoClusterAuthors

Figura 8. – Dois autores de sobrenome 'Mamiya'. (Clicando nos ícones adequados se abrem novas janelas que mostram os co-autores e/ou as publicações)

Figura 9. – Coautores de Mamiya, Edgar Nobuo.

Na *Figura 5*, encontra-se representada a arquitetura do sistema atual, que representa um aprimoramento significativo, em relação às versões anteriores (ROBREDO; CANTANHEDE, 2006). Depois de carregados, os registros podem ser acessados pela aplicação que faz a carga (*InfoCluster,* árvores de autores e edição dos registros). Entretanto, por ser mais leve e acessível, foram construídas interfaces web em *php* que exibem os *clusters* em páginas HTML. As interfaces são o *InfoClusterAuthor*, exclusiva para clusters de autores e o *InfoClusterTerm* para os termos (nomes) presentes nos campos de autoria e nos títulos das publicações.

A página de abertura do *InfoCluster* apresenta-se com um novo visual (*Figura 6*), assim como o a página de abertura do *InfoClusterAuthors* (*Figura 7*). Quando se insere um nome na janela superior da tela, após escolher a opção *InfoClusterAuthor* na tela de abertura, por exemplo 'mamiya', temos o resultado mostrado na *Figura 8*, onde observamos que existem duas pessoas de igual sobrenome, embora apresentadas com grafias diferentes.

Ao clicar na coluna 'Cluster' na linha de 'Mamiya, Edgar Nobuo' obtemos uma relação dos autores que publicaram com ele – com os nomes grafados diferentemente – acompanhados de ícones que nos dão acesso aos *clusters* dos diversos colaboradores, bem como às respectivas publicações individuais ou em co-autoria *(Figura 9)*.

As *Figuras 10*, *11* e *12* mostram, respectivamente, um fragmento da relação das 20 publicações de 'Mamiya, Edgar Nobuo'; a lista de autores que assinaram trabalhos com 'Savi, Marcelo Amorin'; e publicações conjuntas de 'Mamiya, Edgar Nobuo' e 'Savi, Marcelo Amorin'.

INFOCLUSTER AUTHORS

Type author initial:

35	FT ENM95(06)010
99	Produção Científica
100	MAMIYA, EDGAR NOBUO SILVA, E.P. DA
245	Pseudoelastic behavior under partial cycling: a rate independent one-dimentional model (Comportamento pseudoelástico em ciclos parciais: modelo unidimennsional independente de taxa. COBEM-CIDIM/95 - 13. Congresso Brasileiro e 2. Congresso Ibero Americano
35	FT ENM95(06)011
99	Produção Científica
100	MAMIYA, EDGAR NOBUO SILVA, E.P. DA
245	Viscous one-dimentional model for pseudoelastic materials (Modelo unidimensional viscoso para materiais pseudoelásticos). COBEM-CIDIM/95 - 13. Congresso Nacional e 2. Congresso Ibero Americano de Engenharia Mecânica. Anais (cd-rom). Belo Horizonte, MG, 1
35	FT ENM95(06)012
99	Produção Científica
100	MAMIYA, EDGAR NOBUO SILVA, E.P. DA
245	One-dimentional rate independent model for pseudoelastic materials (Modelo unidimensional independente de taxa para materiais pseudoelásticos). CILAMCE 95 - Congresso Ibero Latino Americano de Mecânica Computacional em Engenharia. Anais. Curitiba, PR, p.
35	FT ENM95(06)025
99	Produção Científica
100	MAMIYA, EDGAR NOBUO SIMO, J.C.
245	Assumed enhanced strain method for the description of strain discontinuities within a finite element (Método de pressão intensificada adotado para descrição de discontinuidades de pressão em um elemento finito). CILAME 95 - Congresso Ibero Latino American

Next >>>>

Figura 10. – Publicações de Mamiya, Edgar Nobuo (Fragmento).

Eis uma ferramenta que permite identificar automaticamente, de forma simples e flexível, os grupos de pesquisa de uma instituição, quais seus temas dominantes, a permanência e/ou a mobilidade de seus integrantes, etc. (ROBREDO, 2007), (ROBREDO; CANTANHEDE, 2005; 2006).

Figura 11. – Autores que publicaram com Savi, Marcelo Amorim, partir do Ícone do Cluster' da oitava linha da Figura 9).

Figura 12.- Publicações de Marcelo Amorin Savi, e Edgar Nobuo,em coautoria

Adiantando o que será exposto na Seção seguinte, acrescentamos uma sucinta informação sobre uma nova funcionalidade introduzida recentemente na aplicação (ROBREDO; CANTANHEDE, 2008), que permite tratar os nomes dos autores como tópicos (`*Terms*') que se relacionam diretamente com os vocábulos que figuram nos campos de *tag* 245, quando seus nomes foram registrados como orientadores.

Assim, ao digitar na janela que figura na parte superior da página de abertura, o sobrenome (ou parte dele) ou o nome completo de um autor, este pode ser recuperado dos campos 100 e 245, quando registrado na base, como mostra a *Figura 13*..

Figura 13. – Ao pesquisar como tema pelo nome de um autor que publicou trabalhos e também tem atuado como orientador, o InfoclusterWeb retorna o número de ocorrências que correspondem a cada caso.

Dessa forma pode-se obter, simultaneamente, a relação das publicações de um autor e/ou das teses e dissertações por ele orientadas, como mostram as *Figuras 14* e *15*

A pesquisa de nomes como temas abre outras possibilidades sobre as quais voltaremos na Seção seguinte, onde abordaremos a apresentação de outras funcionalidades do *InfoCluster*, centrando nossa atenção no cálculo da probabilidade de co-ocorrência de pares de termos, em um ou mais documentos, ou, mais propriamente, em suas representações armazenadas na base de dados.

Figura 14. Publicações de Mamiya, Edgar Nobuo (Fragmento).

Figura 15. – Orientações de Mamiya, Edgar Nobuo (Fragmento).

6.3.2 Identificação automática de clusters temáticos, ontologias e mapas de tópicos

A freqüência de co-ocorrência entre pares de termos é calculada pela ferramenta infométrica informatizada desenvolvida, utilizando a fórmula mostrada mais adiante. Se *Fi* representa a freqüência de um termo *i*, e *Fj* a freqüência de outro termo *j*, a freqüência de associação binária (co-ocorrência; *cowording*, em inglês) dos termos *i* e *j* poderá ser representada por *Fij*.

Para calcular a probabilidade de ocorrência' da associação binária (co-ocorrência), representada por *Eij*, pode-se utilizar a fórmula seguinte (POLANCO, 1993; 1995, ROBREDO e CUNHA, 1998):

$$Eij = Fij^2 / Fi * Fj \qquad [6]$$

O cálculo da ocorrência das associações binárias do termo central (por exemplo, *i*) com outros termos semanticamente relacionados *j, k, l, m*..., ou seja, *Fij*, *Fik, Fil, Fim*..., permite organizar os *clusters* temáticos desejados.

Ao clicar na página de abertura mostrada na *Figura 6*, sobre a opção *'InfoClusterTerms'*, é apresentada a página *'InfoClusterTerms'*, reproduzida na *Figura 16a*, de aspecto semelhante à página '*InfoClusterAuthors*' da *Figura 7*. Ao clicar em *'Opções de filtro'* é aberta uma nova janela *(Figura 16 b)* que permite selecionar as opções desejadas, evitando assim que determinados termos sejam exibidos nos *clusters* temáticos.

(a) (b)

Figura 16. – Janela 'InfoClusterTerms, para indicar o termo que será o centro do cluster temático (a), com a possibilidade de filtrar a exibição de termos indesejáveis (b).

A *Figura 17* reproduz um fragmento do *cluster* temático formado a partir do termo '*Trypanosoma*', introduzido na janela da página mostrada na *Figura 16*. O número que acompanha cada termo indica a probabilidade de coocorrência (*Eij*) de cada termo com *Trypanosoma,* nos documentos da base, calculado pela ferramenta, a partir da equação [6]. Os ícones junto a cada termo são sensíveis ao clique do mouse, quando conectado à Internet (manter '*Ctrl*' presionado) e permitem exibir o *cluster* do termo ou a referência às publicações.

Figura 17. – Termos relacionados semanticamente com Trypanosoma (Fragmento).

A *Figura 18* é uma representação de um *cluster* temático com o termo central '*protease*' que se encontra no canto inferior direito do cluster da *Figura 17*. Observe-se que em ambas as figuras – como é natural – a probabilidade de co-ocorrência dos termos '*trypanosoma* ' e '*protease* ' é a mesma (0,036, ou seja 3,6 por cento).

Os(as) leitores(as) podem abrir esse *cluster*, ou outros de sua escolha, clicando num termo, mantendo pressionada a tecla 'control', após selecionar o *cluster* escolhido, mediante um simples clique dentro da moldura que o enquadra*,* o que fará aparecer quatro pequenos círculos no meio das linhas que enquadram o *cluster*. Desnecessário lembrar que, para que as coisas funcionem, é necessário estar conectado à Internet.

Observe-se, ainda, que, em nossos exemplos, trabalhamos somente com simples referências bibliográficas dos trabalhos e que, como começa a ser praxe nos repositórios sérios, a inclusão de metadados, de resumos enxutos (ou seja, com menos *stopwords*), e de descritores / palavras-chave realmente significativos, bem como o uso de listas de sinônimos, e outras pequenas coisas como a identificação automática das variantes dos nomes dos autores, aliada ao uso de radicais eliminariam (ou neutralizariam) os efeitos de desinências e flexões (*stemming*), mas conservariam o significado dos vocábulos. São, pois, essas práticas, trilhas a serem abertas e exploradas, para melhorar a qualidade, rapidez e precisão dos processos de busca e recuperação. Os motores de busca, mineradores de texto, etc. vão agradecer, e nós também.

Encerramos a Seção com mais dois exemplos, que mostram como as rotinas precedentes podem contribuir, respectivamente, para a construção de listas de autoridade de nomes de pessoas (*Figura 19*), e de agrupamentos de variantes de termos (*Figura 20*).

Figura 18. – Cluster temático com o termo central 'protease'.

Não parece necessário insistir sobre os possíveis benefícios na normalização de tesauros e léxicos especializados, da representação e organização de informações e conhecimentos e, também, da qualidade de sua recuperação.

Chagas	211	**Tratamento**	280	**Trypano**cidal	1	**Infect**ada	2
Chagases	1	**Tratamento**s	22	**Trypano**soma	131	**Infect**adas	11
Chagasi	8			**Trypano**somatids	2	**Infect**ado	5
Chagasic	6			**Trypano**somes	3	**Infect**ados	35
Chagásica	15			**Tripano**somiasis	7	**Infect**ed	9
Chagásicas	1			**Trypano**ssoma	4	I**nfect**er	35
Chagásico	19					**Infect**ing	16
Chagásicos	33					**Infect**ion	62
Chagasin	2					**Infect**ions	8
Chagástica	1					**Infect**ious	16
						Infective	2
						Infectividade	2
						Infectologia	17
						Infectrions	8

Figura 19. - Exemplos de radicais de termos (mesmo com grafias erradas) relacionados com a doença de Chagas, e freqüências de ocorrência (ROBREDO; CANTANHEDE, 2006).

TEIXEIRA, A.R.L. (4)
TEIXEIRA, A.R.L.C. (10)
TEIXEIRA, ANTÔNIO, R.L. (15)
TEIXEIRA, ANTÔNIO, RAIMUNDO L.C. (12)
TEIXEIRA, ANTÔNIO, RAIMUNDO LIMA (6)
TEIXEIRA, ANTÔNIO, RAIMUNDO LIMA CRUZ (76)

Figura 20. – Identificação automática da forma mais abrangente do radical das diversas variantes do nome de um autor (incluídos erros de grafia) e freqüência de ocorrência de cada variante
(ROBREDO; CANTANHEDE, 2006).

Observe-se, ainda, que, em nossos exemplos, trabalhamos somente com simples referências bibliográficas dos trabalhos e que, como começa a ser praxe nos repositórios sérios, a inclusão de metadados, de resumos enxutos (ou seja, com menos *stopwords*), e de descritores / palavras-chave realmente significativos, bem como o uso de listas de sinônimos, e outras pequenas coisas como a identificação automática das variantes dos nomes dos autores, aliada ao uso de radicais eliminariam (ou neutralizariam) os efeitos de desinências e flexões (*stemming*), mas conservariam o significado dos vocábulos. São, pois, essas práticas, trilhas a serem abertas e exploradas, para melhorar a qualidade, rapidez e precisão dos processos de busca e recuperação. Os motores de busca, mineradores de texto, etc. vão agradecer, e nós também.

Encerramos a Seção com mais dois exemplos, que mostram como as rotinas precedentes podem contribuir, respectivamente, para a construção de listas de autoridade de nomes de pessoas (*Figura 19*), e de agrupamentos de variantes de termos (*Figura 20*). Não parece necessário insistir sobre os possíveis benefícios na normalização de tesauros e léxicos especializados, da representação e organização de informações e conhecimentos e, também, da qualidade de sua recuperação.

A forma de relacionar, na prática, as variantes de nomes pessoais poder-se-ia resolver com facilidade intoduzindo no motor de busca uma do tipo:

Search..... '**TEIXEIRA, A**NTÔNIO, RAIMUNDO LIMA CRUZ'
OR 'TEIXEIRA, ANTÔNIO, RAIMUNDO LIMA'
OR 'TEIXEIRA, ANTÔNIO, RAIMUNDO L.C.'
OR 'TEIXEIRA, A.R.L.C.'
OR 'TEIXEIRA, A.R.L.'

ou, bem mais simplesmente,

Search..... '**TEIXEIRA, A**&' (onde '&' indica truncagem),

o que não parece tarefa difícil para um informático experiente.

Considerando que, a cada dia que passa, a língua inglesa vem se firmando como a *língua franca* para a comunicação global na Web, valeria a pena considerar seriamente a possibilidade de se inspirar em modelos semelhantes aos citados muito criteriosamente por Dagobert Soergel (1997, p.212), dentre os quais se destacam o tesauro AGROVOC, do sistema AGRIS, patrocinado pela FAO – em cujo desenvolvimento e consolidação o Brasil teve uma participação significativa (ver Seção 6.2) – e o *Dictionary of Glass-Making* internacional e multilingüe.[45] Aí temos um nicho para pesquisa sobre aprimoramento e/ou inovação de ferramentas e rotinas que muito podem contribuir para o melhor uso dos motores de busca e recuperação, e das aplicações de mineração de texto.

6.3.3 Análise de citações e outras cientometrias

Embora a análise de citações não seja, propriamente uma aplicação nascida com a Web, dedicamos a ela um espaço dento das metrias da informação em ambiente Web, pois já se aplicava antes do advento da Internet, é com ela, que ganhou significativo desenvolvimento, tornando-se um poderoso auxiliar para estabelecer indicadores, que ajudem a monitorar, por exemplo, o impacto das publicações científicas, os títulos dos periódicos preferidos pelos especialistas, os congressos mais procurados, etc.

[45] O referido *Interntional Dictionary of Glass-Making* (ICG, 1965, 1967, 1968, 1974), na Nota 40, mais mereceria ser chamado de '*International Dictionaries*', vistos seus desdobramentos em diversas línguas, nos quais as entradas de cada termo remetem aos equivalentes – e às suas definições – em inglês. Trata-se, de fato, de um exemplo precursor do que vem sendo chamado de tesauros especializados (ou por domínios) plurilíngües, que muito podem contribuir para o aprimoramento da recuperação da informação em general e, mais particularmente, das ferrramentas de mineração de dados e textos, e/ou de descoberta de conhecimento. E significativo que, cinqüenta anos depois de seu desenvolvimento e implementação, essa experiência mereceu a atenção da prestigiosa editora Elsevier dos organismos internacionais Unesco e União Européia, que promoveram novas e sucessivas edições multilingües (ICG, 1983, 1992). A esse sucesso é inevitável associar os produtos contemporâneos e/ou antecessores, de outras experiências bem sucedidas, que ajudaram a estabelecer sistemas colaborativos, mesmo entre paises separados pela chamada 'cortina de ferro', de triste memória. Poder-se-ia citar, dentre elas, como exemplos premonitórios de uma 'pré-globalização', antes da hora, o *Thesaurus Verrier*, com seus esquemas – ou mapas? – com suas setas que, partindo de um descritor central, apontavam para os dscritores de um certo domínio, mostrando suas relações hierárquicas ou semânticas (ROBREDO, 2005 p.161-162), (INSTITUT DU VERRE., 1972 p.26-27), (LAUREILHE, 1972); a '*Extension Verre*' da CDU, que permitia indexar, com os mesmos códigos numéricos, os mesmos conceitos que se expressariam com termos diferentes, nas diferentes línguas (FID, 1984); e o *Pool of Abstracts*, que promovia o intercâmbio de resumos informativos tanto de artigos e comunicações técnico-científicas, como de patentes entre diversos países (IPGA, s.d.), (VERRE Online, 2007). Essa experiência, agora com uma 'roupagem' contemporânea, continua viva com o nome de '*GlassFile*'.

A análise de citações abrange numerosas mensurações quantitativas que são cruciais para:
- aprimorar a recuperação de documentos;
- determinar do que trata um documento;
- revelar relações entre documentos, seus assuntos e seus autores.

Alguns dos métodos utilizados são:
- a determinação da idade das citações (obsolescência);
- a análise das co-citações;
- a determinação do fator de impacto;
- a análise do acoplamento ou emparelhamento de co-citações (*co-citation coupling*);
- a análise do acoplamento ou emparelhamento bibliográfico (*bibliographic coupling*);
- a analise das citações de outros autores;
- o índice de autocitação;
- o indice de imediação (*immediacy index*);
- a identificação de grupos colaborativos; etc.

As análises de citações encontram numerosas aplicações, que derivam do pressuposto que, quando um autor cita outro autor, algum tipo de relação ou associação se estabelece entre eles, entre suas obras, entre os periódicos onde se publicaram seus trabalhos, entre suas áreas de interesse, e assim por diante. Um desdobramento interessante é a determinação do fator de impacto de um autor em sua área de atuação, contando o número de vezes que ele foi citado por outros autores. É claro que a seleção das citações deve ser criteriosa, pois, com observam Osareh (1996, *apud* UTEXAS [s.d.]) e Raber (2003, p. 78, *apud* JAYROE, 2008), às citações usadas para corroborar uma afirmação ou confirmar um resultado, pode-se misturar alguma outra com um viés negativo sobre o autor citado. Também é claro que, em algum caso, a intenção pode ser determinar um fator de impacto negativo.

Gráfico 7. – Curva genérica de citações (AMIN; MABE, 2000).

Num interessante artigo, Amin e Mabe (2000), apresentam um gráfico que mostra o que os autores denominam "curva generalizada de citações", reproduzida no *Gráfico 7*.

A "meia-vida de citação" (*cited half-life*) é uma medida da velocidade de declínio da curva de citações. Definida como o número de anos necessários para que o número de citações se reduza em 50 por cento do valor inicial[46], é uma medida do tempo em que os trabalhos publicados em uma revista, continuam a ser citados.

No *Gráfico 8*, os autores apresentam o fator de impacto de três tipos de publicações ('cartas à redação', artigos de periódicos e publicações periódicas). Para os artigos curtos (do tipo 'cartas à redação') o pico de citações da curva se situa nos dois anos depois da publicação, com um fator de impacto medido por um número de citações relativamente baixo.

Gráfico 8 – Fator de impacto e tipo de publicação (AMIN; MABE, 2000).

Para os artigos originais, o pico se situa próximo aos três anos depois da publicação (mais afastado do ponto de origem) e, conseqüentemente, o fator de impacto será menor. Para as revistas, o pico máximo de citações aparece, ainda, mais atrasado, mas o elevado número absoluto de citações se traduz por um fator de impacto também mais elevado.

O fator de impacto, o índice de imediação – ou de contigüidade – e a meia-vida de citação são as metrias mais comumente utilizadas para medir e comparar o interesse das publicações científicas, sua repercussão, atualidade ou obsolescência, etc. Num trabalho de Baudoin *et al*. (2004), em que aproveitam o modelo representacional de Amin e Mabe (2000), já citado, para discutir as realidades, mitos e tendências dos indicadores bibliométricos e/ou cientométricos, na área médica e, de forma mais específica, em biomedicina, áreas essas que, junto com a econometria são grandes usuárias tradicionais de indicadores. Baudoin *et al*. (2004) definem o fator de impacto de uma publicação periódica como a "*medida num tempo 't' de número de citações registrado no decorrer dos anos 't_1' e 't_2' para os*

[46] No *Gráfico 6*, o valor representado para a meia-vida é de 6 anos. O conceito de meia-vida foi tomado emprestado da física nuclear, para medir a velocidade de declínio de um material radioativo, e se define como o tempo necessário para que a radioatividade desse material seja reduzida à metade de seu valor em um determinado momento. (Ver, por exemplo: Burton e Kebler (1960), Brookes (1970) e Száva-Kováts (2002)). A comparação com a física nuclear não é senão uma metáfora, pois o declínio da radioatividade com o tempo não parte de zero até atingir um pico e, logo, decrescer, como é o caso da curva da obsolescência. Registremos, também, que a tradução de "*half-life*" por 'vida média' é incorreta.

artigos publicados em esse período"[47] e calculam o fator de impacto para a revista *Nature*, em 2002, como "o quociente do número de citações dos artigos publicados em 2001 e 2002 pelo número de artigos publicados em 2000 e 2001", ou seja:

$$(33448 + 25955)/(1013 + 939) = 30,432$$

onde (33448 + 25955) são os respectivos números de citações em 2002 e 2001, e (1013 + 939), os respectivos números de artigos publicados em 2001 e 2000. As revistas com maior fator de impacto são:

- *Nature*, com fator de impacto ~ 29 (ver acima), e
- *Science*, com fator de impacto ~ 26.

Para poder calcular o fator de impacto dos periódicos brasileiros e/ou de outros países latino americanos seria necessário dispor de uma base de dados semelhante à elaborada pela *Colorado State University* a partir do periódico *The Journal Report (JCR)*, infelizmente de acesso restrito e que, além do mais, deixa de incluir uma grande parte de revistas significativas que não publicam em língua inglesa. No caso de *Web of Science* ou de *Thomson Science*, a situação não é muito mais brilhante.[48]

O índice de imediação (ou contigüidade, proximidade – "*immediacy*", em inglês) de um periódico se calcula dividindo o número de vezes que o periódico foi citado num de terminado ano pelo número de artigos que publicou nesse ano, ou seja, é a média do número de citações por artigo. O índice mede a rapidez com que são citados os itens publicados nesse periódico, desde a data de sua publicação. Permite identificar os periódicos que publicam artigos mais "quentes". Para a maioria dos periódicos científicos esse índice varia entre 0 e 1. Somente duas revistas, *Nature* e o *Journal of Experimental Medicine*, apresentam um índice maior que 7 (BAUDOIN *et al*., 2004).

O conceito de obsolescência, introduzido por Charles Gosnell (1943), é uma aplicação bibliométrica à qual, em nossa opinião, não se presta a devida atenção – talvez porque não sempre bem entendida –, que poderia trazer significativos benefícios na gerência de bibliotecas, e mais especialmente na gestão de acervos.

O acoplamento ou emparelhamento de co-citações (*co-citation coupling*) é usado para estabelecer as similaridades entre dois documentos. Se dois trabalhos *A* e *B* são citados por um trabalho *C*, pode-se dizer que os dois primeiros guardam alguma relação, mesmo se estes não se citam entre si. A relação entre os trabalhos *A* e *B* será tanto maior, quanto maior o número de outros trabalhos que os citem.[49]

O acoplamento ou emparelhamento bibliográfico (*bibliographic coupling*) segue o mesmo princípio do emparelhamento de co-citações, mas de uma forma que é a imagem no espelho deste último. O emparelhamento bibliográfico estabelece um elo entre dois trabalhos que citam os mesmos trabalhos. Ou seja, se os trabalhos *A* e *B* citam o trabalho *C*, pode-se dizer que guardam uma relação, mesmo se *A* e *B* não se citam mutuamente. Quanto maior o número de outros trabalhos que os citem, maior será seu inter-relacionamento (UTEXAS [s.d.]).

---

[47] Observe-se que o momento '*t*' será o ano completo seguinte ao ano 't_2', para poder completar a contagem das citações dos artigos publicados em 't_2'.

[48] A título de exemplo, convidamos o leitor a consultar as seguintes referências: Danchin ([s.d.]), Meneghini e Packer (2007), Almeida (2008), Bressan et al. (2008), Ferraz *et al*. (2008), Marques (2008), Strehl e Stumpf (2009), Tess *et al*. (2009), Anastasiadis *et al*. (2009), Castro (2009).

[49] Na literatura brasileira sobre bibliometria, o termo inglês "*coupling*" parece, até agora, ter sido traduzido exclusivamente por 'acoplamento', sem perceber que este vocábulo, em vernáculo, encontra um uso maciço na linguagem técnica, em física, mecânica, eletricidade, etc., em expressões tais como 'acoplamento elétrico, acoplamento mecânico, e por aí vai. Cabe levantar a questão se outros termos, tais como, entre outros, 'emparelhamento', 'coadunação', 'junção', 'combinação', não poderiam ser objeto de uma reflexão mais profunda. A título de curiosidade, sem nenhuma pretensão conclusiva, assinalamos que, numa pesquisa na Web, a questão 'acoplamento E/OU bibliográfico', no Brasil, rendeu 60.900 páginas, sendo que, a partir da segunda página, começam a aparecer páginas ligadas a diversos aspectos tecnológicos, inclusive com certo viés de propaganda. Para dirimir uma dúvida, foi repetida a busca usando a expressão composta "acoplamento bibliográfico", no Brasil, o que rendeu 88 páginas/respostas, a primeira metade das quais com respostas razoavelmente adequadas. Finalmente, a busca 'emparelhamento AND (bibliográfico OR bibliografia)', em suas diversas combinações, rendeu entre 20 e 33 respostas. Dois exemplos: um correspondente a uma dissertação de pós-graduação, da Universidade Paulista (UNIP), de São Paulo, na qual se lê "*O modelo de co-citação difere significativamente de modelos de* ***emparelhamento bibliográfico*** [o grifo é nosso], *conforme explica Small (1973), pois este é um relacionamento fixo e permanente...*" (VARGAS, 2009; p.21); outro, absolutamente genérico, na construção de experimentos, com numerosos desdobramentos práticos, onde se lê: "*...um* ***experimento*** *é 'um tipo de pesquisa científica no qual o pesquisador manipula e controla uma ou mais variáveis independentes e observa a variação nas variáveis dependentes concomitantemente à manipulação das variáveis independentes.' O propósito de manipular e medir as variáveis no experimento é captar causalidade (relação entre causa e efeito). As variáveis independentes são responsáveis pelas possíveis causas, e as variáveis dependentes sinalizam os efeitos. Duas variáveis podem ter altíssima correlação, mas não necessariamente uma é causa da outra...*" , e a plavra **emparelhamento** aparece oito vezes (WIKIPEDIA, 2009). Fica em aberto a questão da escolha do melhor termo.

Eugene Garfield, já citado (THOMSON REUTERS, 2006)[50], criou em 1960 o *Citation Index*, que mostrava que os trabalhos publicados eram citados ou referenciados por uns poucos trabalhos, e como estes tendem a se concentrar em torno de uma fonte ou publicação periódica, e – o que é mais importante – que a quantidade de citações não guarda, necessariamente, uma relação com o valor do trabalho para área a que se refere, ou seja, com o fator de impacto (HERTZEL, 2003, p. 319, 320, *apud* JAYROE, 2008).

Aqui encerramos nosso sobrevôo sobre as metrias da informação, antes da Web, não sem antes lembrar um artigo de Edson Nery da Fonseca (1973) – oportuno na data de sua publicação e, talvez, ainda mais oportuno hoje – que consideramos leitura obrigatória para quem deseje reconstruir a história dos primórdios da bibliometria. História, na qual a marcante presença do Brasil é um fato que, lamentavelmente, se esquece pelo afã de olhar só para os 'modelos' externos.[51]

6.3.4 Mineração de textos em bases de periódicos científicos especializados

Na Seção 6.1.1, foram expostos alguns exemplos de aplicação de técnicas infométricas ao monitoramento da produção científica, utilizando a base de dados ABCDM, que hoje reúne mais de cinco mil registros bibliográficos de artigos e comunicações, publicados em periódicos de arquivologia, biblioteconomia, ciência da informação e museologia, desde 1972 até hoje. Alguns aspectos subjacentes, entretanto, merecem ser destacados. Aspectos esses decorrentes da metodologia participativa de alimentação da base de dados, que privilegia os aspectos colaborativos, a continuidade do trabalho, o aproveitamento e o enriquecimento dos resultados e contribuições das turmas precedentes, juntamente com a compreensão do processo como um todo.

Em soma, a construção de uma linha de trabalho continuado, com uma visão crítica, em um ambiente participativo e colaborativo, que gera resultados e se constitui numa semente que pode fazer brotar novas vocações de pesquisa e ampliar horizontes.

Os alunos das turmas sucessivas de graduação em biblioteconomia, que vêm cursando a disciplina 'Planejamento e elaboração de bases de dados', aprendem a elaborar, com o rigor necessário, os registros bibliográficos que alimentam a base de dados, entendem a organização dos dados dos registros em campos marcados com etiquetas (*tags*) apropriadas, dentro da base, e a importância de representar os conteúdos temáticos com palavras-chave ou descritores adequados, suscetíveis de garantir a qualidade da recuperação de informações. Na medida em que os alunos avançam em seus estudos, abrem-se para eles novos horizontes e se consolida a importância dada à interdisciplinaridade. Esse foi o caso de uma aluna que, uma vez formada, se inscreveu no 'Curso de Especialização em Engenharia Elétrica / Área de Gestão da Tecnologia da Informação', e de cuja monografia de finalização, elaborada em comjunto com outra aluna da área de computação – orientada por um professor do Departamento de Energia Elétrica, da Faculdade de Tecnologia da UnB, e co-orientada por um dos autores deste Capítulo[52] – passaremos a falar, como exemplo de interação interdisciplinar e colaborativa (SOUZA; NAGLIS, 2008).

A monografia apresenta um estudo do processo de mineração de dados, utilizando a ferramenta *Rapid Miner* (RAPID I, 2009) com a base de dados ABCM. Foram analisados os títulos dos artigos, identificados os assuntos mais relevantes, nas décadas de 1970 a 1990 e dos anos 2000 a 2007, e calculados os índices *tf-idf*, que definem a importância de cada termo na coleção de documentos.

Segundo as autoras, citando Araújo Júnior (2007), a mineração de textos apresenta-se como uma ferramenta capaz de '*sumarizar*' um conjunto de documentos em agrupamentos (*clusters*), apresentando-os na forma de grafos indicativos das relações semânticas dos termos. Assim, o usuário obtém uma idéia mais clara do assunto de que trata a coleção de páginas, sem precisar lê-las todas. Num processamento preliminar, a coleção é carregada, processada e transformada numa representação estruturada dos documentos, denominada '*bag of words*' (BOW), segundo Pires (2008, *apud* SOUZA; NAGLIS, 2008). O processamento posterior depende do objetivo que se deseja atingir com a mineração da *BOW*. As etapas do pré-processamento, até o calculo dos índices *tf-idf,* são descritas detalhadamente na monografia. O cálculo da freqüência foi realizado da seguinte forma:

- Após contagem dos termos para cada documento da coleção, é determinada sua freqüência;

[50] Ver também o artigo "Citation Index", da Wikipedia (2010), e o artigo do próprio Garfileld, publicado em *Nature* (GARFIELD, 1968).

[51] A título de ilustração deixamos aqui o registro que, numa recente pesquisa no Google, com a estratégia 'infometria OR bibliometria AND "fator de impacto" AND citações', no Brasil, retornou mais de 11.700 respostas. Convidamos os(as) leitores(as) a refazer a busca.

[52] Jayme Leiro Vilan Filho.

- A cada termo t_i de cada documento da BOW, é atribuído um peso; esse peso é o número de ocorrências do termo no documento *(tf)*, modificado segundo uma escala de importância do termo *(tdf)* chamada freqüência inversa do documento;
- O *tf* do termo t_i no documento é:

$$(t_i) = \frac{n_i}{\sum_{j=1}^{k} n_i}$$

onde n_i é o número de ocorrências do termo no documento e o denominador é o somatório de todos os número de ocorrências de todos os termos de um documento;

- O *idf* do termo t_i é:

$$idf(t_i) = \log \frac{N}{df(t_i)}$$

onde N é o número total de documentos do conjunto e $df(t_i)$ é o número de documentos onde o termo t_i aparece, ou seja, ≠ 0. Então:

$$tfidf(t_i) = tf(t_i) idf(t_i)$$

Para obter uma freqüência com o *tf-idf* alto, ou seja, o termo ser representativo para o documento, é necessário que ele tenha uma ocorrência alta no documento e baixa dentro na coleção. Após o cálculo da freqüência dos termos de cada documento, o processo de transformação da coleção em dados numéricos estará concluído.[53] O resultado encontra-se na *Tabela 10*.

Quatro breves comentários, após uma análise super-rápida dos resultados mostrados na *Tabela 10* :

1) O valor dos índices *tf-idf* em cada linha, da esquerda para a direita, aumenta sensivelmente, não somente porque as atualizações sucessivas da base aumentaram o número de registros nela contidos, mas também, e sobre tudo, porque a quantidade de publicações na área também aumentou bastante;

2) o termo 'documentação', importado da Europa na década de 70 do passado século, que ocupa uma posição importante na primeira coluna, desaparece nas décadas seguintes devido à influência crescente da expressão 'ciência da informação';

3) O termo 'informação' ganha destaque crescente no decorrer do tempo, e

4) termos como 'conhecimento', 'sociedade', 'gestão', 'tecnologia', 'pesquisa', 'redes' e 'digital' ganham destaque e/ou aparecem pela primeira vez entre os mais importantes dos quinze termos mais freqüentes, no período 2000-2007.

Os dados da *Tabela 10* serviram para um estudo mais aprofundado de um dos autores (VILAN FILHO, 2009). Do conjunto de 15 termos de cada década, alguns dos quais se repetem, foram feitas buscas na ABCDM para obter o número de títulos em que o termo e as expressões derivadas ocorrem com percentual maior ou igual a 0,5% do total da década (*Tabela 11)*.

Cabe destacar que o termo 'Biblioteconomia' e o termo 'Ciência da Informação' apresentam comportamentos opostos. Enquanto 'Ciência da Informação' ocorre em 3,2% dos artigos da década de 1980, passando para 3,9% nos anos 1990 e atingindo 7,8% nos anos 2000, 'Biblioteconomia' faz o caminho inverso com percentuais de 10,2% (década de 1980), 6,8% (década de 1990), e 3,5% (década de 2000). Também diminuem os percentuais dos termos 'bibliotecas' (20,7% para 12,2% e 9,6%), e 'bibliotecas públicas' (2,5% para 1% e 0,8%). No caso do termo 'bibliotecário(s)' há uma queda entre os anos 80 (5,3%) e 90 (3,3%) com estabilização nos anos 2000 (3,4%), com o termo 'profissional(ais) da informação' seguindo o caminho inverso (0,8% para 1% e 2,2%). Foram identificados, ainda, como termos de

[53] Calma pessoal! Nada de pânico com as fórmulas. A ferramenta calcula tudo.

interesse crescente o termo 'conhecimento' (0,8% para 3,1% saltando para 7,7%), 'gestão' (0,5% para 1,5% saltando para 5,1%), com destaque para a combinação de ambos 'gestão do conhecimento' (0% para 0,2% chegando a 1,7%), além de 'gestão da informação' com percentuais iguais exceto na década de 2000 (0% para 0,2% e 1,1%).

Tabela 10. – Índices tf-idf dos 15 termos mais relevantes de cada década.

1970-1979		1980-1989		1990-1999		2000-2007	
TERMOS	**tf-idf**	**TERMOS**	**tf-idf**	**TERMOS**	**tf-idf**	**TERMOS**	**tf-idf**
biblioteca(s)	30,3609	biblioteca(s)	40,4659	informação(ões)	46,3521	informação	59,0041
informação(ões)	18,6440	informação(ões)	33,8240	biblioteca(s)	36,9688	biblioteca(s)	39,2129
pública(s)	11,8287	biblioteconomia	19,5343	biblioteconomia	20,8046	ciência(s)	34,9069
brasileira(s)	9,8589	bibliotecário(s)	17,7470	Brasil	19,8594	conhecimento	30,4812
biblioteconomia	8,9636	arquivo(s)	17,2708	ciência(s)	18,5023	sociedade	24,1498
ciência(s)	8,3253	pesquisa	16,8996	bibliotecário(s)	15,4466	gestão	23,3840
nacional	8,0801	estudo(s)	16,3498	serviço(s)	14,9858	profissional(is)	22,7289
bibliografia(s)	7,7739	ciência(s)	14,9195	estudo(s)	14,4245	tecnologia(s}	21,1875
Brasil	6,8850	brasileira	13,7330	profissional(is)	14.0407	estudo(s)	20,7921
documentação	6,8022	avaliação	13,6661	pesquisa(s)	13,0568	Brasil	19,7284
bibliotecário(s)	6,2041	Brasil	13,5634	conhecimento	13,0555	análise	19,4812
pesquisa(s)	6,1914	usuário(s)	12,3234	memória	12,7800	pesquisa	19,1501
sistemas	5,9677	serviço(s)	12,1712	desenvolvimento	12,1067	científica	16,9950
classificação(ões)	5,8067	análise	11.4095	educação	12,0771	rede(s)	15,4974
literatura	5,6497	ensino	11,2936	científica	11,1646	digital	14,8435

Além desses, crescem os temas ligados aos termos 'sociedade' (1% para 1,8% até 4,7%) e 'sociedade da informação' (0%, 0,3% e 2,8%), bem como crescem os percentuais dos termos 'memória' (1,2%, 2,2% e 2,7%) e 'redes' (1,2%, 2,1% e 2,3%). Parece também diminuir o interesse pelos temas relacionados com os termos 'ensino' (3,6% para 2,5% caindo para 2%) especialmente 'ensino de biblioteconomia' (0,9% para 0,7% caindo para 0,1%).

Portanto, há crescimento evidente de artigos com termos mais ligados à ciência da informação (sociedade da informação e profissionais da informação), em detrimento dos ligados à biblioteconomia (biblioteca(s), bibliotecários(s) e ensino de biblioteconomia).

6.3.5 Aplicações inferenciais

A importância das análises de freqüência de uso de termos significativos sós ou associados com outros (co-ocorrências) pode ser ilustrada mediante um exemplo próximo de todos nós. Imaginemos que alguém, nas décadas de 60 e 70, buscasse informações que falassem alguma coisa sobre transporte e álcool. Certamente encontraria informações sobre o transporte de álcool caminhões ou por trem. De repente, uns anos mais tarde, a quantidade de publicações localizadas a partir dos mesmos temas, não somente aumentou, mas apontou para um fato novo: o álcool como combustível usado nos meios de transporte. Havia nascido o Pró-Álcool. Poucos anos depois, a literatura sobre o assunto torna-se mais escassa. O Pró-Álcool tinha sido desmontado... Mais na frente, agora, novo aumento de publicações e... novos termos associados. É a ressurreição do álcool como combustível para todo tipo de veículo e surgimento de novos combustíveis (gás, biodiesel, etc.). Em resumo, o acompanhamento da variação da freqüência de uso de determinados temas definidos por um pequeno número de descritores ou palavras-chave, junto com a variação dos grupos em que estes se reúnem no transcurso do tempo, permite monitorar as atividades de um setor, e/ou descobrir tendências.

Esse exemplo nos leva a falar das curvas estocásticas. Elas baseiam-se no princípio de que todos os processos e sistemas – de fato, pensando bem, tudo pode-se associar a algum tipo de processo ou sistema –, surgem, crescem até um certo nível a partir do qual o crescimento torna-se cada vez mais lento, até parar e, depois de um certo tempo... ninguém lembra mais de que existiram. A *Figura 6* é uma representação genérica desse tipo de curvas, extremamente útil para acompanhar inúmeros processos no decorrer do tempo. Em outros termos, para montar e interpretar séries históricas.

Tabela 11. – Termos mais freqüentes nos títulos de artigos de periódicos das áreas de informação por década em ordem alfabética (1980-2007).

TERMO\DÉCADA	1980		1990		2000		1980-2000	
	n° art.	%	n° art.	%	n° art.	%	n° art.	%
Análise	43	4,9%	36	2,9%	90	5,3%	169	4,4%
Arquivo(s)	47	5,3%	37	2,9%	51	3,0%	135	3,5%
Avaliação	44	5,0%	32	2,5%	40	2,4%	116	3,0%
Biblioteca(s)	183	20,7%	153	12,2%	162	9,6%	498	13,0%
<Biblioteca(s) Pública(s)>	22	2,5%	13	1,0%	14	0,8%	49	1,3%
Bibliotecário(s)	47	5,3%	41	3,3%	57	3,4%	145	3,8%
Biblioteconomia	90	10,2%	85	6,8%	59	3,5%	234	6,1%
Brasil	46	5,2%	89	7,1%	84	5,0%	219	5,7%
Brasileira	26	2,9%	22	1,7%	30	1,8%	78	2,0%
Ciência	41	4,6%	65	5,2%	172	10,2%	278	7,3%
<Ciência da Informação>	28	3,2%	49	3,9%	131	7,8%	208	5,4%
Ciência(s)	48	5,4%	74	5,9%	184	10,9%	306	8,0%
Científica	37	4,2%	38	3,0%	69	4,1%	144	3,8%
Conhecimento	7	0,8%	39	3,1%	129	7,7%	175	4,6%
Desenvolvimento	28	3,2%	39	3,1%	31	1,8%	98	2,6%
Digital	0	0,0%	4	0,3%	50	3,0%	54	1,4%
Educação	18	2,0%	38	3,0%	48	2,8%	104	2,7%
Ensino	32	3,6%	31	2,5%	33	2,0%	96	2,5%
<Ensino de Biblioteconomia>	8	0,9%	9	0,7%	1	0,1%	18	0,5%
Estudo(s)	58	6,6%	57	4,5%	78	4,6%	193	5,0%
Gestão	4	0,5%	19	1,5%	86	5,1%	109	2,8%
<Gestão da Informação>	0	0,0%	2	0,2%	19	1,1%	21	0,5%
<Gestão do Conhecimento>	0	0,0%	3	0,2%	29	1,7%	32	0,8%
Informação	166	18,8%	379	30,1%	555	32,9%	1.100	28,7%
<Informação Científica>	16	1,8%	2	0,2%	8	0,5%	26	0,7%
Informação(ões)	177	20,0%	397	31,6%	583	34,6%	1.157	30,2%
Memória	11	1,2%	28	2,2%	45	2,7%	84	2,2%
Pesquisa	47	5,3%	41	3,3%	73	4,3%	161	4,2%
Pesquisa(s)	55	6,2%	44	3,5%	75	4,4%	174	4,5%
Profissional(is)	25	2,8%	41	3,3%	64	3,8%	130	3,4%
<Profissional(is) da Informação>	7	0,8%	12	1,0%	37	2,2%	56	1,5%
Rede(s)	11	1,2%	27	2,1%	39	2,3%	77	2,0%
Serviço(s)	35	4,0%	47	3,7%	38	2,3%	120	3,1%
<Serviço(s) de Informação(ões)>	6	0,7%	25	2,0%	10	0,6%	41	1,1%
Sociedade	9	1,0%	23	1,8%	80	4,7%	112	2,9%
<Sociedade da Informação>	0	0,0%	4	0,3%	48	2,8%	52	1,4%
Tecnologia(s)	23	2,6%	56	4,5%	71	4,2%	150	3,9%
Usuário(s)	34	3,9%	22	1,7%	33	2,0%	89	2,3%
Total de Artigos na Base	883	-	1.258	-	1.686	-	3.827	-

Observação. – Os percentuais são calculados pelo total de artigos da respectiva década. Foram relacionados apenas os 15 termos mais freqüentes em cada década e seus termos derivados (no formato <termo>). O total de artigos da base não é a soma dos artigos de cada termo, mas o número total de registros na base de dados para cada década.

Observe-se que, em qualquer ponto da curva, pode surgir um fato novo suscetível de modificar completamente o traçado da curva do processo que estava sendo estudado. A detecção dessas variações, aliada à análise infométrica das ocorrências dos termos significativos em publicações técnicas ou nos registros de patentes, bem como a possibilidade de 'seguir' e monitorar os caminhos do mercado ou de um concorrente, abrem um espaço de grande interesse para a infometria inferencial, e para as vigílias tecnológicas e estratégicas. Em suma, para uma inteligência competitiva, com bases mais sólidas para orientar decisões e escolhas

Figura 6.- Curva estocástica ou curva em S.

Os resultados de um estudo realizado no início da década de 70, financiado pela indústria francesa do vidro, mostrou o interesse de identificar parâmetros para tomada de decisão quanto à conveniência de investir em pesquisa para lançamento de novos produtos, em função do grau de avanço dos concorrentes na corrida para a introdução no mercado de um produto similar (ROBREDO,1970). A pesquisa restringiu-se a um certo número de produtos e processos, em relação aos quais foram registradas, no decorrer de meses e anos, a variação no número de patentes requeridas e as datas de lançamento no mercado de novos produtos.

O estudo foi desenvolvido com base na premissa de que o número de patentes solicitadas por uma companhia industrial cresce desenhando uma curva em forma de *S*, até atingir um ponto crítico – definido de acordo com o nível de consistência e maturidade de seus processos de pesquisa e capacidade de desenvolvimento e inovação – que é seguido por
um decréscimo regular. Foi possível estabelecer assim uma correlação entre o número de patentes requeridas e o ponto da curva em que o lançamento dos novos produtos ocorria, e como conseqüência definir um programa otimizado de pesquisa e inovação,... o que deu certo.

Lembremos, ainda, para encerrar esta Seção, que a infometria, com seus aglomerados (ou *clusters*) temáticos, é um poderoso auxiliar na conceituação e construção dos mapas de tópicos, chamados a desempenhar um papel importantíssimo no processo de desenvolvimento da *Web* semântica.

7. CONSIDERAÇÕES FINAIS DO CAPÍTULO 10

Todas essas idéias, exemplos e realizações perfeitamente se enquadram na sistematização elaborada por Derek de Solla Price (1976) sobre a evolução da ciência, que ele mesmo sintetiza assim: "Se as coisas evoluem de acordo com um quadro conhecido previamente e se podemos determinar em que ponto da curva evolutiva situa-se determinado fenômeno, em certo momento, [...] então seremos capazes de prever razoavelmente - supondo que fatos inesperados não intervenham - como as coisas provavelmente acontecerão."

Encerraremos com um convite à reflexão, que já tinhamos formulado antes (ROBREDO, 2000) (ROBREDO; CANTANHEDE, 2006).

Se comparados os respectivos desenvolvimentos da infometria e da econometria é impressionante o impacto desta última na sociedade contemporânea: seis econometristas laureados com o Prêmio Nobel entre 1969 e 2003![54].

Será que teremos um dia um cientista da informação, com um forte viés infométrico, entre os laureados com o Prêmio Nobel? Ou teremos ainda que esperar muito tempo para que os políticos entendam que mais importante que o capital é o conhecimento e a inteligência?

8. BIBLIOGRAFIA DO CAPÍTULO 10

ACKLAND, Robert. Using Facebook as a Data Source and Platform for e-Researching Social Networks. Disponível em: http://voson.anu.edu.au/papers/paper-Ackland-2008.pdf. Acesso em setembro 2009.

AGUILLO, I.F. Propuesta para la incorporación de indicadores cibermétricos al sistema de metadatos Dublin Core. V Taller Iberoamericano e Interamericano de Indicadores de Ciencia y Tecnología. Montevideo (Uruguay), 15-18 October 2001. http://www.ricyt.edu.ar/interior/interior.asp?Nivel1=6&Nivel2=2&IdTaller=8&Idioma=

AGUILLO, I.F. Cibermetría con motores de búsqueda. Indicadores Web de Universidades Iberoamericanas. Tercer Taller Iberoamericano de Bibliometría. Madrid, Spain, 3-5 March 2003

AGUILLO, I.F. Evolución reciente de los Indicadores cibermétricos de las Universidades Iberoamericanas. Congreso: II Seminario Internacional sobre Estudios Cuantitativos y Cualitativo de la Ciencia y la Tecnología "Prof. Gilberto Sotolongo Aguilar" INFO 2004. La Habana, Cuba, 15 April 2004

AGUILLO, I.F. Indicadores de contenidos para la web académica iberoamericana. *BiD: textos universitarios de biblioteconomia i documentació*, n.15, desembre 2005a. Disponível em: http://www2.ub.edu/bid/consulta_articulos.php?fichero=15aguil2.htm. Acesso em setembro 2009.

AGUILLO, I.F.; Granadino, B.; Llamas, G. Posicionamiento en el Web del sector acadêmico iberoamericano. *Interciencia*, v.30, n.12, 2005c, p.1-5.

AGUILLO, Isidro F.; GRANADINO, Begoña; ORTEGA, Jose Luis; PRIETO, Jose Antonio. What the Internet says about Science – Universities can be ranked based on web indicators. *The Scientist*, v.19, n.14, 2005b, p.10. Disponível em: http://www.the-scientist.com/article/display/15616/. Acesso em: setembro 2009.

AGUILLO, Isidro F. Cibermetría: Introducción teórico-práctica a una disciplina emergente. 2008. Digital.csic (Acceso abierto a documentos digitales). Disponível em: http://hdl.handle.net/10261/4143. Acesso em: setembro 2009.

ALMEIDA, Gil Lúcio. Brazilian Journal of Physical Therapy: A history of dedication, hard work and success (Revista Brasileira de Fisioterapia: uma história de dedicação, trabalho e sucesso) – Editorial. *Revista Brasileira de Fisioterapia*, v.12, n.5, 2008, p. v-vi. ISSN 1413-3555. Disponível em: http://www.scielo.br/pdf/rbfis/v12n5/en_a01v12n5.pdf. Acesso em setembro 2009.

ALMIND, Thomas C.; INGWERSEN, Peter. Informetric analysis on the World Wide Web: A methodological approach to "webometrics". *Journal of Documentation*, v.53, n.2, 1997, p.404-426.

AMIN, M.; MABE, M. Impact factors: use and abuse. Perspectives in Publishing (News letter for journal editors).*Elsevier Science,* 2000, 1, p.1-6). Disponível em: http://www.elsevier.com/framework_editors/pdfs/Perspectives1.pdf. Acesso em setembro 2009.

ANASTASIADIS, Aristoklis D.; ALBUQUERQUE, Marcelo P. de; ALBUQUERQUE, Marcio P. A. characterization of the scientific impact of Brazilian institutions. *Brazilian Journal of Physics*, v.39, n.2A, 2009, p.511-518. Disponível em: http://www.sbfisica.org.br/bjp/files/v39_511.pdf. Acesso em setembro 2009.

ANNUAL Review of Information Science and Technology – ARIST, v.30, 1996. ISBN 0066 4200 1996.

ANNUAL Review of Information Science and Technology – ARIST, v. 39**,** p. 1-609, 2005. DOI**:** 10.1002/aris.1440390101.

[54] A econometria pode ser definida como 'a medida da economia'. É um domínio dessa ciência social aplicada que se serve de métodos estatísticos para o estudo empírico de teorias e relações econômicas. É uma combinação da economia, matemática, estatística, estatística econômica e teoria econômica. (WIKIPEDIA, 2009). Infometria e cientometria também poderiam ser definidas, respectivamente, como a 'medida da informação' e do 'comportamento da ciência'. São domínios da ciência da informação, que também é uma ciência social aplicada que se serve de métodos estatísticos e matemáticos, mas com uma profunda fundamentação teórica. Os ganhadores do Prêmio Nobel em Ciências Econômicas (Nobel Memorial Prize in Economic Sciences), na área de econometria, são: Jan Tinbergen, Professor na Erasmus University Rotterdam, e Ragnar Frisch em 1969; Lawrence Klein, Professor de economia na University of Pennsylvania, em 1980; Trygve Haavelmo, em 1989, pelo seu artigo "The Probability Approach to Econometrics", publicado em 1944 em *Econometrica*; Daniel McFadden e James Heckman, da University of California, Berkeley, em 2000; e Robert Engle e Clive Granger, da University of California, San Diego, em 2003.

ARAÚJO, Carlos Alberto. Biblioimetria: Evolução histórica e questões atuais. *Em Questão*, v.12, n.1, jan/jun 2006, p.11-32. Disponível em: http://www.usp.br/siicusp/Resumos/17Siicusp/resumos/2244.pdf. Acesso em janeiro 2010.

ARAÚJO JÚNIOR. Rogério Henrique. *Precisão no processo de busca e recuperação da informação*. Brasília DF: Thesaurus, 2007.

ARCHUBY, César Aprendiendo Matemática con Bradford: propuesta metodológica de enseñanza del uso de los elementos básicos de la matemática en Bibliotecología y Ciencia de la Información (ByCI). *In*: VII Encuentro de la Asociación de Educadores e Investigadores de Bibliotecología, Archivología, Ciencias de la Información y Documentación de Iberoamérica y el Caribe. Marília SP, 22-24 nov 2006 (VII EDIBCIC). *Anais*. Marília SP: UNESP, 2006. Disponível, em: http://edibcic.org/VII_EDIBCIC/VIIEDIBCICProgramacion.htm. Acesso em outubro 2009.

ATHERTON, Pauline. Handbook fr information systems and services. Paris: UNESCO, 1977, 259 p. (*apud* Pinheiro, 2005)

BAKER, S.L.; LANCASTER, F.W. (1968b). *Information retrieval systems; characteristics, testing, and evaluation.* New York, Wiley.

BAR-ILAN, Judit. Informetrics at the beginning of te 21st century – A review. *Journal of Informetrics*, v.2, n.1, January 2008, p.1-52; DOI: 10.1016/j.joi.200711.001. Disponível em: http://www.citeulike.org/user/LEONRANGEL/article/2394624. Acesso em setembro 2009.

BARCELLOS, Silvia de Oliveira; KURAMOTO, Helio; LEIRO, Jayme. *Implemen tação do GERIR: um novo software para gerenciar informações referenciais*. Brasília: IBICT, 198-?. 30 p. Memória Técnica do IBICT; MT/0368.

BARTS, Nicolas. *Infocentre Recherche: Un système d'information, outil d'aide o pilotage de la recherche d'un établissement de recherche. Aplication à l'Université Paul Cézanne Aix-Marseille III*. Thèse pour obtenir le grade de Docteur de l'Université Paul Cézanne Aix-Marseille III, Faculté des Sciences et des Techniques (Discipline: Sciences de l'Information et de la Communication. 19 décembre 2008. 197 p. Disponível em: http://crrm.u-3mars.fr/web/IMG/pdf/Nicolas_Bark_T.pdf. Acesso: ago 2009.

BATES, M. The invisible substrate of information science. *Journal of the American Society for Information Science,* v. 50, n. 12, 1999, p. 1043–1050 [Electronic version]. (*Apud* Jayroe, 2008).

BAUDOIN,Lesya; HAEFFNER-CAVAILLON, Nicole; PINHAS, Nicole; MOUCHET, Suzy; KORDON, Claude. Indicateurs bibliométriques: Réalités, mythes et prospective. *Medecine/Sciences,* v.20, 2004, p.909-915. Disponível em: http://pdf2.edk.fr/archive/ms/2004/10/909-915.pdf. Acesso em setembro 2009.

BIREME. Interface WWWISIS, versão 5.0. 2000. Disponível em: http://www.bireme.br. *Apud* Robredo e Oliveira (2005).

BJÖRNERBORN, L.; INGWERSEN, P. Towards a basic framework for Webometrics. *Journal of the American Society for Information Science and Technology*, v.55, n.14, 2004, p. 1216-1227 [Electronic version].

BLISS Classification Association. The Bliss Bibliographic Classification: Outline of Bliss Classification. 2nd ed. 2007. Disponível em: http://www.blissclassification.org.uk/bcoutline.htm Acesso em stembro 2009.

BLISS, H.E. *The Organization of Knowledge in Libraries and the subject-approach to books*. New York: The H. W. Wilson Company, 1934.

BLISS, H.E. *A bibliographic classification.* Volumes. 1-4. [BC1]. New York: H. W. Wilson, 1950-1953 (*apud* HJørland, 2006).

BORGMAN, Christine L.; FURNER, Jonathan. Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, v.36, p. 4-59, 2002.

BRAMBILLA, S.D.S.; LAIPERT, R.C.F.; STUMPF, I.R. C.; CAREGNATO, S.E. Interfaces entre os campos da Comunicação e da Informação. *Comunicação & Informação* (UFG), v.10, p.21-33, 2007.

BRITO, Sandra Paula de. *Aperfeiçoamento da Base de Artigos de Periódicos Científicos das Áreas de Arquivologia, Biblioteconomia, Ciência da Informação e Documentação (ABCID): inclusão de palavras-chave e resumos nos registros da Revista de Biblioteconomia de Brasília*. 2006. Trabalho de Conclusão de Curso. (Graduação em Biblioteconomia) - Universidade de Brasília. Orientador: Jayme Leiro Vilan Filho.

BROOKES,B.C. The fundamental problem of information science. *In* V. Horsnell (Ed.), *Informatics 2*: Proceedings of a Conference Held by the ASLIB Coordinate Indexing Group. London: ASLIB, 1975, p. 42-49.

BRADFORD, Samuel C. Sources of information on specific subjects. *Engineering,* v.26, 1934, p. 85-86 (*apud* Hjørland, 2006).

BRAGA, Gilda Maria. *Relações bibliométricas entre a frente de pesquisa (research front) e revisões da literatura: estudo aplicado à ciência da informação*. Rio de Janeiro: IBBD, 1972.

BRAGA, Gilda Maria. Bibliometria: teoria e prática. *Ciência da Informação*, v. 16, n. 1, 1987, p. 103 – Recensões (FONSECA, Edson Nery da (org). *Bibliometria: teoria e práticaI: textos de Paul Otlet, Robert Estivais, Victor Zoltowski, Eugene Gartield.* São Paulo, Cultrix; Editora da USP, 1986. 141p.).

BRESSAN, Rodrigo Affonseca; LAFER, Beny; ROHDE, Luis Augusto Marcelo; FLECK, Pio de Almeida. The challenge of new RBP editors. *Revista Brasileira de Psiquiatria*, v.30, n.2, 2008 - Editorial; DOI: 10.1590/S1516-44462008000200001; *Print version* ISSN 1516-4446. Disponível em: http://www.scielo.br/scielo.php?pid=S1516-44462008000200001&script=sci_arttext&tlng=en. Acesso em setembro 2009.

BROOKES, B. C. The growth, utility, and obsolescence of scientific periodical literature. *Journal of Documentation*, London, v. 23, n. 4, p.283-294, Dec. 1970.

BROOKES, B.C. Developing cognitive viewpoint in information science: *In*: M. de Mey, R. Pinxten, M. Poriau, and F. Vandamme (eds.). *International Workshop on the Cognitive Viewpoint*, Belgium, 1977. Ghent, Belgium: University of Ghent, 1977, p. 195-203.

BROOKES, Bertram C.The Foundations of Information Science. Part i. Philosophical aspects. *Journal of Information Science*, v.2, .3-4, 1980, p. 125-133; Part ii. Quantitative aspects: classes of things and the challenge of human individuality. *Ibidem*, v.2, n.5, p. 209-221; Part III. Quantitative aspects: objective maps and subjective landscapes. *Ibidem*, v,2, n.6, p. 269-275; Part IIII. Information science: the changing paradigm. *Ibidem*, v.3, n.1, 1981, p.3-12.

BROOKES, B.C. Biblio-, Sciento-, Infor-metrics??? What are we talking about. *In*: L. Egghe and R. Rousseau (eds.). *Informetrics 89/90*. Proceedings of the 1st International Conference on Bibliometrics and Technical Aspects of Information Retrieval. Limburg, Belgium, 1989. Disponível em: http://doclib.uhasselt.be/dspace/bitstream/1942/857 ou, também, em: http://hdl.handle.net/1942/857. Acesso em agosto 2009.

BURTON, R.E.; KLEBER, R.W. The "half-life" of some scientific and technical literatures. *American Documentation,* v. 2, n. 1, p. 18-22, Jan. 1960.

CAFÉ, L.; BRASCHER, M. Organização da informação e bibliometria. *Encontros Bibli.* (UFSC), Número especial, 2008, p. 54-35.

CALLON, M.; COURTIAL, J.P.; LAVILLE, F. Co-word analysis as a tool for describing the network of interaction between basic and rechnological research: The case of polymer chemistry. *Scientometrics*, v.22, 1991, p. 155-205. ISBN 0138-9130 (Print) 1588-2861 (Online).

CARDOSO, João Luiz. *GIBI: Gerenciador Integrado de Biblioteca*, IBICT. Brasília: IBICT, 1988. Memória Técnica do IBICT - MT/0977 (Monografia de Graduação em Processamento de dados. UNB/ICE, 1988. 462 p. – Orientadores: Renato da Veiga Guadagnim e Jayme Leiro Vilan Filho).

CASTRO, Fábio de. A Question of quality. *Revista Brasileira de Cirurgia Cardiovascular*, v.24, n.2, 2009; DOI: 10.1590/S0102-76382009000200028; *Print version* ISSN 0102-7638.
Disponível em: http://www.scielo.br/scielo.php?pid=S0102-76382009000200028&script=sci_arttext&tlng=en. Acesso em setembro 2009.

CHASTINET, Y.S. *Lista de periódicos brasileiros na área de ciências agrícolas produtores de artigos em nível técnico-cienífico (core list)*. FAO, Roma (Italy): 1975, 15 p. Report No.: FAO-GIL – UNDP/BRA/72/020; FAO-GIL – DOC/TEC/75/025.

CHASTINET, Y.S.; FONSECA, A.F.M.; LOBO, P.R.A; ROBREDO, J. Análise da expansão do Serviço de Bibliografias Personalizdas em Agricultura (BIP / AGRI) – Um serviço brasileiro de disseminação seletiva da informação. *In*: International Seminar on Selective Disemination of Information, Ottawa, Ontário (Canadá), 2 out. 1978, 19 p.

CHASTINET, Y.S.; ROBREDO, J.; COUTINHO, M.E. de A.; PONCE, C. de A.; DINIZ, M. de F.P. Estabelecimento da lista básica de periódicos agrícolas através da análise crítica da dispersão da literatura. *In*: Worldwide International Association of Agricultural Librarians and Documentalists Congress, *Proceedings*...5, México City (México),14 Apr 1975. Publicado, também, em *AIBDA Boletín Técnico*, v.14, 1975, p.1-39.

CHEN, Y.; LEIMKUHLER, F. A relationship between Lotka's law, Bradford's law, and Zipf's law *Journal of the American Society for Information Science,* v.37, n. 5, 1986, p. 307–314. [Electronic version]. (*Apud* Jayroe, 2008).

COSTA, S.M.; BRASCHER, M. MADEIRA, F.; SCHIELSS. Tem Yeras of ElPub: An Analysis of its Major Trends. *In*: 10th International Conferenc on Electronic Publishung, held in Bansko, Bulgaria, 14- 18 Juune 2006, Proceedings..., 2006, p. 395-400.

COURTIAL, J.P.; CALLON, M. Indicators for the identification of strategic tehemes within a researc programme. *Scientometrics*, v.21, 1991, p. 447-458. ISBN 0138-9130 (Print) 1588-2861 (Online).

COURTIAL, J.P.; CALLON, M.; SIGOGNEAU, M. Is indexing trustworthy? Classification of articles through co-word analysis. *Journal of Information Science*, v.9, 1994, p.47-56.

CRAWFORD, Susan Y. Scientific Communication and the growth of Big Science. *In*: CRAWFORD, Susan Y. (org.) *From print to electronic*. [s.l.]: ASIS, 1996, p. 1-8.
CUTTER, C.A. *Rules for a Dictionary Catalog.* Washington, DC: Government Printing Office, 1876 (2.ed. 1889). (*Apud* Hjørland, 2005)
DANCHIN, Antoine. A Note on Bibliometrics. *Site of biological philosophy / Epistemology, ethics, history / Bibliometrics*. Disponível em: http://www.normalesup.org/~adanchin/site_map.hml#map, ou diretamente em: http://www.normalesup.org/~adanchin/bibliography/bibliometrics.html. Acesso em setembro 2009.
DE HAAN, J. Authorship patterns in Dutch sociology. *Scientometrics,* v.39, n.2, 1997, p.197-208.
DIODATO, Virgil. *Dictionary of bibliometrics*. Binghampton NY: The Haworth Press, 1994.
DUBIN, David. The Most Influential Paper Gerard Salton Never Wrote**.** *Library Trends,* v. 52, 4, 2004, p. 748–764.
DOBROV, G. M. (Gennadii Mikhailovich) *Wissenschafts Wissenschaft : Einfuhrung in die Allgemeine Wissenschaftswissenschaft / G. M. Dobrov; herausgegeben und mit einem Vorwort verseben von Dr. G?Lotz.* Berlin : Akademie-Verlag, 1969. Q158 D53 LAK. Normal loan.
DOBROV, G.M.; KARENNOI, A.A. The informational basis of scientometrics. Mikhailov, A.I. *et al*. (eds.) *On theoretical problems of informatics*. VINITI. FID 435, 1969, p. 165-191 (*apud* BROOKES, 1989)
DOU, Henri. Inteligência Competitiva e Tecnologia (*Intelligence compétitive et technologie*). Palestra. *In*: III Workshop Internacional de Ciência da Informação (III WICI). Brasília DF, 1-2 setembro 2008.
EGGHE, Léo; ROUSSEAU, Ronald. *Introduction to informetrics*. Elsevier, Amsterdam, 1990. ISBN 0-444-88493-9.
EGGHE, Léo; ROUSSEAU, Ronald. *Elementary Statistics for effective Library and Information Management*. London (UK): Aslib, 2001. ISBN 0-8514-2451-1.
ESTIVALS, R. La statistique bibliographique. *Bulletin des Bibliothèques de France*, v.14, n.12, 1969, p.481-502. Disponível em: http://ec3.ugr.es/publicaciones/Jimenez_Contreras,E;_Responde_Surface_Methodology_and_its_Application_in_Evaluating_Scientific_activity.pdf. Acesso em em agosto 2009.
FAO – United Nations Organization for Food and Agriculture. *FAO Online Catalogues* [Entrada Robredo, J]. (1972 ->). Disponível em: http://www4.fao.org/cgi-bin/faobib.exe?vq_query=A%3DRobredo,%20J.&database=faobib&search_type=view_query_search&table=mona&page_header=ephmon&lang=eng. Acesso em agosto 2009.
FERRAZ, Valéria Cristina Trindade; AMADEI, José Roberto Plácido; SANTOS, Carlos Ferreira. The evolution of the Journal of Applied Oral Science: a bibliometric analysis. *Journal of Applied Oral Science*, v.16, n.6, 2008. DOI: 10.1590/S1678-77572008000600012. *Print version* ISSN 1678-7757. Disponível em: http://www.scielo.br/pdf/jaos/v16n6/a12v16n6.pdf. Acesso em setembro 2009.
FERREIRA, Maria Mary. O profissional da informação no mundo do trabalho e as relações de gênero. *Transinformação*, v.15, n.2, mai/ago 2003, p.189-201.
FID – FEDERATION Internationale de Documentation. *Extension de la Classification Décimale Universelle dans le domaine du verre.* FID Publication 350. Charleroi, Belgique : Institut National du Verre, 1984, 102 p.
FIGUEIREDO, L.M. de. *Distribuição da literatura geológica brasileira: estudo bibliométrico*. Rio de Janeiro, CNPq/IBBD-UFRJ, 1972. 120p. Dissertação de Mestrado.
FONSECA, Edson Nery da. Bibliografia Estatística e Bibliometria: Uma Reivindicação de Prioridades. *Ciência da Informação*, v.2, n.1, 1973, p.5-7. Disponível em: http://revista.ibict.br/index.php/ciinf/article/viewFile/1625/1234. Acesso em setembro 2009.
FONSECA, Edson Nery da. (org.) *Bibliometria:* teoria e prática; textos de Paul Otlet, Robert Estivais, Victor Zoltowski, Eugene Gartield. São Paulo, Cultrix; Editora da USP, 1986. 141p.
FUJINO, Asa; NORONHA, Daisy Pires; POBLACIÓN, Dinah A de Mello Aguiar; SILVA, José Fernando Modesto da. Comunicação e produção científica: avaliação e perspectivas. *In*: Marilda Lopes Ginez de lara; Asa Fujino; Daisy Pires Noronha. (orgs.). Informação e Contemporaneidade: perspectivas. Recife: NÉCTAR: PPGCI, 2007, p.199-223.
GARFIELD, E.; SMALL, H. Michael J. MORAVCSIK: Multidimensional scholar and hero of Third World science. *Scientometrics*, v.20, n.1, 1991, p.19-23. ISSN 0138-9130 (Print) 1588-2661 (Online).

GARFIELD, E. "Science Citation Index": A New Dimension in Indexing. *Science,* v.144, n.3619, May 8 1964, p.649-654. Reprinted by *Essays of an Information Scientist*. v.7, 1984, p.525-535. Disponível em: http://garfield.library.upenn/essays/v7p525y1984.pdf. Acesso em janeiro de 2010.
GLOTTOMETRICS, 2000, vols. 3, 4, 5. Home page and Table of contents: Disponível em: http://www.ram-verlag.de/journal.htm. Acesso em setembro 2009.
GORKOVA, Valentina I. *Informetrics*. Moscow: VINITI, 1988. *Informatics* 10 (*apud* BROOKES, 1989).
GOSWANI, Prashant; SHARMA, Umesh; SHUKIA, Anil Kumar. The Webometrics. *International CALIBER*-2008, p.656-660. Disponível em: http://ir.inflibnet.ac.in:8080/jspui/bitstream/123456789/520/1/CALIBER%202008(70).pdf.
GOSNELL, Charles F. *The Rate of Obsolescence in College Library Book Collection*. PhD Dissertation. New York University, 1943.
GOSNELL, Charles F. Obsolescence of Books in College Libraries. College Res. Libr., v.5, n.2, 1944, p. 116
HAYASHI, M.C.P.I. ; FERREIRA JÚNIOR, A. ; BITTAR, M. ; HAYASHI, C.R.M.; SILVA, M.R. da. História da Educação: a produção científica na biblioteca eletrônica Scielo. *Educação e Sociedade*, v. 29, p. 181-211, 2008.
HAYASHI, Maria Cristina Piumbato Innocentini; CABRERO, Rodrigo de Castro; COSTA Maria da Piedade Resende da; HAYASHI, Carlos Roberto Massao. Indicadores da participação feminina em Ciência e Tecnologia. *Transinformação*, v.19, n.2, mai/ago 2007, p.169-187.
HERTZEL, D.H. Bibliometrics History. *In* : Drake, M. (ed.), *Encyclopedia of library and information science.* New York, NY: Marcel Dekker, Inc., 2003, p. 288–328. (*Apud* JAYROE, 2008)
HINKLE, Karyn. *"another unusual figure in the history of computer science": George Kingsley Zipf, 1902-1950.* (2006 ?). Disponível em: http://prattsils.wikispaces.com/page/pdf/karynh. Acesso em setembro 2009.
HJØRLAND, Birger. *Core Concepts in Library and Information Science (LIS)*, 2005. Disponível em: http://www.db.dk/bh/Core%20Concepts%20in%20LIS/articles%20a-z/wersig.htm. Acesso em agosto 2009.
HJØRLAND, Birger. *Core Concepts in Library and Information Science (LIS)*, 2006. Disponível em: http://www.db.dk/bh/Core%20Concepts%20in%20LIS/Clusters%20of%20LIS/information_science_biography.htm. Acesso em agosto 2009.
HJØRLAND, Birger. *Core Concepts in Library and Information Science (LIS)*, 2006. Disponível: http://www.db.dk/bh/Core%20Concepts%20in%20LIS/articles%20a-z/vickery.htm.
HJØRLAND, Birger. *Bradford, Samuel C. (1878-1948).* Disponível em: http://www.db.dk/bh/core%20concepts%20in%20lis/articles%20a-z/bradford.htm. Acesso em setembro 2009.
HOOD, William W.; WILSON, Conception S. The Literature of Bibliometrics, Scientometrics and Informetrics. *Scientometris*, v.52, n.2, 2004, p. 291-314. ISSN 0138-9130 (Print) 1588-2661 (Online).
IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia. Formato IBICT. Brasília DF: CNPq/IBICT, 1984.
IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia. *Diretrizes para Elaboração de Tesauros Monolíngües*. Brasília DF: IBICT, 1984. Projeto Coordenado por Hagar Espanha Gomes.
IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia. Programa de elaboração de tesauros em microcomputador (TECER). Brasília DF: IBICT, 1989, 76 p.
IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia; CNEN-CIN – Comissão Nacional de Energia Nuclear / Centro de Informações Nucleares. LINCE: uma proposta de padronização para linguagens de acesso a bases de dados. *Revista de Biblioteconomia de Brasília*, v.18, n.2, jul/dez 1990, p.203-211. Disponível em: http://164.41.122.25/portalnesp/ojs-2.1.1/index.php/RBB/article/view/651/648. Acesso em outubro 2009.
ICG – International Commission on Glass. *Home.* (Last update: 21.10.2009). Disponível em: http://www.icg.group.shef.ac.uk/ Acesso em out 2009.
ICG – International Commission on Glass. *Dictionary of Glass-Making* : Dictionnaire de verrerie; Glas-Fachwörterbuch. Charleroi, Belgium: Secretariat of Subcommittee A1 c/o Institut National de Verre. 1965, 233p.
ICG – International Commission on Glass. *Dictionary of Glass-Making* : Dizionario di terminologia vetraria (Suplemento italiano). Charleroi, Belgium: Secretariat of Subcommittee A1 c/o Institut National de Verre, 1967a. 97p.

ICG – International Commission on Glass. *Dictionary of Glass-Making* : Vocabulário vitrotenológico. (Suplemento español). Charleroi, Belgium: Secretariat of Subcommittee A1 c/o Institut National de Verre, 1967b. 97p.
ICG – International Commission on Glass. *Dictionary of Glass-Making* : Glastechnische woordenlijst (Nederlands supplement). Charleroi, Belgium: Secretariat of Subcommittee A1 c/o Institut National de Verre, 1968a, 95p.
ICG – International Commission on Glass. *Dictionary of Glass-Making* : Glasteknick ordbok (Swenskt supl.) Charleroi, Belgium: Secretariat of Subcommittee A1 c/o Institut National de Verre, 1968b, 87s.
ICG – International Commission on Glass. *Dictionary of Glass-Making* : Glas-Fachwörterbuch; Slovnik Szklarski. Warszawa: Sekretariat Polskiej Grupy Roboczej. Instytut Szkla, 1974, 241p.
ICG – International Commission on Glass. *Dictionary of Glass-Making* : Dictionnaire de verrerie; Glas-Fachwörterbuch (Japanese Edition). [s.l. : s.e. , s.d.].
ICG – International Commission on Glass. *Dictionary of Glass-Making* : Odborý Sklarský Stekla; Словарь по Технлогии Стекла [s.l.: s.e. , s.d.].
ICG – International Commission on Glass. *Dictionary of Glass-Making* : In English, French and German. Amsterdam: Elsevier Science, 1983, 424 p. ISBN-10: 0-444-42048-7; ISBN-13: 9780444420480. Disponível em: http://www.book-info/isbn/0-444-42048-7.htm.
ICG – International Commission on Glass. *Dictionary of Glass-Making* : In English, Spanish and Portuguese. Amsterdam: Elsevier Science, 1992, 372p. (In collaboration with the European Community and UNESCO). ISBN-10: 0-444-89622-8; ISBN-13: 9780444896223. Disponível em: http://www.book-info/isbn/0-444-89622-8.htm.
INDIGOSTAR Software. MicroWeb. 2002. Disponível em: http://www.indigostar.com. *Apud* Robredo e Oliveira (2005).
INGWERSEN, Peter. *Information Retrieval Interaction*. London: Taylor Graham, 1992. 246 p. ISBN 0-947568-54-9.
INGWERSEN, Peter; JÄRVELIN, Kalervo. *The Turn: Integration of Information Seeking and Retrieval in Context.* Dordrecht, The Netherlands: Springer, 2005. 472 p. ISBN 1-4020-3850-X.
INSTITUT DU VERRE. 'Thésaurus verrier'. Paris : Institut du Verre, 1971 – 2 vol. *In*: Guéniot, Yvonne (analyste). *Bulletin de Documentation Bibliographique*, 2e Partie. Analyses d'ouvrages et d'articles français et étrangers préparées par la Direction chargée des Bibliothèques et de la Lecture Publique, 1972, p.26-27 (Analyse 163). Disponível em: http://bbf.enssib.fr/consulter/bbf-1972-01-0013-001.pdf e também em: http://74.125.47.132/search?q=cache:4n0IoK1jQiUJ:bbf.enssib.fr/consulter/bbf-1972-01-0013-001.pdf+institut-du-verre+bulletin-de-documentation-bibliographique+th%C3%A9saurus+OR+Gu%C3%A9niot+%22sch%C3%A9ma+fl%C3%A9ch%C3%A9+%22&cd=1&hl=pt-BR&ct=clnk&gl=br&lr=lang_fr. Acesso em outubro 2009.
IPGA – International Pool of Glass Abstracts. *A propos de l'IPGA*. Disponível em: http://www.verreonline.fr/ipga/index.php. Acesso em setembro 2009.
IPGA – International Pool of Glass Abstracts. *Glass file (Base de données verrière)*. DATA. Disponível em: http://www.verreonline.fr/ipga/glassfile.ipga/glassfile.php. Acesso em setembro 2009.
ISO – International Organization for Standardization. ISO 2709:2008 – *Information and documentation – Format for information exchange*. 2008. Disponivel em: http://www.iso.org/iso/catalogue_detail.htm?csnumber=41319. Acesso em out 2009.
JANSEN, B.J.; SPINK, A.; SARACEVIC, T. Failure analysis in query construction: Data and analysis from a large sample of Web queries. Proceedings of the 3rd ACM Digital Libraries Conference, 289-290. 1998, Disponível em: http://comminfo.rutgers.edu/~tefko/ProcDL1998.pdf. Acesso em setembro 2009.
JANSSENS, Frizo; GLÄNZEL, Wolfgang; DE MOOR, Bart. A hybrid mapping of information science. *Scientometrics,* v.75, n.3, 2009, p.607-631. DOI: 10.1007/s11192-007-2002-7. Disponível em: ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/frizo/reports/Janssens_F_Glanzel_W_DeMoor_B_AHybridMappongOfIS.pdf. Acesso em setembro 2009.
JAYROE, Tina. *Bibliometrics for Dummies*. [s.l.] : Shimelis G. Assefa, 2008 (Apostila). Disponível em: https://portfolio.du.edu/portfolio/getportfoliofile?uid=122400. Acesso em em setembro 2009.
JOY MOYER, L. Bibliography of publications of Belver C. Griffith. *Scientometrics*, v. 51, n. 3, 2002, p. 469-479. ISSN 0138-9130 (Print) 1588-2661 (Online). Disponível em: http://www.ingentaconnect.com/content/klu/scie/2002/00000051/00000003/00400599. Acesso em agosto 2009.
KATZ, J.S.; MARTIN, B.R. What is research collaboration? *Research Policy*, v.26, p.1–18, 1997.

KAVATRA, P.S. *Texbook of Information Science*. New Delhi: A.P.H. Publishing Corporation, 2000, 323 p. ISBN 81-76-48-163-7.
KUHN, Thomas S. Reflexões sobre meus críticos. *In*: LAKATOS, Imre; MUSGRAVE, Alan (Orgs.). *A crítica e o desenvolvimento do conhecimento*. São Paulo: Editora de Universidade de São Paulo, 1979. p.285-343.
LANCASTER, F.W.; FAYEN, E.G. *Information Retrieval On-Line*. Los Angeles: Melville Pub. Co. 1973.
LANCASTER, F.W.; SMITH, L.C. *Compatibility issues affecting information systems and services. Prepared for the General Information Programme and UNISIST*. Paris: United Nations Educational, Scientific and Cultural Organization. 1983.
LANCASTER, F.W. *Toward Paperless Information Systems*. New York: Academic Press. 1978.
LANCASTER, F.W. L*ibraries and Librarians in an Age of Electronics.* Arlington, Va.: Information Resources Press. 1982.
LANCASTER, F.W. (ed.). *Libraries and the future; essays on the library in the twenty-first century.* New York : Haworth Press. 1993.
LANCASTER, F.W. *The Measurement and Evaluation of Library Services.* 2nd ed. Arlington, Va.: Information Resources Press. 1977. (2nd edit. 1991).
LAUREILHE, Marie-Thérèse. 'Troisième complément à la bibliographie des thesauri et index par matières parus depuis 1960 (au 31 décembre 1971)'. *Bulletin des Bibliothèques de France – BBF.* Paris, 1972, t. 17, n° 2 - 6. Sciences appliquées, médecine, technique. INSTITUT DU VERRE. Paris. - Thesaurus Verrier...1972. Disponível em: http://bbf.enssib.fr/consulter/bbf-1972-02-0067-002. Acesso em out 2009.
LE COADIC, Yves François. *La science de l'information*. Paris: Presses Universitaires de France, 1994. (Coll. '*Que sais-je'*, 2873). Disponível edição brasileira: *A Ciência da Informação*. 2ª ed. rev. atual., Brasília DF: Briquet de Lemos / Livros, 2004, 124 p.
LETA, J.; LANNES, D.; MEIS, L. A formação de recursos humanos e a produção científica no Brasil. In: PALATNIK, Marcos *et al.* (org.). *A pós-graduação no Brasil*. Rio de Janeiro: Lidador, 1997, p. 99-108.
LEYDESDORFF, Loet; MEYER, Martin. The Scientometrics of a Triple Helix of University-Industry-Government Relations (Introduction to the Topical Issue), *Scientometrics*, v.70, n.2, 2007, p.207-222. ISBN 0138-9130 (Print) 1588-2861 (Online). Disponível em: http://users.fmg.uva.nl/lleydesdorff/sci06th5/index.htm. Acesso em agosto 2009.
LIMA, R.A.; VELHO, L.M.L.S. Indicadores ibero-americanos de atividade científica em bioprospecção. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 6, p. 1-14, 2008.
LÓPEZ, Maricela. La Cibermetria, uma Nueva Alternativa para Evaluar la Visibilidad de la Publicación Académica Electrónica. El caso de la REDIE, *Razón y Palabra*, n.58, Agosto-Septiembre 2007. Disponível em: http://www,razonypalabra.org.mx/anteriores/n58/mlopez.html.
MAPA – Ministério da Agricultura, Pecuária e Abastecimento/BINAGRI, 2009). Disponível em: http://www.agricultura.gov.br/portal/page?_pageid=33,959024&_dad=portal&_schema=PORTAL. Acesso em agosto 2009.
MARCOVITCH, J.; OHAYON, P. Évaluation en science et technologie au Brésil. In: COLLOQUE INTERNATIONAL METHODOLOGIES ÉVALUATIVES DE LA RECHERCHE : *Les Programmes de Recherche*, 1. 1985, Paris. *Anais***...** Paris : CPE-MRT, 1984. v. 1. p. 63-79.
MARQUES, Fabrício. Taking a breather: Incongruous results in two different rankings open up the debate about the limits of growth in Brazilian academic production. *Pesquisa FAPESP Online*. Print edition 150 - August, 2008. (Policies/Indicators).Disponível em: http://revistapesquisa.fapesp.br/?art=2252&bd=1&pg=1&lg=en. Acesso em setembro 2009.
MARTIN, B.R.; IRVINE, J., *Research Foresight: Priority-Setting in Science. London and New York:* Printer Publishers, 1988. Publicado, também, como IRVINE, J.; MARTIN, B.R. *Research Foresight: Creating the Future*. Zoetermeer: Netherlands Ministry of Education and Science. 1989 (*apud* PAVITT, 1997).
MARTYN, J.; LANCASTER, F.W. *Investigative methods in library and information science; an introduction.* Arlington, Va.: Information Resources Press. 1981.
McCAIN, K.W.; SALVUCCI, L.J. How influential is Brooks' Law: A longitudinal citation context analysis of Frederick Brooks 'The Mythical Man-Month. *Journal of Information Science,* v. 32, 2006, p. 277-295.
McCAIN, K.W. Neural Networks Research in Context: A Longitudinal Journal Cocitation Analysis of an Emerging Interdisciplinary Field. *Scientometrics*, v.41, n. 3, 1998, p. 389-410. ISBN 0138-9130 (Print) 1588-2861 (Online).
McCAIN, K.W. Sharing Digitized Research-Related Information on the World Wide Web. *Journal of the American Society for Information Science,* v. 51, n. 14, 2000, p.

1321-1327 (Invited paper).
MDIC/STI – Ministério de Desenvolvimento, Indústria e Comércio Exterior/Secretaria de Tecnologia Industrial. *Cronologia do desenvolvimento científico e tecnológico brasileiro 1950-2000.* Brasília DF: MDIC/STI; CNI; SEBRAE, 2002, 413 p. (*apud* Pinheiro, 2005).
MEADOWS, A.J. *A comunicação científica*. Brasília: Briquet de Lemos, 1999.
MEIS, L.; LETA, J. *O perfil da ciência brasileira*. Rio de Janeiro: Editora UFRJ, 1996.
MENEGHINI, R. Indicadores alternativos de avaliação de desempenho científico: aplicação a uma instituição. *Ciência e Cultura*, São Paulo, v. 40, p. 548-558, 1988.
MENEGHINI, R. The key role of colaborative work in the growth of braziliana science in the last ten years. Scientometrics. v.35, n.3, 1996, p.367-373.
MENEGHINI, R.; LETA, J. Produção científica. *In*: LANDI, F.R. (org.). *Indicadores de ciência, tecnologia e inovação tecnológica*. São Paulo: FAPESP, 2002, p. 3-21.
MENEGHINI, Rogerio; PACKER, Abel L. Is there science beyond English? Initiatives to increase the quality and visibility of non-english publications might help to break down language barriers in scientific communication. *EMBO Reports,* v.8, n2, 2007, 112–116 (View point). DOI:10.1038/SJ.EMBOR.7400906. Disponível em: http://www.nature.com/embor/journal/v8/n2/pdf/400906.pdf. Acesso em setembre 2009.
MERTON, Robert K. The Matthew Effect in Science. *Science,* v.5, n. 3810, 5 January 1968, pp. 56-63. DOI: 10.1126/science.159.3810.56. Disponível em: http://www.sciencemag.org/cgi/content/abstract/159/3810/56. Acesso em agosto 2009.
MIKHAILOV, A.I.; GILYAREVSKII, R.S. *An introductory course on informatics/documentation*. Revised and enlarged edition. The Hague: FID, 1971.
MIKHAILOV, A.I. Scientific and technical information and effectiveness of science. Rivista dell'Informazione, v. 3, n. 1, 1972, p. 74-77.
MIKHAILOV, A.I. (ed.). *Theoretical problems of informatics: forecasting of the development of scientific information activities*. Moscow: International Federation for Documentation, 1979 (FID 563).
MOED, Henk, F. *Moed, H.F. – Contact Information.* Universiteit Leiden, Centre for Science and Technology Studies., 2008. Disponível em: http://www.socialsciences.leiden.edu/cwts/staff/moedhf.html Acesso em agosto 2009.
MORAES, M.F.; OHAYON, P.; LEITAO, D.M. Indicadores de avaliação no CENPES / Petrobras: situação atual e perspectivas. *Revista de Administração*, São Paulo, v. 24, n. 2, p. 122-128, 1989.
MUELLER, S.P.M. O estudo do tema comunicação científica e tecnológica no Brasil: tendências e perspectivas na área de Ciência da Informação. Tendências da Pesquisa Brasileira em Ciência da Informação, v.1, p. n.7, 2008a.
MUELLER, S.P.M. Métricas para a ciência e tecnologia e o financiamento da pesquisa: algumas reflexões. *Encontros Bibli* (UFSC), v. especial, p. 2, 2008b.
MUELLER, S.P.M. Metodologia para Avaliacãao de Listas Básicas de Periódicos. *Ciência da Informação*, Brasília, v. 20, n. 2, p. 111-118, 1991.
MUELLER, S.P.M.; PASSOS, E.J. (Orgs.). *Comunicação científica*. Brasília DF: Ciencia da Informação, 2000. 144p.
MUELLER, Suzana Pinheiro Machado; CAMPELLO, Bernadete Santos; DIAS, Eduardo Wense. Disseminação da pesquisa em ciência da informação e biblioteconomia no Brasil. *Ciência da Informação*, v.25, n3, 1996.
MUGNAINI, R.; PACKER, A. L.; MENEGHINI, R. Comparison of scientists of the Brazilian Academy of Sciences and of the National Academy of Sciences of the USA on the basis of the h-index. *Brazilian Journal of Medical and Biological Research*, Ribeirão Preto, v. 41, n.4, p. 258-262, 2008.
MURRAY, Robin R.B. Literature of the Science and Technology of Ceramics, Including Enamels and Glass. *Literature of Chemical Technology*, Chapter 5, p.71-104. Chapter DOI: 10.1021/ba-1968-0078.ch005. *Advances in Chemistry*, v.78, 1968. ISBN 13: 9780841200791 e ISBN: 9780841222588. Disponível em: http://pubs.acs.org/doi/abs/10.1021/ba-1968-0078.ch005. Acesso em setembro 2009.
NALIMOV, V.V.; MULCHENKO, Z.M. *Naukometriya.* Moscow: Nauka,1969 (*apud* BRAUN, 2001).
NARIN, Francis. Technology Indicators and Corporate Strategy. *Review of Business*, v. 14, 1993. Disponível em: http://www.questia.com/googleScholar.qst;jsessionid=KhgRDg12TLY6XlKZyJJgmv8vc4SDrfQ1MbdQX01RM5zHqNJ2nJdz!846852835!-1813158499?docId=5000197307. Acesso em agosto 2009.
NEELAMEGHAN, A. Libraries and Information Services in Third World Countries. *In*: F.W. Lancaster (ed.). *Libraries and the Future: Essays on the Library of the Twenty-first Century*. New York: Haworth, 1993, p.85-106.

NEELAMEGHAN, A. International Cooperation and Assistance. *In*: Yves Courrier (General Editor). *World Information Report 1997-1998*. Paris: UNESCO, 1997. ISBN 92-3-1033411-7. Chapter 7, p. 361-380. http://unesdoc.unesco.org/images/0010/001062/106315e.pdf. Acesso em agosto 2009.

OCLC. *How one library pioneer profoundly influenced modern librarianship.* 2009. Disponível em: http://www.oclc.org/americalatina/pt/dewey/resources/biography/default.htm. Acesso em setembro 2009.

ODDONE, Nanci. O IBBD e a informação científica: uma perspectiva histórica para a ciência da informação no Brasil. *Ciência da Informação*, v.35, n.1, 2006; DOI: 10.1590/S0100-19652006000100006; *Print version* ISSN 0100-1965. Disponível em: http://www.scielo.br/scielo.php?pid=S0100-19652006000100006&script=sci_arttext. Acesso em: agosto 2009.

OHAYON, P.; ROSENBERG, G. Avaliação de Programas de Pesquisa, Desenvolvimento Tecnológico e Inovação: modelo proposto para o PADCT/MCT. In: AFONSO, A.J.; MELO, M.M. (orgs.). *Avaliação na Educação*. Pinhais: Ed. Melo, 2007, p. 177-186.

OSAREH, Farideh. Bibliometrics, Citation Analysis and Co-Citation Analysis: A Review of Literature I. *Libri*, v. 46, 1996, p.149-158; *ibidem*, II, v. 46, 1996, p. 217-225 (*apud* UTEXAS [s.d.]).

OTLET, Paul. *Traité de documentation: le livre sur le livre, théorie et pratique*. Reimpression de l'édition de 1934. Liège: Centre de Lecture de la Communauté Française de Belgique, 1989

PAVITT, Keith, Comments on John Irvine and Ben B. Martin, Recipients of the 1997 Derek de Solla Price Award, *Scientometrics,* v.4o, n.3, 1997, p. 363-366. ISBN 0138-9130 (Print) 1588-2861 (Online). Disponível em: http://www.garfield.library.upenn.edu/smetrics/irvine_martin.pdf. Acesso em agosto 2009.

PINHEIRO, L.V.R. Infra-estrutura da pesquisa em Ciência da Informação. *Datagramazero.* Rio de Janeiro RJ, v. 1, n. 6, 2000, p. 2.

PINHEIRO, L.V.R.; LOUREIRO, J.M.M. Traçados e limites da Ciência da Informação. *Ciência da Informação*, Brasília, v. 24, n. 1, p. 42-53, jan./jul. 1995. Disponível em: http://revista.ibict.br/index.php/ciinf Acesso em: 17 set. 2007.

PINHEIRO, L.V.R.; LOUREIRO, J.M.M. . Políticas públicas de C&T, ICT e de pós-graduação e o surgimento da Ciência da Informação. In: V Encontro Nacional de Ciência da Informação (V CINFORM), 2004, Salvador BA. *Anais do V Encontro Nacional de Ciência da Informação (V CINFORM)*, 2004. p. 187-202. Disponível em: http://www.cinform.ufba.br/v_anais/artigos/vaniajose.html.

PINHEIRO, Lena Vânia Ribeiro; BRASCHER, Marisa; BURNIER, Sonia. Ciência da Informação: 32 anos (1972-2004) no caminho da história e horizontes de um periódico científico brasileiro. *Ciência da Informação,* v. 34, n. 3, 2005

PINHEIRO, L.V.R.; BRASCHER, M.; BURNIER, S.; MORAES, J.N.L. de. Os impactos das redes eletrônicas nas metodologias e modelos para análise de periódicos na comunicação científica. *In*: Conferência Ibero-Americana de Publicações Eletrônicas no Contexto da Comunicação Científica – CIPECC 2006. Brasília DF, *Anais...* Campo Grande: UNIDERP, 2006, p. 129-134.

PINHEIRO, Lena Vania Ribeiro. Evolução e tendências da ciência da informação, no exterior e no Brasil: Quadro comparativo a partir de pesquisas históricas e empíricas, *In:* Encontro Nacional de Pesquisa em Ciência da Informação – VI ENANCIB, 2005, Florianópolis, SC.

PINHEIRO, Lena Vania Ribeiro; SILVA, Giselle Santos. Cartografia histórica e conceitual da bibliometria/informetria no Brasil. *In:* Conferência Ibero-Americana de Publicações Eletrônicas no Contexto da Comunicação Científica – II CIPECC, Rio de Janeiro RJ, 17-21 nov 2008. Sub-Tema 2 – Metrias da comunicação científica: da bibliometria/informetria à Webmetria.

PIRES, Marina Melo. *Agrupamento incremental e hierárquico de documentos*. Rio de Janeiro RJ: Universidade Federal de Rio de Janeiro, 2008 (Dissertação de mestrado em em Engenharia Civil). Disponível em: http://wwwp.coc.ufrj.br/teses/mestrado/Novas_2008/teses/PIRES_MM_08_t_M_int.pdf. Acesso em em out 2009.

POBLACIÓN, Dinah Aguiar. *Análise quantitativa da produção cientifica do corpo docente da área da saúde da USP – Campus de São Paulo: 1980-1983*. 1ª ed. São Paulo: ECA/USP, 1986. v. 2.

POBLACIÓN, Dinah Aguiar. Visibilidade da produção científica gerada pelos docente e egressos dos Programas de Pós-graduação em Ciência da Informação e as interfaces com os Grupos de Pesquisa da área, constantes do Diretório do CNPq. *Datagramazero*, Rio de Janeiro, v. 6, n. 1, p. 1-10, 2005.

POBLACIÓN, Dinah Aguiar; WITTER, Geraldina Porto; SILVA, José Fernando Modesto da. (orgs.). *Comunicação e Produção Científica: contexto, indicadores e avaliação*. São Paulo: Angellara, 2006.
POLANCO, X. Scientometric analysis of the cognitive sciences in PASCAL. *INIST Info*, n.7, jul 1993.
POLANCO, X. Aux sources de la scientométrie (*Solaris,* nº 2, "Les sciences de l'information: bibliométrie, scientométrie, infométrie". Presses Universitaires de Rennes, 1995 (ISSN: 1265-4876), p. 13-79.) Disponível em: http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/2polanco1.html. Acesso em agosto 2009.
POLANCO, X.; GRIVEL, L.; FRANÇOIS, C.; BESAGNI,D. L'infométrie, un programme de recherche. *Journées d'études sur les systèmes d'information élaborée* de la SFBA, Île Rousse, Corse, 1993. Document n° 3 des Actes, 9p. (*apud* ROBREDO, 2000)
POLANCO, Xavier. *Experiencia del INIST en la producción de indicadores infométricos: Indicadores y Vigilancia científica y tecnológica,* Observatorio Colombiano de Ciencia y Tecnología, Bogotá, 2001, 8p.
POTTER, William Gray. " 'Of Making Many Books There is No End': Bibliometrics and Libraries." *The Journal of Academic Librarianship* v.14, September 1988), p.238a-238c (insert between 238 and 239).
PRICE, Derek J. de Solla. *Little Science, Big Science*. New York and London: Columbia University Press, 1963, 118 p. [Trad. Brasileira: *O desenvolvimento da ciência: análise histórica, filosófica, sociológica e econômica*. Rio de Janeiro: Livros Técnicos e Científicos, 1976. 98 p.].
PRICE, Derek J. de Solla. The Structures of Publication in Science and Tecnology*. In*: H. Gruber e D.G. Marquis (eds.) *Factors in the Transfer of Tecnology*. Cambridge, Mass: The MIT Press, 1969, p. 91-104.
PRICE, Derek J. de Solla. Networks of scientific papers. *Science*, v.149, 1965: p. 510-515.
PRICE, Derek J. de Solla. A general theory of bibliometrics and other cumulative advantage processes. *Journal of the American Society for Information Science*, v.27, n.1, 1976. p. 73-74.
PRITCHARD, A. Statistical Bibliography or Bibliometrcs. *Journal of Documentation*, v.25, n.4. 1969, p.348-349.
QUONIAM, Luc; TARAPANOFF, Kira; ARAÚJO JÚNIOR, Rogério Henrique; ALVARES, Lilian. Inteligência obtida pela aplicação de 'data mining' em base de teses francesas sobre o Brasil. *Ciência da Informação*, v.20, n.2, mai/jun 2001, p.20-28.
RABER, D. *The problem of information: An introduction to information science*. Lanham, MD: Scarecrow Press, Inc., 2003. (*apud* JAYROE, 2008).
RAJAN, T.N. *Curriculum Development in the Changing Context of Information: INSDOC Experience.* 1983. ERIC #: ED254239. Microfiche.
RAJAN, T.N. *FID Informetrics Committee (FID/IM).* International Forum for Information and Documentation, 11, 1986. *Apud* VALDERRAMA ZURIÁN *et al*., 1986.
RAVICHANDRA RAO, I.K. *Métodos Quantitativos em Biblioteconomia e Ciência da Informação.* Brasília DF: Associação dos Bibliotecários do Distrito Federal (ABDF), 1986. 270 p. ISBN 85-230-0334-0 (Tradução do original inglês: *Quantitative Methods for Library and Information Science.* New Delhi, Bangalore, Bombay, Calcutta: Wiley Eastern Limited, 1983, 272 p. ISBN 0 85226 7495).
REIS, Luciana Monteiro de Barros. *Periódicos científicos brasileiros das áreas de informação (1972-2007): representatividade das áreas do conhecimento nos artigos*. 2009. Trabalho de Conclusão de Curso. (Graduação em Biblioteconomia) - Universidade de Brasília. Orientador: Jayme Leiro Vilan Filho.
ROBREDO, J. *La documentation scientifique e technique en tant qu'outil pour une prospective en matière de rechereche*. Paris, Institut du Verre, 1970.
ROBREDO, J. La dispersion des informations dans la litterature verrière. *Verres et Réfractaires*, Paris, France, v.27, n.3, 1973a, p.117-140.
ROBREDO, J. Analyse critique de l'indexage en fonction de l'efficacité de la recupération de l'information. *ANRT Information et Documentation*, Paris, France, v.5, n.11, 1973b, p.3-36.
ROBREDO, Jaime. Infometria e Ciência da Informação. *In*: Nídia M.L. Lubisco; Lídia M. B. Brandão. (orgs.) *Informação & Informática*. Salvador BA: EDUFBA, 2000, 307 p. (pp. 81-118).
ROBREDO, Jaime. *Documentação de hoje e de amanhã*. 4ª ed. revista e ampliada. Brasília DF: Edição de autor, 2005, 410 p. ISBN 85-905920-1-4.
ROBREDO, Jaime. Utilização de técnicas infométricas na organização de bancos de conhecimento, com base em ontologias, visando a gestão otimizada das políticas de pesquisa e inovação. *In*: Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação – XXII CBBD. Brasília DF, 8-11 jul 2007. *Anais*... Brasília

DF: FEBAB, 2007 (CD ROM). Área: Gestão; Sub-área: Gestão da Informação e Inteligência Competitiva.
ROBREDO, J.; BOTELHO, Tânia Mara G.; RAMOS E CORTE, Adelaide. Aplicação dos resultados de um estudo Delfos ao desenvolvimento e revisão de currículos em Biblioteconomia e Ciência da Informação. *Revista de Biblioteconomia de Brasília*, v.16, n.2, jul./dez. 1988, p.157-177. Disponível em: http://164.41.122.25/portalnesp/ojs-2.1.1/index.php/RBB/article/view/457. Acesso em em agosto 2009.
ROBREDO, J.; BRISNER, O. Automatic indexing tests for the ALLIANCE Project. *In*: Meeting of the Sub-Committee A12 "Documentation and Information Retrieval". Juan-les-Pins, France, 1973c. *Travaux*. International Commission on Glass, 1973c.
ROBREDO, J.; CANTANHEDE, R.S. Identificação automática de grupos de pesquisa mediante o estudo infométrico de clusters de autores e co-autores. In: VI Encontro da Associação Brasileira de Pesquisa e Pós-graduação em Ciência da Informação – VI ENANCIB, Florianópolis SC, 28-30-nov 2005. *Anais...* Florianópolis SC: UFSC, 2005. (CD ROM). Grupo Temático 7: Informação para Diagnóstico, Mapeamento e Avaliação
ROBREDO, J.; CANTANHEDE, R.S. Utilização de técnicas infométricas na organização de bancos de conhecimento sobre recursos humanose aplicações nos processos decissórios. *In*: VII Encontro Nacional de Pesquisa em Ciência da Informação – VII ENANCIB. Marília SP,19-22 nov 2006. Marília SP: UNESP, 2006. *Anais*... ISBN 85-86718-35-2 (CD ROM). Grupo Temático 7 – Produção e Comunicação da Informação em Ciência e Tecnologia.
ROBREDO, J.; CANTANHEDE, R.S. A probabilidade de coocorrência de termos com afinidade semântica como base para o desenvolvimento de funcionalidades mais performantes na organização do conhecimento e acréscimo de valor agregado da recuperação de informações. *In*: VII Encuentro de la Asociación de Educadores e Investigadores de Bibliotecología, Archivología, Ciências de la Información y Documentación de Iberoamérica y El Caribe – VII EDIBCIC. Marília SP, 22-24 nov 2006. I... Marília SP: UNESP, 2006. *Anais*... ISBN 85-86738-35-2 (CD ROM).
ROBREDO, J.; CANTANHEDE, R.S. Identificação automática de grupos de pesquisa e geração de agrupamentos temáticos: Uma ferramenta global para planejamento e monitoramento de políticas científicas e tecnológicas. *In*: IX Encontro Nacional de Pesquisa em Ciência da Informação – IX ENANCIB. São Paulo SP, 28-30 set 2008. *Anais*... São Paulo SP: USP/ECA, 2008. (CD ROM). Tema
ROBREDO, J.; CHASTINET, Y.S.; PONCE, C. de A. *Metodologia para a elaboração da lista Básica (core list) dos periódicos nacionais em ciências agrícolas e estudo da dispersão da literatura agrícola brasileira*. FAO, Roma (Italy), 1974. Report No.: FAO-GIL – UNDP/BRA/72/020; FAO-GIL – DOC/TEC/74/030.
ROBREDO, J.; CHASTINET, Y.S.; LOBO, P.R.A. Elaboración de um thesaurus agrícola baseado em critérios de eficiência del lenguaje em el proceso de comunicación. *In*: Worldwide International Association of Agricultural Librarians and Documentalists Congress, *Proceedings*...5, México City (México), 14 Apr 1975.
ROBREDO, J.; CHASTINET, Y.S.; LOBO, P.R.A. Construção de um núcleo de thesaurus em agricultura baseado no uso real dos descritores. *In*: Reunião Brasileira de Ciência da Informação, 1. Rio de Janeiro RJ, 15 jun, 1975. *Anais*..., 1975, 17 p.
ROBREDO, J.; CHASTINET, Y.S.; LOBO, P.R.A. *The AGRIS data base as a support for the selective dissemination service BIP / AGRI*. FAO, Roma (Italy), 1976. Report No.: FAO-GIL – UNDP/BRA/72/020; FAO-GIL – DOC/TEC/76/019 (Rev.1).
ROBREDO, J.; OLIVEIRA, L.R. de. Novo modelo de processamento e difusão da produção científica e artística da Universidade de Brasília *In*: V Encontro Nacional de Pesquisa em Ciência da Informação – V ENANCIB. Belo Horizonte MG, 10-14 nov 2003. *Anais*..., Belo Horizonte MG: Escola de Ciência da Informação da UFMG, 2003. (CD ROM). Tema 2 – Representação do Conhecimento; Indexação; Teoria da Classificação.
ROSTAING, Hervé. La bibliométrie et ses techniques. *Sciences de la Société*, (coll. 'Outils et méthodes'), 1996 (*apud* BARTS, 2008).
ROSTAING, *Hervé. Veille technologique et bibliométrie: concepts, outils, applications.* Université de Droit et des Sciences d'Aix-Marseille – Faculté des Sciences et Techniques de St Jérôme, 1993 (Thèse).
ROUSSEAU, Ronald. Relations between continuous versions of bibliometric laws. *Journal of the American Society for Information Science,* v. 41, n. 3, 1990, p. 197–203 [Electronic version]. (*Apud* Jayroe, 2008).
ROUSSEAU Ronald. George Kingsley Zipf: life, ideas, his law and informetrics, diponível em: Glottometrics, v.3, 2002, p.62-80.
ROUSSEAU, Ronald. *Home page*. 2009. (CV detalhado e lista completa de publicações) Disponível em: http://users.telenet.be/ronald.rousseau/html/english_homepage.html#CV. Acesso em agosto 2009.
SALTON, G. *Automatic Information Organization and Retrieval*. New York: McGraw-Hill, 1968.

SALTON, G.; McGILL, J.M. *Introduction to Modern Information Retrieval*. New York, N.Y.: McGraw-Hill 1983 (*apud* HJørland, 2006).
SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. *Information Processing and Management, 24*(5), 1996, p. 513-523.
SANTOS, R.N.M.; KOBASHI, Nair Yumiko; IGAMI, Mery Piedad Zamudio ; BRESSIANI, J.C. Análise cienciométrica da produção científica por meio de dissertações e teses: uma experiência brasileira. In: VII Congreso de Indicadores de Ciencia y Tecnologia, 2007, São Paulo. Nuevos Indicadores para Nuevas Demandas de Información, 2007.
SARACEVIC, T. Information science. *Journal of the American Society for Information Science ,* v. 50, n.12, 1999. p.1051-1063. Disponível em: http://comminfo.rutgers.edu/~tefko/JASIS1999.pdf. Acesso em setembre 2009.
SARACEVIC, T. (2008). Effects of inconsistent relevance judgments on information retrieval test results: A historical perspective. *Library Trends*, v. 56, n. 4, 2008, p. 763-783. Disponível em: http://comminfo.rutgers.edu/~tefko/LibraryTrends2008.pdf. Acesso em setembro 2009.
SARACEVIC, T. (2000). Digital Library Evaluation: Toward Evolution of Concepts. *Library Trends*, v. 49, n. 2, 2000, p. 350-369. (Special issue on Evaluation of Digital Libraries.) Disponível em: http://comminfo.rutgers.edu/~tefko/LibraryTrends2000.pdf. Acesso em setembro 2009.
SARACEVIC, T. *Bibliografical Sketch*. 2009. Disponível em: http://comminfo.rutgers.edu/~tefko/biosketch.htm. Acesso em setembro 2009.
SEM, B.K; DUTTA, Bidyarthi; DAS, Anup Kumar. INSDOC's Contribution to Bibliometrics. Paper presented at the XII IASLIC Conference, Trivandrum, 4-7 Dezember 2001. *Annals of Library and Information Studies*, v.49, n.1, 2002, p. 1-6. Disponível em: http://drtc.isibang.ac.in:8080/xmlui/bitstream/handle/1849/198/Sen_Dutta_Das_ALIS02.pdf?sequence=1. Acesso em agosto 2009.
SHAW, Alexandra. B.C. Brookes and the development of information science: a bibliography. *Journal of Information Science*, v.16, n.1, 1990, p.3-7.
SILVA, Gilda Olinto do Valle. Capital cultural, classe e gênero em Bourdieu. *Informare*, v.1, n.2, jul/dez 1995.
SOERGEL, Dagobert. Multilingual thesauri and ontologies in cross-language retrieval. *In*: AAAi-97 Spring Symposium, Stanford, CA, March 24-26, 1997. *Papers*, Published in the Symposium Tecnical Report, pela American Association for Artificial Intelligence – AAAI (AAAI Technical Report SS-97-05), 1997. Disponível em: http://www.aaai.org/Papers/Symposia/Spring/1997/SS-97-05/SS97-05-026.pdf. Acesso em setembro 2009.
SOUZA. Held Barbosa de; NAGLIS. Daniela Leite. *Mineração de dados: uma aplicação na base de dados de artigos de periódicos científicos das áreas de informação (ABCDM)*. Brasília DF: Universidade de Brasília / Faculdade de Tecnologia, 2008, iv + 67p. (Monografia de Especialização em Gestão de Tecnologia da Informação). Romualdo Alves Pereira Júnior, Orientador; Jayme Leiro Vilan Filho, Co-orientador. Disponível em: http://bdm.bce.unb.br/bitstream/10483/158/1/2007_HeldSouza_DanielaNaglis.pdf. Acesso em out 2009.
SOUZA, Sonia Regina Burnier de. *Sistema GIBI: Manual de Serviço*. Brasília DF: IBICT, 1991. 1 v. (Memória Técnica do IBICT ; MT/0685).
SPÄRCK JONES, K. Some thoughts on classification for retrieval. *Journal of Documentation*, v.26, n.2, 1970, p. 89-101. Disponível em: http://www.db.dk/bh/core%20concepts%20in%20lis/Sparck%20Jones_1970.pdf. Acesso em setembro 2009.
SPÄRCK JONES, K. *Automatic keyword classification for retrieval*. London: Butterworth. 1971.
SPÄRCK JONES, K. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, v. 28, n. 1, 1972, p. 11-21.
SPÄRCK JONES, K. Index terms weighting. *Information Storage and Retrieval*, v. 9, 1973, p. 619-633.
SPINK, A.; WOLFRAM, D.; JANSEN, B.J.; SARACEVIC, T. (2001). Searching the Web: The public and their queries. *Journal of the American Society for Information Science and Technology,* v. 52, n. 3, 2001, p. 226-234. Disponível em: http://comminfo.rutgers.edu/~tefko/JASIST2001.pdf. Acesso em setembro 2009.
STREHL, Leticia; STUMPF, Ida Regina C. The Iimpact of Brazilian Scientific Articles Published in Domestic and Foreign Journals Indexed in Web of Science. *In*: 12th International Conference in Scientometrics and Informetrics – ISSI 2009. Rio de Janeiro RJ, Brazil, 14-17 July 2009. Poster session – Board 47.
STUMPF, I.R.C.; MESQUITA, R.M.A. Estudo de Citações de Documentos Eletrônicos on-line em Revistas da Área de Comunicação. Em Questão, Porto Alegre, v. 10, n. 2, p. 261-274, 2004.
STUMPF, I.R.C.; WEBER, M.H. Comunicação e Informação: conflitos e convergências. *In*: Maria Immacolata Vassallo de Lopes (org.). *Epistemologia da Comunicação*. São Paulo SP: Edições Loyola, 2003, p. 121-134.

STUMPF, Ida Regina Chittó ; CAREGNATO, S. E. ; VANTI, Nádia ; VANZ, Samile Andréa de Souza ; CORRÊA, Cynthia Harumy W ; CRESPO, Isabel Merlo ; GALDINO, Karina ; GOMES, Janaina . Uso dos Termos Cienciometria e Cientometria pela Comunidade Científica Brasileira. *In*: Dinah Aguiar Poblacion; Geraldina Porto Witter; José Fernando Modesto da Silva. (orgs.). *Comunicação e Produção Científica: contexto, indicadores e avaliação*. São Paulo: Angellara, 2006, p.341-365.

SZÁVA-KOVÁTS, Endre. Unfounded attribution of the half-life index-number of literature obsolescence to Burton and Kebler: A literature science study. *Journal of the American Society for Information Science and Technology,* v.53, n.13, 2002, p. 1098-1105

TESS, Beatriz Helena; FURUIE, Sérgio Shiguemi; CASTRO, Regina Célia Figueiredo; BARRETO, Maria do Carmo Cavarette; NOBRE, Moacyr Roberto Cuce. Assessing the scientific research productivity of a Brazilian healthcare institution: a case study at the heart institute of São Paulo, Brazil. *Clinics,* v.64, n.6, 2009, p.571-576, June 2009. DOI: 10.1590/S1807-59322009000600013. *Print version* ISSN 1807-5932. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1807-59322009000600013. Acesso em setembro 2009.

THELWALL, Mike. Bibliometrics to webometrics. *Journal of Information Science*, v.34, 2008, 605. DOI: 101177/0165551507087238. Disponível em: http://sagenpub.com/cgi/content/abstract/34/4/605. Acesso em: setembro 2009.

THELWALL, M.; AGUILLO, I. F. La salud de las web universitarias españolas. *Revista Española de Documentación Científica*, 2003, vol 26(3): 291-305

THELWALL, M.; VAUGHAN, L. Webometrics: An introduction to the special issue. *Journal of the American Society for Information Science,* v.55, n.4, 2004, p.1213–1215 [Electronic version].

THELWALL, M.; VAUGHAN, L.,; BJÖRNEBORN, L. (2005). Webometrics *Annual Review of Information Science and Technology,* v.39, n.1, 2005, p.81–135 [Electronic version].

THOMSON REUTERS. Dr. Eugene Garfield Wins the "Online Information Lifetime Achievement Arward". *Press Room – Science.* London UK – November 30, 2006 Disponivel em: http://science.thomsonreuters.com/press/2006/8351497/. Acesso em agosto 2009.

TODOROV, R. Comments on Jan Vlachý, recipient of the 1989 Derek de Solla Price Award. *Scientometrics,* v.19, n.3-4, 1990, p.157-158. ISBN 0138-9130 (Print) 1588-2861 (Online). Disponível em: http://www.springerlink.com/content/m36r764636234177/. Acesso em agosto 2009.

TRZESNIAK, Piotr. Indicadores quantitativos: Reflexões que antecedem o seu estabelecimento. *Ciência da Informação*, v.27, n.2, mai/ago 1998, p. 159-164. Disponível em: http://www.usp.br/siicusp/Resumos/17Siicusp/resumos/2244.pdf. Acesso em outubro 2009.

UNESCO. *Mini-micro CDS/ISIS – Refrenece Manual*. Versão 2.3 -3.071. Paris, França: Unesco, 1993.

UNESCO. *CDS/ISIS – Windows Version.* Paris, França: Unesco, 1997. Disponível em: ftp://ftp.unesco.org.pub/winisis. *Apud* Robredo e Oliveira (2005).

URBIZAGÁSTEGUI ALVARADO, Rubén. A bibliometria no Brasil. *Ciência da Informação*, v. 13, n. 2, 1984, p. 91-105. Disponível em: http://revista.ibict.br/index.php/ciinf/article/viewArticle/1444. Acesso em agosto 2009.

USP/ECA. Núcleo de Pesquisa de Produção Científica – ECA/USP. Disponível em: http://www.eca.usp.br/nucleos/pc/ProjeodoNPCnaComunidadeCientfica.htm. Acesso em: 15 set. 2007

UTEXAS – University of Texas. *Biblimetrics* [s.d.]. Disponivel em: http://www.ischool.utexas.edu/~palmquis/courses/biblio.html. Acesso em agosto 2009.

VAISHNAVI, Vijay K.; VANDENBERG, Art; ZHANG, Yanqing; DURAISAMY, Saravanaraj**.** Towards design principles for effective context- and perspective-based web mining. 4th International Conference on Design Science Research in Information Systems and Technology, Philadelphia, Pennsylvania, 2009. *Proceedings,* New York, NY, USA: ACM, 2009. ISBN 978-1-60558-408-9. Disponível em: http://doi.acm.org/10.1145/1555619.1555643. Acesso em setembro 2009.

VALDERRAMA ZURIÁN, Juan Carlos; GISBERT TIO, Amparo; TERRADA FERRANDIS, Maria Luz. *Bibliografia española e internacional de estudos bibliométricos*, 1996, 210 p. (Cadernos de Documentação e Informática Biomédica, XIV).

VAN RAAN, A.F.J. *Hanbook of Quantitative Studies of Science and Technology*. Amsterdam: Elsevier, 1988.

VANTI, Nadia Aurora Peres. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. *Ciência da Informação,* v.31, n.2, 2002; DOI: 10.1590/S0100-19652002000200016; *Print version* ISSN 0100-1965. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652002000200016. Acesso em: setembro 2009.

VANZ, S.A.S.; STUMPF, I.R.C. A Colaboração científica intra e inter-institucional no Brasil. *In*: IX Encontro Nacional de Pesquisa em Ciência da Informação, 2008, São Paulo SP. *Anais ENANCIB*. São Paulo SP: ANCIB, 2008.
VANZ, Samile Andréa de Souza ; CAREGNATO, S. E. . Estudos de citação: uma ferramenta para entender a comunicação científica. *Em Questão* (UFRGS), Porto Alegre, v. 9, n. 2, p. 295-307, 2003.
VARGAS, Maria Helena Simões. *A formação do campo denominado capacidade dinâmica: um estudo bibliométrico*. São Paulo SP: Universidade Paulista/Instituto de Ciências Humanas.2009. (Trabalho de Conclussão de Pós-Graduação). Disponível em: http://www3.unip.br/ensino/pos_graduacao/strictosensu/administracao/download/adm_mariahelenasimoesvargas.swf. Acesso em setembro 2009.
VELHO, L. Como medir a ciência? *Revista da Sociedade Brasileira de Ciência e Tecnologia de Alimentos*, Brasília, v. 16, n. 1, p. 35-41, 1985.
VELHO, L. The meaning of citation in the context of a scientifically peripheral country. *Scientometrics*, Amsterdam, v. 9, n. 1-2, p. 71-89, 1986.
VELHO, L.; KRIGE, J. Publication and citation practices of Brazilian Agricultural scientists. *Social Studies of Science*, London, UK, v. 14, n. 2, p. 45-62, 1984.
VERRE Online – Le portal français du verre. *L'industrie du verre*: *Organismes officiels*. Disponível em: http://www.verreonline.fr/v_gene/orga_07.php. Acesso em outubro 2009.
VICKERY, B.C. *Faceted Classification. A Guide to the Construction and Use of Special Schemes*. London: ASLIB, 1960.
VICKERY, B.C. *Information systems*, London: Butterworths, 1973.
VICKERY, B.C.; VICKERY, A. *Information science in theory and practice*, 3rd edition. Munich: Saur, 2004). (1st ed.:1987; 2nd ed: 1992.) (*Apud* WIKIPEDIA, the free encyclopedia. *Citation index*. Last modified on 4 February 2010. Didponível em: http://en.wikipedia.org/wiki/Citation_index. Acesso em fevereiro 2010.
VILAN FILHO, Jayme Leiro. Base de artigos de periódicos científicos do Brasil e Portugal nas áreas de informação. In: VI Congresso de Arquivologia do Mercosul, 2005, Campos do Jordão. *Anais do VI CAM*. São Paulo SP: Associação de Arquivistas de São Paulo - ARQ-SP, 2005.
VILAN FILHO, Jayme Leiro. Autoria Múltipla em Artigos de Periódicos Científicos das Áreas de Informação no Brasil: o gênero dos autores e o tipo de autoria. *In*: IV Workshop Internacional em Ciência da Informação (IV WICI). Brasília DF. *Anais*, 2009 (Pôster).
VILAN FILHO, Jayme Leiro. Autoria Múltipla em Artigos de Periódicos Científicos das Áreas de Informação no Brasil: identificação de fatores de interferência. *In*: III Workshop Internacional em Ciência da Informação (WICI). Brasília DF, 2008 (Pôster).
VILAN FILHO, Jayme Leiro; SILVA, Sebastião Dimas Justo da; VIEIRA, José Ronaldo. Base de Artigos de Periódicos Científicos do Brasil e Portugal nas Áreas de Informação. *In*: II Congreso Mundial de Usuarios del CDS/ISIS, 2005, Salvador BA, 2005 (Pôster).
VILAN FILHO, Jayme Leiro. Base de artigos de periódicos em português nas áreas de informação. *In*: XXI Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação – CBBD. Curitiba PR, 2005. Curitiba PR: FEBAB; Associação Bibliotecária do Paraná, 2005 (Pôster).
VILAN FILHO, Jayme Leiro; SOUZA, Held Barbosa de; MUELLER, Suzana Pinheiro Machado . Artigos de periódicos científicos das áreas de informação no Brasil: evolução da produção e da autoria múltipla. *Perspectivas em Ciência da Informação*, v.13, p.2-17, 2008.
VILAN FILHO, Jayme Leiro. Periódicos científicos brasileiros de arquivologia: os artigos e suas autorias (1972-2007). 2009. Pendente de publicação.
VILAN FILHO, Jayme Leiro. *Autoria múltipla em artigos de periódicos científicos das Áreas de Informação no Brasil: identificação de fatores de interferência*. Brasília: Universidade de Brasília DF, 2007 (Projeto de tese, qualificado em 17 de dezembro de 2007).
VILAN FILHO, Jayme Leiro. *Artigos de Periódicos das Áreas de Informação no Brasil: Evolução dos temas.* Brasília DF: Universidade de Brasília/Departamento de Ciência da Informação e Documentação (UnB/CID). 2009 (Pesquisa em andamento).
VINKLER, Péter. Introducing the Contemporary Contribution Index for Characterizing the Recent,Relevant Impact of Journals. 11th International Conference of the International Society for Scientometrics and Informetrics (ISSI 2007), Madrid 25-27 June. *In*: Daniel Torres-Salinas and Henk F. Moed. *Proceedings*. Madrid: Consejo Superior de Investigações Científicas, Centro de Informação e Documentação Científica (CSIC/INDOC), 2007, p.753-760. Disponível em:
VINKLER, Peter. Tudomanimetriai Kutatások Magyarországon [Scientometric Investigations in Hungary]. *Magyar Tudomány* [*Hungarian Science*], 2008/11, p. 1372-1380. Disponível em: http://www.iif.hu/2008-11,pdf. Acesso em agosto 2009.

WERSIG, G. *Informationssoziologie. Hinweise zu einem informationswissenschaftlichen Teilbereich.* Frankfurt am Main: Athenäum Fischer Taschenbuch Verlag, 1973. (*apud* Hjørland, 2005).
WERSIG, G. Information science and theory: A weaver bird's perspective. In: P. Vakkari; B. Cronin (eds.), *Conceptions of library and information science: Historical, empirical and theoretical perspectives*, London: Taylor Graham, 1992, p. 201-217.
WERSIG, G. Information science: The study of postmodern knowledge usage. *Information Processing & Management*, v. 29, n. 2, 1993.
WERSIG, G. Information theory. *In* : J. Feather, & P. Sturges (eds.), *International encyclopedia of library and information science*. London & New York: Routledge, 2003, p. 310-319. (*apud* Hjørland, 2005)
WIKIPEDIA – The Free Encyclopedia. *Econometrics – Notable econometricians.* (Page last modified on 18 October 2009). Disponível em: http://en.wikipedia.org/wiki/Econometrics. Acesso em out 2009.
WIKIPEDIA – The Free Encyclopedia. *Gernot Wersig*, 2005. Disponível em: http://de.wikipedia.org/wiki/Gernot_Wersig. Acesso em agosto 2009.
WHITE, Howard D.; BATES, Marcia; WILSON, Patrck. *For Information Specialists: Interpretations of Reference and Bibliographic Work*. Norwood NJ: Ablex,1992. 320 p.
WHITE, Howard D. *Brief Tests of Collection Strength; A Methodology for All Types of Libraries.* Westport CT: Greenwood Press, 1995. 210 p.
WHITE; McCAIN. Visualizing a Discipline: An Author Co-Citation Analysis of Information Science, 1972-1995. *Journal of the American Society for Information Science,* v. 49, 1998, p. 327-355.
WIKIPEDIA the free encyclopedia. *Veille technologique*. (Dernière modification de cette page le 8 août 2009). Disponível em: http://fr.wikipedia.org/wiki/Veille_technologique Acesso: agosto 2009.
WIKIPÉDIA, a enciclopédia livre*. Experimentação*. (Página modificada pela última vez em 30 de maio de 2009). Disponível em: http://pt.wikipedia.org/wiki/Experimenta%C3%A7%C3%A3o. Acesso, setembro 2009.
WORMELL, Irene. Informetria: explorando bases de dados como instrumentos de análise. *Ciência da Informação*, Brasília, v. 27, n. 2, p. 210-216, maio/ago. 1998.
WOUTERS, Paul. Scientometrics by Hand. The Ups and Downs of Scientometrics in Russia [Communication] *ADEST (Association pour la mesure des sciences e des techyniques).* March 1988. URL: http://www.upmf-grenoble.fr/seminaires/wouters.html. CYBERmetrics. International Journal of Scientometrics, Informetrics and bibliometrics. ISSN 1137-5019 Papers & Abstracts:Papers. Disponível em: http://cybermetrics.cindoc.csic.es/articulos.asp?art=445. Acesso em setembro 2009.
ZIPF, Gerge Kingsley. *Selective Studies and the Principle of Relative Frequency in Language.* Cambridge, Mass.: Harvard University Press, 1932. (*Apud* HINKLE, 2006?).
ZIPF, Gerge Kingsley. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology.* Cambridge, Mass.: Addison-Wesley, 1949. *Apud* HINKLE, 2006 ?).

ANEXO 1

Exemplo de aplicação da lei de Zipf à análise de um resumo de 3.100 palavras, procedente de um artigo de um artigo de autoria de um dos autores do presente Capítulo (ROBREDO *et al.*, 1988).

A experiência foi realizada em cinco etapas, que serão descritas a seguir com o detalhamento necessário:

Etapa 1. – Escolha do texto a analisar;
Etapa 2. – Extração e ordenação das palavras do texto;
Etapa 3. – Reunião das palavras por radicais, desinências e flexões, contagem das respectivas freqüências de ocorrência;
Etapa 4. – Ordenação dos dados para montagem do gráfico;
Etapa 5. – Montagem do gráfico

ETAPA 1. – Escolha do texto a analisar

Jaime **Robredo**; Tânia Mara G. **Botelho**; Adelaide **Ramos e Corte**.
Aplicação dos resultados de um estudo Delfos ao desenvolvimento e revisão de currículos em Biblioteconomia e Ciência da Informação. *Revista de Biblioteconomia de Brasília*, v.16, n.2, jul./dez. 1988, p.157-177. Disponível em: http://164.41.122.25/portalnesp/ojs-2.1.1/index.php/RBB/article/view/457.

O método Delfos tem sido amplamente utilizado, nos países industrializados, para detectar tendências nas necessidades do mercado de trabalho e para orientar novas estruturas curriculares em Biblioteconomia e Ciência da Informação. Comparativamente, só alguns exemplos de aplicação desse tipo encontram-se na literatura dos países em desenvolvimento. O Conselho Federal de Educação estabeleceu recentemente novas normas para o desenvolvimento dos currículos das escolas de Biblioteconomia. Dentro dessas normas, cada escola possui autonomia e flexibilidade para desenvolver seu próprio currículo, de acordo com as características ambientais. No Departamento de Biblioteconomia da Universidade de Brasília, foi criado um grupo para elaborar uma proposta de currículo novo que representaria os sentimentos e anseios do corpo docente. Paralelamente, um dos autores deste artigo coordenou o trabalho de um grupo de professores do departamento visando a realizar um cada estudo Delfos no Distrito Federal, ou seja, na área geográfica de influência da Universidade de Brasília. O objetivo deste estudo foi detectar as falhas e lacunas na formação atual dos profissionais com vistas à orientação de conteúdos programáticos específicos que reforcem determinados aspectos do perfil, de acordo com a demanda previsível do mercado de trabalho da informação nos próximos anos. As opiniões dos profissionais e dos especialistas mostravam uma forte convergência sobre a necessidade de reforçar, no novo currículo, algumas áreas específicas, tais como as aplicações da informática e o uso de bases de dados, as técnicas de indexação e de recuperação de informação, telecomunicações, práticas gerenciais e métodos quantitativos. No caso da pós-graduação, após cinco anos de experiência, foram também introduzidas, no currículo do curso de Mestrado em Biblioteconomia e Documentação, algumas modificações visando a satisfazer à demanda do mercado. A mesma tendência foi observada em estudos sobre desenvolvimento de currículos em outros países em desenvolvimento, alguns dentre eles auspiciados por organismos internacionais. Poderia afirmar-se, de fato, que existe consenso generalizado sobre os princípios que devem guiar os novos currículos em Biblioteconomia e Ciência da Informação, nos países em desenvolvimento. Entretanto, não foi encontrada até agora uma receita geral para implementar de forma adequada as mudanças que exigem as novas tendências, no que diz respeito aos hábitos, habilidades e mentalidades de professores. A consciência do problema, embora essencial, não significa a capacidade de resolvê-lo. Poderíamos ter alguma esperança se, nos próximos anos, os organismos internacionais e as escolas de Biblioteconomia dos países em desenvolvimento trabalhassem juntos para estabelecer um programa de treinamento e reciclagem acelerado dos professores.
NOTA: O artigo reúne elementos essenciais de uma comunicação apresentada ao *"Seminar on Information Manpower Forecasting"*, patrocinado pelo *"Education and Training Committee"* da Federação Internacional de informação e Documentação (FID/ET), celebrado em Espoo, Finlândia de 24 a 27 de agosto de 1988. Os dados referentes às estruturas dos currículos dos cursos de graduação e de pós-graduação do Departamento de Biblioteconomia da Universidade de Brasília foram devidamente atualizados.

SUMÁRIO GERAL SUMÁRIO DO CAPÍTULO 10

ETAPA 2. – Extração e ordenação das palavras do texto

1. Termos compostos

[Ciência da Informação]
[Ciência da Informação]
[Ciência da Informação]
[Ciência da Informação]
[Conselho Federal de Educação]
[conteúdos programáticos]
[corpo docente]
[Departamento de Biblioteconomia]
[Departamento de Biblioteconomia]
[desenvolvimento de currículos]
[desenvolvimento de currículos]
[Distrito Federal]
[Education and Training Committee]
[Federação Internacional de Informação e Documentação]
[Information Manpower Forecasting]
[mercado de trabalho]
[Mestrado em Biblioteconomia e Documentação]
[métodos quantitativos]
[organismos internacionais]
[países em desenvolvimento]
[países em desenvolvimento]
[países em desenvolvimento]
[países industrializados]
[revisão de currículos]
[Universidade de Brasília]
[Universidade de Brasília]
[Universidade de Brasília]

2. Termos simples

1988
24
27
a
a
a
A
A
a
a
à
à
acelerado
acordo
acordo
adequada
afirmar-se
agora
agosto
alguma
algumas
algumas
alguns
alguns
ambientais
amplamente
and
anos
anos
anos
anseios
ao
ao
aos
Aplicação
aplicação
aplicações
após
apresentada
área
áreas
artigo
artigo
as
as
As
as
as
as
as
às
aspectos
até
atual
atualizados
auspiciados
autonomia
autores
bases
Biblioteconomia
Biblioteconomia
Biblioteconomia
Biblioteconomia
Biblioteconomia
Biblioteconomia
Biblioteconomia
Biblioteconomia
Brasília
Brasília
Brasília
cada
capacidade
características
caso
celebrado
celebrado
Ciência
Ciência
Ciência
cinco
com
com
com
Committee
como
Comparativamente
comunicação
consciência
Conselho
consenso
conteúdos
convergência
coordenou
corpo
criado
curriculares
currículo
currículo
currículo
currículo
currículos
currículos
currículos
currículos
currículos
curso
cursos
da
Da
da
da

da	de	Distrito	e
da	de	diz	e
da	de	do	e
da	de	do	Educação
da	de	do	Education
da	De	do	elaborar
dados	de	do	elementos
dados	de	do	eles
das	de	do	em
de	de	do	em
de	de	do	em
de	de	docente	em
de	Delfos	Documentação	em
De	Delfos	Documentação	em
de	Delfos	dos	em
de	demanda	dos	em
de	demanda	dos	em
de	dentre	dos	em
de	Dentro	dos	embora
de	Departamento	dos	encontrada
de	departamento	dos	encontram-se
de	Departamento	dos	Entretanto
de	desenvolver	dos	escola
de	desenvolvimento	dos	escolas
de	desenvolvimento	dos	escolas
de	desenvolvimento	e	especialistas
de	desenvolvimento	e	específicas
de	desenvolvimento	e	específicos
de	desenvolvimento	e	esperança
de	dessas	e	Espoo
de	desse	e	essenciais
de	deste	e	essencial
de	deste	e	estabelecer
de	detectar	e	estabeleceu
de	detectar	e	estruturas
de	determinados	e	estruturas
de	devem	e	estudo
de	devidamente	e	estudo

SUMÁRIO GERAL SUMÁRIO DO CAPÍTULO 10

estudo
estudos
exemplos
exigem
existe
experiência
falhas
fato
Federação
Federal
Federal
FID/ET
Finlândia
flexibilidade
foi
foi
foi
foi
foram
foram
Forecasting
forma
formação
forte
generalizado
geográfica
geral
gerenciais
graduação
grupo
grupo
guiar
habilidades
hábitos
implementar
indexação
industrializados
influência
Informação
Informação
Informação
informação
Informação
Informação
informática
Information
internacionais
internacionais
Internacional
introduzidas
juntos
lacunas
literatura
Manpower
mentalidades
mercado
mercado
mercado
mesma
Mestrado
método
métodos
modificações
mostravam
mudanças
na
na
na
não
não
nas
necessidade
necessidades
no
no
no
No
no
normas
normas
nos
nos
nos
nos
novas
novas
novas
novo
novo
novos
O
O
o
o
O
o
O
objetivo
observada
On
Opiniões
organismos
organismos
orientação
orientar
os
os
os
os
os
Os
ou
outros
países
países
países
países
para
para
para
para
para
para
para
Paralelamente
patrocinado
pelo
perfil
Poderia
Poderíamos
por
pós-graduação
pós-graduação
possui
práticas
previsível
problema
professores
professores
professores
profissionais
profissionais
programa
programáticos
proposta
próprio
próximos

SUMÁRIO GERAL SUMÁRIO DO CAPÍTULO 10

próximos
quantitativos
que
que
que
que
que
que
realizar
receita
recentemente
reciclagem
recuperação
reforcem
reforçar
resultados
satisfazer
se
seja
Seminar
sentimentos
seu
sido
referentes
reforçar
resolvê-lo
respeito
significa
só
sobre
sobre
sobre
tais
também
tem
trabalhassem
técnicas
telecomunicações
trabalho
trabalho
trabalho
Training
treinamento
um
um
um
um
uma
uma
uma
uma
Universidade
Universidade
Universidade
uso
utilizado
visando
visando
vistas

ETAPA 3. – Reunião das palavras por radicais, desinências e flexões, contagem das respectivas freqüências de ocorrência [Os números à esquerda indicam a freqüência das palavras]

1. Termos compostos

1-[Conselho Federal de Educação]
1-[conteúdos programáticos]
1-[corpo docente]
1-[Distrito Federal]
1-[*Education and Training Committee*]
1-[Federação Internacional de Informação e Documentação]
1-[*Information Manpower Forecasting*]
1-[mercado de trabalho]
1-[Mestrado em Biblioteconomia e Documentação]
1-[métodos quantitativos]
1-[organismos internacionais]
1-[países industrializados]
1-[revisão de currículos]
2-[Departamento de Biblioteconomia]
2-[desenvolvimento de currículos]
3-[Ciência da Informação]
3-[países em desenvolvimento]
3-[Universidade de Brasília]

2. Palavras reunidas por radicais, dessinências e flexões

1-1988
1-24
1-27
1-acelerado
1-adequada
1-afirmar-se
1-agora
1-agosto
1-ambientais
1-amplamente
1-*and*
1-anseios
1-após
1-apresentada
1-às
1-aspectos
1-até
1-atual
1-atualizados
1-auspiciados
1-autonomia
1-autores
1-bases
1-cada
1-capacidade
1-características
1-caso
1-celebrado
1-cinco
1-*Committee*
1-como
1-Comparativamente
1-comunicação
1-consciência
1-Conselho
1-consenso
1-conteúdos
1-convergência
1-coordenou
1-corpo
1-criado
1-dentre
1-Dentro
1-Distrito
1-diz
1-docente
1-elaborar
1-elementos
1-eles
1-embora
1-encontrada/ encontram-se
1-Entretanto
1-especialistas
1-esperança
1-Espoo
1-Essenciais/essencial
1-exemplos
1-exigem
1-existe
1-experiência
1-falhas
1-fato
1-Federação
1-Federal
1-FID/ET
1-Finlândia
1-flexibilidade
1-*Forecasting*
1-forma
1-formação
1-forte
1-generalizado
1-geográfica
1-geral
1-gerenciais
1-graduação
1-guiar
1-habilidades
1-hábitos
1-implementar
1-indexação
1-industrializados
1-influência
1-informática
1-introduzidas
1-juntos
1-lacunas
1-literatura
1-*Manpower*
1-mentalidades
1-mesma
1-Mestrado
1-métodos
1-modificações
1-mostravam
1-mudanças
1-objetivo
1-observada
1-*on*
1-Opiniões
1-ou
1-outros
1-Paralelamente
1-patrocinado

SUMÁRIO GERAL SUMÁRIO DO CAPÍTULO 10

1-pelo
1-perfil
1-Poderia
1-Poderíamos
1-por
1-possui
1-práticas
1-previsível
1-princípios
1-problema
1-proposta
1-próprio
1-quantitativos
1-realizar
1-receita
1-recentemente
1-reciclagem
1-recuperação
1-referentes
1-representaria
1-resolvê-lo
1-respeito
1-resultados
1-satisfazer
1-se
1-*Seminar*
1-sentimentos
1-seu
1-tais
1-também
1-técnicas
1-telecomunicações
1-tipo
1-significa
1-só
2-à 2-determina/detrminados
2-área/áreas
2-acordo
2-artigo
2-curso
2-dados
2-demanda
2-detectar
2-devem/devidamente
2-estabelecer/estabeleceu
2-Documentação
2-Educação/*Education*
2-específicas/específicos
2-estruturas
2-Federal
2-grupo
2-método
2-não
2-necessidade
2-normas
2-orientação
2-organismos
2-pós-graduação
2-Training/treinamento
2-uso/ utilizado
2-visando/visando/vistas
3-anos
3-ao/aos
3-Delfos
3-Departamento
3-escola
3-mercado
3-na/nas
3-professores
3-sobre
3-tendência/tendências
3-Universidade
4-dessas/desse/deste 3-aplicação
3-Brasília
3-Ciência
3-com
4-estudo/estudos
4-internacionais/internacional 4-nos
4-países
4-trabalhassem/trabalho
5-alguma/algumas/alguns
6-foi/foram
6-no
6-novas
6-que
7-desenvolver/desenvolvimento
7-Informação*/information*
7-para
8-as
9-Biblioteconomia
10-a
10-curriculares/currículo/currículos
10-em
10-um/uma
11-dos
13-o/os
18-e
62-da/das/de/do

ETAPA 4. – Ordenação dos dados para montagem do gráfico

Eixo X = ranking	Eixo Y = freq.
1	142
2	62
3	33
4	18
5	15
6	13
7	11
8	9
9	8
10	7
11	6
12	1

ETAPA 5. – Montagem do gráfico

Capítulo 11

Rede metodológica integrando epistemologia, organização do conhecimento, bibliometria e tesauros: Concepção e construção do Tesauro Brasileiro de Ciência da Informação[1]

Lena Vania Ribeiro Pinheiro[2] e **Helena Dodd Ferrez**[3]

SUMÁRIO DO CAPÍTULO 11

Resumo do Capítulo 11, *p. 260*
Como citar o Capítulo 11, *p. 260*

1. **INTRODUÇÃO DO CAPÍTULO 11**, *p. 261*
2. **UMA LONGA JORNADA DE CONSTRUÇÃO DE ALICERCES TEÓRICOS E EMPÍRICOS DE UM TESAURO**, *p. 261*
   - **2.1 Mapeamentos temáticos da Ciência da Informação: análise da literatura estrangeira e definição de categorias**, *p. 262*
   - **2.2 Mapeamento temático da Ciência da Ciência da Informação: análise da literatura brasileira**, *p. 266*
   - **2.3 Classificação para pesquisa de Chaim Zins**, *p.268*
3. **ANÁLISE COMPARATIVA COM OUTROS TESAUROS E CONSULTA A DICIONÁRIOS E ESPESCIALISTAS**, *p. 269*
   - **3.1 Teauro do IBICT, 1989**, *p. 269*
   - **3.2 Tesauro da ASIST, 2005**, *p. 270*
4. **ESTRUTURA DO TESAURO BRASILEIRO DE CIÊNCIA DA INFORMAÇÃO, 2010**, *p. 271*
5. **DAS PRIMEIRAS REFLEXÕES SOBRE O DESAFIO DA CONTRUÇÃO DE UM TESAURO BRASILEIRO DE CIÊNCIA DA INFORMAÇÃO**, *p. 273*
6. **REFERÊNCIAS DO CAPÍTULO 11**, *p. 274*

---

[1] Pesquisa realizada no âmbito do projeto LABCOM- Laboratório de Pesquisas em Comunicação Científica, financiado pela FINEP- Financiadora de Estudos e Projetos

[2] Doutora em Ciência daInformação**,** Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT. Coordenadora do Projeto Tesauro Brasileiro de Ciência da Informação. CV Lattes: http://lattes.cnpq.br/9613980184982976. E-mail: lena.vania@ibict.br.

[3] Mestre em Ciência da Informação, IBICT-UFRJ. bolsista DTI do IBICT, e primeira autora de um tesauro para museus. (FERREZ, Helena D.; BIANCHINI, Maria Helena S. *Thesaurus para acervos museológicos*. Rio de Janeiro: Fundação Nacional Pró-Memória, Coordenadoria Geral de Acervos Museológicos, 1987. 2v.).

Resumo do Capítulo 11

Descrição e análise do processo de concepção e construção do Tesauro Brasileiro de Ciência da Informação, em versão trilingue (português, inglês e espanhol) e nos formatos impresso e eletrônico, nos seus aspectos teóricos e metodológicos, em articulação da Epistemologia, Organização do Conhecimento e Bibliometria. A estrutura representa a constituição ou uma classificação da Ciência da Informação, com suas disciplinas e subdisciplinas, e foi determinada por mapeamento temático, baseado nas análises de conteúdo da literatura brasileira e estrangeira da área (*ARIST* e revista *Ciência da Informação*, do IBICT) e bibliométrica. A consolidação dos termos foi norteada pela análise comparativa de tesauros estrangeiros e um brasileiro, e de dicionários especializados, fortalecida pela consulta a especialistas. O processo evolutivo e as transformações terminológicas são analisadas e demonstram a complexidade, subjetividade e dinamismo do processo, pelas influências de diferentes línguas e culturas, representadas no discurso de uma comunidade científica.

Palavras-chave: Tesauro Brasileiro de Ciência da Informação**;** Ciência da Informação; tesauros; metodologia; fundamentos; construção de tesauros

Abstract

Methodological network integrating epistemology, knowledge organization, bibliometrics and thesauri: Conception and building up of the Brazilian Thesaurus of Information Science

Description and analysis of the conception and process of construction the Brazilian Thesaurus of Information Science in its theoretical and methodological aspects. It was articulated by Epistemology, Knowledge Organization and Bibliometrics. It is a trilingual thesaurus (Portuguese, English, and Spanish) that will be also published in electronic format. This structure represents an Information Science classification with disciplines and sub-disciplines of the area, and it was defined through a theme mapping based on content and bibliometric analysis of Brazilian and foreign literature of the field (*ARIST* and the journal *Ciência da Informação*). Terms consolidation was based in a comparative analysis of foreign and Brazilian thesauri and specialized dictionaries as well as experts contributions. The evolutive process and the terminology changes analyzed show complexity, subjectivity and dynamism of the process, as well as by the influence of different languages and cultures represented in a scientific community's discourse.

KeyWords: Brazilian Thesaurus of Information Science; Information science; methodology; theoretical foundations; thesauri construction; thesauri

Como citar o Capítulo 11:

PINHEIRO, L. V. R.; FERREZ, H. D. Rede metodológica entre epistemologia, organização do conhecimento, bibliometria e tesauros: Concepção e construção do Tesauro Brasileiro de Ciência da Informação. *In:* Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento.* Brasília DF: IBICT, 2010, 335 p. ISBN: 978-85-7013-072-3. Capítulo 10, p. 259-276. Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

1. INTRODUÇÃO DO CAPÍTULO 11

Se pensarmos nas transformações que a Sociedade da Informação e do Conhecimento ocasionou pela introdução e disponibilidade das tecnologias da informação e comunicação -TICs, especialmente a Internet, juntamente com o processo de globalização que se iniciou pela economia e se expandiu por todos os setores da vida, muitas dúvidas nos cercam e os questionamentos são inevitáveis.

Em áreas como a Ciência da Informação, profundamente afetadas por estes novos cenários e circunstâncias históricas, políticas, culturais e sociais, é natural indagarmos sobre a permanência ou validade, hoje, de alguns recursos, metodologias, serviços e produtos de informação até então existentes. O próprio território da ciência, que desde a revolução científica do século XVII vinha apresentando regularidade no seu desenvolvimento, conforme estudos de Price (1976), mais recentemente passou por profundas transformações, tanto que Boaventura de Sousa Santos (2002) chega a afirmar: " [...] os progressos científicos dos últimos trinta anos são de tal ordem dramáticos que o séculos que nos precederam – desde o século XVI, até o século XIX – não são mais do que uma pré-história longínqua". Assim, questões brotam em torno da constituição ou configuração epistêmica atual da Ciência da Informação, ou especialmente sobre algumas metodologias, no caso, a Bibliometria:

- Os princípios, teorias, metodologias e leis da Bibliometria ainda são válidos atualmente? Qual o seu papel na Ciência da Informação contemporânea?
- Como a Bibliometria, aplicada ao mapeamento temático, pode contribuir para a Organização do Conhecimento? E para tesauros?

A Bibliometria repousa nos fundamentos teóricos da Comunicação Científica, cujos alicerces, por sua vez, se firmam na História da Ciência e nos Estudos Sociais da Ciência, constituindo pilares das pesquisas da área. Mais uma vez, o pensamento de Price (1976a, 1976b) norteia as questões relativas ao caráter cumulativo da ciência, expansão da comunidade científica, padrões de produtividade científica, periódicos científicos, seu crescimento e importância. Suas pesquisas continuam a ser a base do estudo e compreensão da Bibliometria, nas redes científicas e sociotécnicas, por ele denominadas "ciência da ciência".

Nas relações epistêmicas no interior da Ciência da Informação, disciplinas ou subáreas se interconectam, da mesma forma que esta área com outros campos da ciência, na sua tão decantada interdisciplinaridade, questão estudada em pesquisas teóricas e empíricas de uma das autoras deste trabalho (PINHEIRO, 1995, 1997, 1998, 2005, 2006, 2007, 2009), o que justifica a incidência de autocitações no presente texto. A partir daí, uma rede metodológica foi tecida e o tesauro em Ciência da Informação foi sendo estruturado e se desenvolveu, numa trajetória de quase 20 anos.

O objetivo desta pesquisa é descrever e analisar as etapas teóricas e metodológicas da construção do Tesauro Brasileiro de Ciência da Informação, que fundamentaram a definição de sua estrutura e termos, na busca de um instrumento que refletisse a configuração atual da área na Sociedade da Informação e do Conhecimento.

Adotamos neste trabalho a definição de tesauro como "um conjunto de termos semântica e genericamente relacionados, cobrindo uma área específica do conhecimento. É um instrumento da indexação/recuperação de informação" (GOMES, 1996). A sua elaboração tanto visou a construir a estrutura disciplinar da Ciência da Informação, numa classificação que representasse a sua constituição epistêmica, com as respectivas disciplinas ou subáreas e seus desdobramentos em temas ou assuntos, quanto servir de instrumento para indexação e recuperação da informação. Sobre as relações entre tesauros, classificação e terminologia GOMES (1996), especialista brasileira que há longos anos estuda estas questões, afirma que "há características comuns entre terminologia e tesauro, ou seja, ambos lidam com termos de uma área específica. Mas Terminologia - assim, com T maiúsculo para designar a área de estudo...". Na elaboração do Tesauro Brasileiro de Ciência da Informação, ao ser tomada como base estrutural uma classificação de Ciência da Informação, assumimos que "a característica comum entre tesauros e terminologias é que são ambos sistemas de classificação"(GOMES,1996).

Nos procedimentos metodológicos para elaboração do Tesauro Brasileiro de Ciência da Informação o recurso foi o mapeamento temático, por meio da análise da literatura e da Bibliometria para chegar às categorias e à classificação, esta inicialmente desenvolvida para a pesquisa de Chaim Zins (2004-2005) e posteriormente disseminada em artigos (2007), e da qual participou a coordenadora do projeto do tesauro e uma das autoras deste trabalho, Lena Vania Pinheiro, conforme será explicitado no tópico 2.3.

Na base de desenvolvimento do tesauro foram adotadas pesquisas teóricas de enfoque epistemológico, na linha da epistemologia histórica, corrente liderada por Bachelard, cuja finalidade é a descoberta da gênese, da estrutura e do funcionamento dos acontecimentos científicos e o delineamento das "relações susceptíveis de existir entre a ciência e a sociedade, entre a ciência e as diversas instituições científicas ou entre as diversas ciências", na construção de uma epistemologia para a produção dos conhecimentos sob todos os aspectos - lógico, ideológico, histórico (JAPIASSU, 1977).

Na Bibliometria adotamos a definição clássica de Pritchard (1969): "todos os estudos que buscam quantificar os processos de comunicação escrita" ou a "aplicação de métodos matemáticos para livros e outros meios de comunicação".

O estudo de estruturas de tesauros nacionais e estrangeiros, a coleta de termos e a análise comparativa da terminologia e de aspectos de tradução, no seu processo evolutivo no tempo, complementaram a metodologia híbrida que deu forma e conteúdo ao tesauro. A atividade relativa à tradução e equivalências terminológicas foi das mais complexas, uma vez que o tesauro apresenta os termos em três línguas - português, inglês e espanhol – e foi permanente, pois perpassou todas as fases de elaboração do tesauro.

É importante ressaltar que embora no IBICT- Instituto Brasileiro de Informação em Ciência e Tecnologia tenham sido empreendidos esforços para a elaboração de tesauros, o que foi construído e finalizado, em 1989, não chegou a ser publicado e circulou de forma restrita.

2. UMA LONGA JORNADA DE CONSTRUÇÃO DE ALICERCES TEÓRICOS E EMPÍRICOS DE UM TESAURO

A concepção do Tesauro Brasileiro de Ciência da Informação foi, conforme já explicitado, iniciada há muito tempo atrás, num longo exercício teórico e empírico que embasou a metodologia híbrida e moldou a sua configuração e conteúdos atuais.

As inquietações teóricas começaram na dissertação de mestrado de Pinheiro, em 1982, tendo por tema a Lei de Bradford, e se estenderam ao doutorado, iniciado em 1991. Durante a pesquisa-tese (PINHEIRO,1997) foi pensada a primeira demarcação epistemológica da Ciência da Informação, ilustrada por uma espécie de mandala, publicada em artigo de Pinheiro e Loureiro (1995), cujo título traduz os propósitos dos autores: Traçados e limites da Ciência da Informação. A figura representava a Ciência da Informação, tal como constituída no Programa de Pós-Graduação em Ciência da Informação (IBICT-UFRJ- Universidade Federal do Rio de Janeiro), com suas respectivas disciplinas ou subáreas, num exercício de identificação de áreas interdisciplinares em cada uma dessas disciplinas, no que Saracevic (1999) denomina "mapeamento de área". Os autores reconheciam e chamavam a atenção de que esta visão representava um Curso de Pós-Graduação, num determinado momento, no Brasil e, portanto, poderia não corresponder a um escopo mais amplo da Ciência da Informação no mundo. As disciplinas ou subáreas identificadas nesse estudo (PINHEIRO e LOUREIRO,1995), num total de doze, foram as seguintes:

Administração de Sistemas de Informação;
Automação;
Bibliometria;
Comunicação Científica e Tecnológica;
Divulgação Científica;
Economia da Informação;
Epistemologia da Ciência da Informação;
Estudos de Usuários;
Informação , Cultura e Sociedade;
Redes e Sistemas de Informação;
Representação da Informação;
Sistemas de Recuperação da Informação.

As pesquisas sobre o domínio epistemológico e a interdisciplinaridade da Ciência da Informação que se seguiram (PINHEIRO,1997, 1998, 2005, 2006,2007, 2009) foram fortalecidas por pesquisa empírica de mapeamento temático da área no exterior, tendo por fonte o ARIST – *Annual Review of Information Science and Technology* e recorrendo à Bibliometria, o que será abordado no próximo tópico.

2.1 Mapeamentos temáticos da Ciência da Informação: análise da literatura estrangeira e definição de categorias

O "mapeamento de literatura" é uma das três mais importantes idéias na área, nascida da exploração dos índices de citação, nos anos 60, de acordo com Saracevic (1999). Segundo o autor, a estrutura da Ciência da informação é constituída por um grande quadro composto, como em qualquer área, por diferentes e mais amplas disciplinas e subdisciplinas de pesquisa e prática. Este pesquisador recorre à metáfora sócio-geográfica para descrever a Ciência da Informação, afirmando que como todas as coisas que se voltam para fora, a Ciência da Informação parece mais com a Austrália - acentuadamente desenvolvida na costa (fronteiras), mas com esparsa delimitação e ocupação interior (SARACEVIC, 1999).

Na tese de doutorado de Pinheiro (1997) foi delineado um mapa epistemológico, desta vez utilizando métodos bibliométricos e o ARIST como fonte, conforme mencionado. O período coberto foi de 1966 a 1995, portanto, 30 anos, e a análise totalizou 307 artigos de revisão, em 30 volumes. Posteriormente, Pinheiro (1997, 1988, 2002, 2006) atualizou a sua pesquisa e cobriu o período de 1996 a 2004, mais nove (9) anos e 81 artigos de revisão. O quadro a seguir reúne os dados gerais das duas pesquisas.

Quadro 1: Mapeamento temático da Ciência da Informação

Fonte/material de pesquisa	Período	Nº Artigos de revisão
ARIST	1966-1995	307
ARIST	1996-004	81
Total	**38 anos**	**388**

O ARIST foi escolhido como fonte da pesquisa por sua legitimidade na área, pelos seus editores, nomes respeitados como Martha Williams e Carlos Cuadra, posteriormente Blaise Cronin, e autores do mesmo quilate intelectual, em geral convidados, e pela entidade editora, inicialmente a ADI - American Documentation Institute, depois denominada ASIS - American Society for Information Science e finalmente ASIST - American Society for Information Science and Technology. Por outro lado, o conteúdo de artigos de revisão que também de certa forma efetuam o mapeamento de questões e assuntos de uma área, num determinado período, são adequados ao caráter da pesquisa que estava sendo empreendida, por propiciarem uma visão mais completa e integrada de um campo do conhecimento.

Neste ano foi anunciado que o ARIST, a partir de 2011 não será mais publicado em papel e que as revisões de literatura serão mais concisas e incluídas no JASIS - *Journal of American Society for Information Science and Technology*. Sobre o ARIST e sua origem editorial, que poderia refletir uma concepção norte-americana e, portanto, parcial da Ciência da Informação, é oportuno introduzir o pensamento de Saracevic (1992), para quem a área tem natureza internacional, uma vez que "a justificativa básica e conceitos são os mesmos globalmente", embora em cada país as prioridades no seu desenvolvimento e evolução possam ser diferentes.

Assim sendo, independente da possível limitação do ARIST, da qual, aliás, qualquer fonte não escaparia, este aspecto não chega a atingir a sua relevância e representatividade na área, inclusive como fonte ou material de pesquisa, o que pode ser constatado pela manifestação de teóricos como o dinamarquês Birger Hjorland (*JASIS,* May, 2000) que o considera "... a mais importante fonte de estados da arte na Ciência da Informação" ou Laurence P. Heilprin (*JASIS,* July, 1988), para quem o ARIST "ocupa um lugar único e impar" (apud ARIST http://www.asis.org/Publications/ARIST/).

No primeiro mapeamento temático, cobrindo o período 1966-1995, Pinheiro (1997) apresentou os resultados, alcançados após a leitura e análise de conteúdo dos artigos de revisão, o que incluiu os títulos, os resumos, objetivos e o texto completo, este último quando necessário, para que fossem definidos os temas, orientados por uma concepção de área na sua constituição disciplinar, epistemológica, e não de assuntos. Naturalmente, este exercício é marcado por seus propósitos e subjetividade, uma vez que a definição de temas foi mais ampla (macro) e não específica, detalhada (micro). Além disso, numa decisão metodológica necessária à tese, não foram incluídos os temas com freqüência abaixo de oito (8), muito dispersos e que corresponderiam à zona de baixa frequência, embora pudessem representar novos temas eclodindo na Ciência da Informação.

O resultado deste primeiro mapeamento temático é mostrado no *Quadro 2*.

Quadro 2. – Temas de artigos de revisão do ARIST, 1966-1995, por ordem de freqüência

Tema/Assunto	Artigos de revisão nº absoluto	%
1. Sistemas de informação	43	14
2. Tecnologia da informação	28	9,12
3. Disseminação da informação	27	8,79
4. Políticas de informação	23	7,49
5. Necessidades e usos de informação	22	7,16
6. Sistemas de recuperação da informação	20	6,51
7. Computadores e programas	19	6,18
8. Representação da informação	16	5,21
9. Automação de bibliotecas	15	4,88
10.Redes de informação	14	4,56
11.Formação e aspectos profissionais	14	4,56
12.Bases de dados	13	4,23
13.Organização e processamento da informação	13	4,23
14.Administração da informação	12	3,90
15.Teoria da Ciência da informação	11	3,58
16.Processamento automático de linguagem	9	2,93
17.Economia da informação	8	2,60
Total	**307**	**99,93**

Fonte: Pinheiro, Lena Vania Ribeiro. A Ciência da Informação entre sombra e luz... (1997)

As questões terminológicas são importantes e o *Quadro* já enseja algumas observações que ficam mais claras no *Quadro 3*, com o surgimento de novas questões, como uma mudança de terminologia e até de enfoque, conforme veremos a seguir. A primeira observação é sobre Pinheiro (1997), que deixa de denominar "temas" e passa a adotar "disciplinas", mais expressiva considerando seus propósitos e enfoque, a fim de enfatizar o caráter epistemológico do mapeamento, de constituição disciplinar da área, o que na tese foi chamado "domínio epistemológico".

*Quadro 3: Disciplinas segundo artigos de revisão do ARIS por ordem de freqüência. 1996-2004 **

Disciplinas	Nº	%
1. Sistemas de recuperação da informação	15	18,51
2. Representação da informação	9	11,11
3. Tecnologia da informação	8	9,87
4. Sistemas de informação	6	7,40
5. Bibliometria	6	7,40
6. Inteligência competitiva e Gestão do conhecimento	5	6,17
7. Mineração de dados ("data mining")	5	6,17
8. Política de informação	5	6,17
9. Teoria da Ciência da Informação	5	6,17
10. Comunicação científica eletrônica	3	3,70
11. Necessidades e usos da informação	3	3,70
12. Administração de informação	2	2.50
13. Bibliotecas digitais	2	2,50
14. Economia da informação	2	2.50
15. Formação e aspectos profissionais	2	2.50
16. Processamento automático de linguagem	2	2.50
17 Bases de dados	1	1,23
Total	**81**	

Fonte: Pinheiro, Lena Vania R. Ciência da Informação: questões sobre formação, ensino e pesquisa (coluna). 2002
*** Os dados dos anos de 2003 e 2004 foram extraídos dos " print preview" do ARIST, disponíveis na Internet em 2002, data da publicação.*

A primeira observação é sobre Pinheiro (1997), que deixa de denominar "temas" e passa a adotar "disciplinas", mais expressiva considerando seus propósitos e enfoque, a fim de enfatizar o caráter epistemológico do mapeamento, de constituição disciplinar da área, o que na tese foi chamado "domínio epistemológico".

A exclusão de artigos de revisão sobre computadores e programas, que deixam de ser apresentados, é indicadora de que este assunto, que enfatizava a "máquina pela máquina", como fim, o recolocou no seu devido lugar, de meio, de instrumento, melhor exemplificado por mineração de dados, ao mostrar exatamente esta função.

Inversamente, é constatada a inclusão de novas disciplinas como Inteligência competitiva e Gestão do conhecimento, Mineração de dados ("data mining") e Bibliotecas Digitais, emergentes e impulsionadas pela globalização e tecnologias da informação e comunicação -TICs, particularmente a Internet/Web, que marcam a nova era da Sociedade da Informação e do Conhecimento.

A inserção de Inteligência Competitiva e Gestão do Conhecimento, juntos, não significa que sejam sinônimos, embora profundamente interdisciplinares, tanto que nessa ocasião não foram estabelecidas distinções. Posteriormente estas disciplinas foram objeto de um capítulo de coletânea, intitulado "Inteligência competitiva como disciplina da Ciência da Informação e sua trajetória e evolução no Brasil" (PINHEIRO, 2005). Neste texto, no tópico específico sobre questões conceituais e evolução terminológica são expostas as variações terminológicas em diferentes países, bem como sua evolução. Como não é objetivo deste trabalho, neste momento, discutir a distinção entre as duas disciplinas, citamos os conceitos sintetizados por Canongia e colaboradores (2004 apud Pinheiro, 2005): "...a Gestão do Conhecimento promove a codificação e a circulação do conhecimento internamente, enquanto a Inteligência Competitiva fornece meios para adquirir conhecimento sobre o ambiente externo, conhecimento esse que pode ser, em grande parte, introduzido na rede interna de circulação".

Quanto às Bibliotecas Digitais, também há controvérsias conceituais em relação a outros termos similares, sobretudo bibliotecas virtuais, e ainda não foi atingida a consolidação conceitual.

Uma série de análises sobre questões terminológicas e de tradução foram naturalmente sendo efetivadas, mesmo antes do projeto do tesauro ser pensado ou iniciado, fundamentais para sedimentar o seu terreno semântico. A Administração da Informação, por exemplo, foi gradativamente substituída por Gestão da Informação e posteriormente ampliada por Gestão do Conhecimento, disciplina cuja denominação tem sido muito questionada, inclusive por uma das autoras deste trabalho: "Gestão do conhecimento é um conceito incompleto e incorreto, no primeiro caso porque trata de gestão do conhecimento nas (e das) organizações e, no segundo, porque informação não é sinônimo de conhecimento e o correto, portanto, seria gestão da informação nas organizações ou empresas" (PINHEIRO, 2002).

Não pode deixar de ser explicada a presença, somente no segundo levantamento (1996-2004), da Bibliometria, cuja existência data do início do século 20, por motivo do corte ter sido na frequência 8 (oito), conforme mencionado, e a Bibliometria ter obtido freqüência 2 (dois), no primeiro mapeamento (1966-1995). Cabe explicitar, ainda, que o termo adotado no ARIST foi mantido na pesquisa, de Bibliometria, primeiro e mais antigo, mesmo com a introdução dos novos termos Informetria e Webmetria, além de outras métricas, como Cientometria. Tal como na interdisciplinaridade, aqui temos também uma "família" de termos, neste caso, unidos pelo sufixo metrias, sobre cuja genealogia não há consenso qual o termo mais abrangente, para alguns autores Cientometria, para outros Informetria.

Ao final dos resultados empíricos, Pinheiro (1997) criou categorias que pudessem reunir disciplinas ou temas por sua natureza, não aprofundadas na tese, mas essenciais para a concepção da estrutura do tesauro e que foram desenvolvidas e desmembradas quando participou da pesquisa de Zins (2004-2005), a ser abordada no tópico 2.3, e construiu uma classificação para a Ciência da Informação:

- disciplinas estruturais;
- disciplinas de representação ou instrumentais ;
- disciplinas gerenciais ;
- disciplinas tecnológicas ; e
- disciplinas sócio-culturais ou de transferência da informação.

Nas disciplinas estruturais seriam incluídas a Teoria e Epistemologia da Ciência da Informação, a Bibliometria, ensino e pesquisa em Ciência da Informação e as diversas teorias como a da informação, a teoria geral de sistemas e assim por diante.

Corresponderiam às disciplinas de representação, "as relativas à descrição, classificação e indexação, metodologias de tesauros e vocabulários controlados, normas e padrões nacionais e internacionais de intercâmbio da informação", entre outras.

Como disciplinas gerenciais Pinheiro (1997) destacou: administração de redes e sistemas de informação, organização e processamento da informação, gestão da informação, economia da informação, sistemas gerenciais de informação (MIS).

As disciplinas tecnológicas abrangeriam "automação, sistemas de informação, redes de comunicação e informação, tecnologia de informação, produção e acesso a bancos e bases de dados, sistemas de recuperação da informação".

E, finalmente, entre as disciplinas sócio-culturais ou de transferência da informação seriam consideradas: "necessidades e usos de informação, comunicação científica, tecnológica, artística e cultural, política de informação, divulgação científica e disseminação da informação". (PINHEIRO,1997)

Na elaboração do tesauro esses temas ou disciplinas foram aglutinados em categorias mais amplas, e em dois casos foram desmembradas, o que será detalhado no ítem 4.

Na sequência de pesquisas empíricas descritas e analisadas foi empreendido um estudo similar sobre a produção científica brasileira da área e, neste caso, utilizando como fonte um periódico nacional, objeto do próximo tópico.

2.2 Mapeamento temático da Ciência da Informação: análise da literatura brasileira

Após os mapeamentos da literatura estrangeira, quase que se impôs a replicação da pesquisa voltada à produção científica nacional. A fonte escolhida foi o periódico *Ciência da Informação*, entre outros publicados no Brasil, considerando a sua legitimidade, apoiada nos seguintes aspectos: a instituição editora, o IBICT, a sua existência desde 1972, sem interrupções, e o seu caráter científico assegurado pelo comitê editorial, editor científico e avaliadores, incumbidos do processo de avaliação dos artigos submetidos à publicação na revista.

Este mapeamento ocorreu como parte de pesquisa apoiada pelo CNPq[4] e foi publicado no artigo "Ciência da Informação: 32 anos (1972-2004) no caminho da história e horizontes de um periódico científico brasileiro (PINHEIRO; BRASCHER; BURNIER, 2005). Os resultados são mostrados no *Quadro 4*, a seguir.

Da mesma forma que o mapeamento no ARIST foi desenvolvido em duas etapas, o que possibilitou visualizar o movimento das disciplinas no tempo, o da revista Ciência da Informação foi distribuído por décadas e ambos abrangeram um período suficientemente longo, permitindo um acompanhamento e inferências devidamente respaldados.

Os temas/disciplinas praticamente se repetem, tanto na literatura estrangeira quanto na brasileira. Numa análise ampla dos resultados, podemos ratificar as afirmativas de Saracevic (1999) já mencionadas, de que na Ciência da Informação, embora internacional, há prioridades que diferem em cada região ou país. Assim é que, enquanto nos resultados do ARIST predominam as preocupações tecnológicas, no Brasil a ênfase está nos estudos teóricos, sociais e políticos, fruto das necessidades de países em desenvolvimento, em todos esses aspectos.

Por outro lado, certos resultados estão relacionados à pós-graduação do IBICT como, por exemplo, o alto índice de artigos sobre Bibliometria, uma vez que esta foi uma questão muito estudada desde o início do mestrado, em 1970, pela introdução dessa disciplina por Saracevic, professor do Curso e orientador das primeiras dissertações, juntamente com outros professores estrangeiros, como F. Wilfrid Lancaster (PINHEIRO; BRASCHER; BURNIER, 2005 c).

[4] PINHEIRO, Lena Vania Ribeiro. A Ciência da Informação no Brasil: historiografia de uma área do conhecimento contemporânea no cenário nacional. Rio de Janeiro:CNPq,2002-2005,2006-2009.

Quadro 4.- Freqüência de temas de artigos da Ciência da Informação distribuídos por décadas

Temas	1972-79	%	1980-89	%	1990-99	%	2000-04	%	Total	%
1. Teoria da Ciência da Informação	4	4,65	8	7,33	23	9,2	14	9,45	49	8,26
2. Bibliometria	18	20,93	11	10,09	14	5,6	5	3,37	48	8,09
3. Representação da informação	8	9,3	9	8,25	15	6	14	9,45	46	7,75
4. Políticas de informação	3	3,48	11	10,09	12	4,8	17	11,48	43	7,25
5.Necessidades e usos de informação	6	6,97	9	8,25	16	6,4	11	7,43	42	7,08
6. Gestão da informação	2	2,32	7	6,42	25	10	7	4,72	41	6,91
7. Comunicação científica	7	8,13	10	9,17	12	4,8	10	6,75	39	6,57
8.Tecnologias da informação	-		1	0,91	18	7,2	18	12,16	37	6,23
9. Formação e aspectos profissionais	6	6,97	9	8,25	15	6	4	2,7	34	5,73
10. Sistemas e redes de informação	6	6,97	6	5,5	17	6,8	3	2,02	32	5,39
11. Disseminação da informação	4	4,65	7	6,42	16	6,4	4	2,7	31	5,22
12. Sistema de recuperação da informação	4	4,65	2	1,83	6	2,4	9	6,08	21	3,54
13. Bibliotecas virtuais / digitais	-		-	-	6	2,4	11	7,43	17	2,86
14. Inteligência competitiva	-		-	-	8	3,2	5	3,37	13	2,19
15. Política de Ciência e Tecnologia	-		3	2,75	7	2,8	2	1,35	12	2,02
16. Bases de dados	3	3,48	4	3,66	2	0,8	2	1,35	11	1,85
17. Organização e processamento da informação	6	6,97	1	0,91	1	0,4	2	1,35	10	1,68
18. Economia da informação	-		2	1,83	5	2	2	1,35	9	1,51
19. Biblioteconomia/Bibliotecas/Livros	2	2,32	3	2,75	-	-	2	1,35	7	1,18
20. Arquivologia	-		-	-	6	2,4	-		6	1,01
21. Processamento automático da linguagem	1	1,16	1	0,91	4	1,6	-		6	1,01
22. Automação de biblioteca	2	2,32	1	0,91	2	0,8	1	0,67	6	1,01
23. Divulgação científica	-	-	1	0,91	2	0,8	1	0,67	4	0,67
24. Ciência e Tecnologia	-	-	-	-	3	1,2	-		3	0,5
25. Gestão do conhecimento	-	-	-	-	-	-	3	2,02	3	0,5
26. Política editorial	2	2,32	1	0,91	-		-	-	3	0,5
27. Sistemas especialistas	-	-	-	-	2	0,8	-	-	2	0,33
28. Comunicação social	-	-	-	-	2	0,8	-	-	2	0,33
29. Imprensa	-	-	-	-	2	0,8	-	-	2	0,33
30. Lingüística	-	-	-	-	2	0,8	-	-	2	0,33
31. Tecnologias	-	-	-	-	2	0,8	-	-	2	0,33
Temas com freqüência 1 no somatório geral	2	2,32	2	1,83	5	2	1	0,67	10	1,68
Total	**86**	**100**	**109**	**100**	**250**	**100**	**148**	**100**	**593**	**100**

Fonte: Pinheiro, Brascher, Burnier (2005)

Diferentemente dos resultados do ARIST, nos quais aparecem juntos Gestão do Conhecimento e Inteligência Competitiva, aqui é assumida a sua identidade disciplinar, por nessa ocasião já haver, no Brasil conhecimento e experiência, principalmente em decorrência dos Cursos de Especialização em Inteligência Competitiva e dos Workshops Brasileiros de Inteligência Competitiva e Gestão do Conhecimento, realizados desde os meados da década de 1990 e geradores de pesquisas nessas subáreas. (PINHEIRO, 2005 a).

Outra particularidade brasileira percebida é a presença da Divulgação Científica, que não consta nos resultados do exterior. No Brasil, especificamente no PPGCI- Programa de Pós-Graduação em Ciência da Informação, do IBICT-UFRJ pode ter sido consequência de um desdobramento ou extensão da disciplina Comunicação Científica, além da presença de mestrandos e doutorando oriundos da Comunicação. Além desta, a inserção da Biblioteconomia e da Arquivologia, campos interdisciplinares à Ciência da Informação, ausentes no mapeamento do exterior, o que pode ter sido ocasionado pela não existência, em nosso País, de mestrados e doutorados específicos nessas áreas e a procura e o acolhimento natural em programas de pós-graduação em Ciência da Informação, de alunos com formação nesses dois campos do conhecimento.

Algumas mudanças terminológicas também aparecem, como a adoção de gestão (Gestão da Informação), no lugar de administração. Inversamente, foi mantida a Representação da Informação, quando a Organização do Conhecimento já galgava seu status disciplinar e reunia uma mescla ou emaranhado terminológico, difícil de separar e tratar autonomamente, o que foi sentido em algumas das categorias do tesauro e exigiu muitas leituras discussões.

Nesse sentido, a consulta a tesauros e dicionários, no caso, os selecionados para a pesquisa, foram essenciais diante dos impasses e necessárias decisões, bem como a discussão com especialistas. Esta parte do processo de elaboração do tesauro é descrita no tópico 3.

2.3 Classificação para pesquisa de Chaim Zins

Em 2004, Chaim Zins, da University of Haifa, em Jerusalém, iniciou uma pesquisa denominada "knowledge map of information science: issues, principles, implications", com a participação de uma das autoras deste trabalho, Pinheiro, selecionada entre os três brasileiros e mais de 40 pesquisadores da comunidade acadêmica da área, de diferentes países, para responder a um questionário, aplicado em três etapas,como parte da técnica adotada nessa investigação, a crítica Delphi (ZINS, 2005).

A pesquisa empreendida por Zins teve como um dos objetivos "clarificar as diferentes concepções de Ciência da Informação" e abrangeu os principais conceitos - dado, informação conhecimento, ciência da Informação - também estudados por Pinheiro (1997). O pesquisador visava, ainda, a "desenvolver um mapa do conhecimento amplo, sistemático e cientificamente válido do domínio do conhecimento da Ciência da Informação" e "[...] fundamentar este mapa em sólidas bases teóricas" (ZINS, 2007 a,b). O objetivo é semelhante ao de Pinheiro na sua tese (1997), na qual pretendia identificar as disciplinas formadoras da Ciência da Informação, no traçado do "domínio epistemológico", com o delineamento de suas internalidades e externalidades, citada anteriormente.

Na etapa final da pesquisa, Zins solicitou aos participantes que elaborassem uma classificação da área, o que levou Pinheiro à retomada das grandes categorias definidas em sua tese e descritas no tópico 2.1.

Como resultado da participação nessa pesquisa, foi possível ampliar o número de disciplinas incluídas em cada categoria, não ultrapassando três níveis hierárquicos. Num exercício daí decorrente, as disciplinas estruturais foram reunidas em Fundamentos da Ciência da Informação; as instrumentais passaram a ser representadas por Organização e Processamento da Informação; as disciplinas gerenciais ganharam o nome de Gestão da Informação; as tecnológicas foram agrupadas pelas Tecnologias da Informação; e as disciplinas sócio-culturais ficaram identificadas por Transferência da Informação.

3. ANALISE COMPARATIVA COM OUTROS TESAUROS E CONSULTA A DICIONÁRIOS E ESPECIALISTAS

Conforme mencionado no início deste trabalho, no IBICT (1989) foi elaborado um tesauro, em versão preliminar e não publicado, que foi um dos consultados para a construção do atual, além do tesauro da ASIST (2005) e o latinoamericano (1999). Portanto, para esta fase de coleta de termos e análise comparativa foram selecionados três tesauros, um brasileiro e dois estrangeiros (um norte-americano e outro latino-americano) e dois dicionários, um brasileiro e outro americano.

Os tesauros foram os seguintes:

- INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. *Tesauro de ciência da informação*: versão preliminar. Brasília : IBICT, 1989. 1 v.;
- NAUMIS PEÑA, Catalina et al. *Tesauro lationamericano en ciência bibliotecológica y de la información* (TELACIBIN). México : Universidad Nacional Autônoma de México, 1999. 307 p. [5];
- REDMOND-NEAL, Alice ; HLAVA, Marjorie M.K. (Org.) *ASIS&T thesaurus of information science, technology and librarianship*. 3. ed. Medford, N.J : publicado para American Society for Information Science and Technology pela Information Today, Inc., c2005. 255 p.
- Os dicionários brasileiro e o norte-americano, este online, são:
- CUNHA, Murilo Bastos da ; CAVALCANTI, Cordélia Robalinho de Oliveira. *Dicionário de biblioteconomia e arquivologia*. Brasília : Briquet de Lemos / Livros, 2008. 451 p.;
- REITZ, Joan M. *ODLIS*: online dictionary for library and information science. Danbury : Western Connecticut State University, 1994- Disponível em: http://lu.com/odlis/ .

É oportuno ressaltar que estas fontes, tal como as pesquisas empíricas, cobrem um período de tempo longo, que vai desde o tesauro brasileiro, de 1989, ao dicionário brasileiro de 2008, quase 20 anos.

Para compreensão dos diferenciados enfoques dos tesauros adotados como fonte e o tesauro em construção, é fundamental reproduzir as respectivas estruturas.

3.1 Tesauro do IBICT, 1989

O Tesauro de Ciência da Informação do IBICT, na versão preliminar, mantida até hoje, é composto por sete (7) categorias, relacionadas a seguir.

A - INFORMAÇÃO: aspectos históricos, teóricos, sociais, legais e filosóficos relacionados à informação, disciplinas e tipos de informação.

B - DOCUMENTO: todo tipo de registro da informação apresentada em qualquer suporte impresso, visual, auditivo, táctil, eletrônico, magnético, etc.

C - UNIDADE DE INFORMAÇÃO: instituições voltadas para a aquisição, processamento, armazenamento e disseminação de informações.

D - PLANEJAMENTO, ORGANIZAÇÃO E ADMINISTRAÇÃO DE UNIDADES DE INFORMAÇÃO: atividades relacionadas com a concepção e funcionamento de unidades de informação.

E - PROCESSOS E SERVIÇOS DE INFORMAÇÃO: atividades de armazenamento, tratamento e recuperação da informação e os serviços originados desses processos envolvendo instrumentos e técnicas utilizados.

[5] Este tesauro foi gentilmente doado pelo Prof. Ulf Baranov, como contribuição ao tesauro.

F - TRANSFERÊNCIA E USO DA INFORMAÇÃO: aspectos relacionados ao estudo do fenômeno da comunicação da informação desde sua geração à sua utilização.

G - PROFISSÃO: aspectos relacionados ao ensino, formação, treinamento, ética e atuação do profissional de informação.

Embora com enfoque diferente no estabelecimento da estrutura, há coincidência de algumas categorias como a de informação e de transferência e uso da informação, e naturalmente estão ausentes algumas questões surgidas a partir da década de 1990, como gestão da informação estratégica.

3.2 Tesauro da ASIST, 2005

O tesauro da ASIST foi publicado em primeira edição no ano de 1994 e a versão utilizada é a terceira, de 2005. Logo que foi lançado, este tesauro podia ser acessado pela Internet, mas posteriormente a ASIST deixou de disponibilizar a versão eletrônica.

O tesauro apresenta a seguinte estrutura básica, com suas respectivas categorias (*Thesaurus headings*):

- PEOPLE AND ORGANIZATIONS
  - *(organizations)*
  - *(persons and informal groups)*
  - *(product and service providers)*
- ACTIONS, EVENTS, AND PROCESSES
  - *(activities and operations)*
  - *(natural processes and events)*
- PHYSICAL OBJECTS
  - *(buildings and facilities)*
  - *(communications networks)*
  - *(hardware, software, and equipment)*
  - *(physical media)*
- THEORETICAL CONCEPTS AND INFLUENCES ON INFORMATION
  - *(attributes)*
  - *(fields and disciplines)*
  - *(language)*
  - *(sociocultural aspects)*
- INFORMATION, INFORMATION DELIVERY FORMATS AND CHANNELS
  - *(communications media)*
  - *(document types)*
  - *(knowledge and information)*
- METHODS OF STUDY
  - *(research and analytic methods)*
- GEOGRAPHIC INFORMATION
  - *(countries and regions)*

A mesma observação feita em relação ao tesauro brasileiro do IBICT, de 1989, aqui se impõe, de abordagem distinta também do tesauro americano da ASIST, na concepção e, consequentemente, de aglutinação dos assuntos, embora sempre apareçam pontos de convergência com o tesauro brasileiro objeto deste trabalho.

Esta análise foi essencial para percepção das mutações de um termo no tempo, o desmembramento ou fusão de questões, além de peculiaridades de tradução. Nesse sentido, por ser um tesauro trilingue, foi possível aprofundar as razões de enfoques distintos e traduções específicas para um mesmo termo, o que reflete mais do que um problema de tradução, de cultura e de prioridades e peculiaridades do desenvolvimento da área num determinado país.

Simultaneamente a estas atividades, concretizadas durante a elaboração do tesauro, por categorias, foram consultados professores, pesquisadores e profissionais de informação, alguns acompanhando o processo como um todo, como Gilda Maria Braga (IBICT), Hagar Espanha Gomes, hoje consultora independente, e Marisa Brascher (Universidade de Brasília/UnB e IBICT), responsável também pelo treinamento da equipe do tesauro no software adotado, o MultiTes. Outros professores contribuíram na sua respectiva área de pesquisa, como Kira Tarapanoff (UnB), Maria Nélida González de Gomez e Rosali Fernandez de Souza, ambas do IBICT. A construção do tesauro contou, ainda, com a participação de especialistas em determinadas questões, como em diretos autorais e propriedade intelectual (Jaury Nepomuceno de Oliveira[6]) e em tecnologias de informação e comunicação – TICs, Linair Maria Campos[7], além de Sonia Burnier, bibliotecária do IBICT, que participou em diferentes momentos da construção do tesauro. O MultiTes é um software para a construção e publicação de tesauros, desenvolvido e aprimorado pela companhia norte-americana Multisystems, desde 1983. Este programa gera um arquivo HTML que permite a navegação no tesauro e sua publicação na Web. A cada nova versão, esse arquivo deverá ser gerado e o tesauro ficará disponível na Internet. Uma das vantagens do MultiTes é que permite a saída dos dados em diferentes formatos impressos, antes da publicação da cópia final. A alimentação de dados do Tesauro no MultiTes é feita por Miguel Amorim Neto[8] e colaboraram, em diferentes aspectos desse programa, Dilza Fonseca da Motta[9] e Rosan Coutinho Tavares[10] .

4. ESTRUTURA DO TESAURO BRASILEIRO DE CIÊNCIA DA INFORMAÇÃO, 2010

O Tesauro Brasileiro de Ciência da Informação apresenta as seguintes categorias:

- 1 EPISTEMOLOGIA DA CIÊNCIA DA INFORMAÇÃO
  - História da Ciência da Informação
  - Teorias
  - Métodos de Pesquisa e Análise
    - Metrias da informação e comunicação/Bibliometria
  - Ensino e Pesquisa em Ciência da Informação e Áreas Afins.
  - Aspectos Profissionais da Ciência da Informação e Áreas Afins

[6] Advogado, funcionário da Fundação Biblioteca Nacional,no Escritório de Direitos Autorais, e atualmente doutorando do PPGCI IBICT-UFRJ
[7] Formada em Tecnologia de Processamento de Dados, exerce funções no NCE -Núcleo de Computação Eletrônica da UFRJ, é atualmente doutoranda da PPGCI IBICT-UFF (convênio já extinto, em fase de conclusão de teses)
[8] Bolsista de iniciação científica do CNPq, hoje já formado em Biblioteconomia pela UNIRIO.
[9] Bibliotecária e consultora independente.
[10] Técnico de Informática da CPMBraxis lotado no IBICT/RJ.

- ORGANIZAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO
  - Organização da Informação
    - Sistemas de Organização do Conhecimento
  - Recuperação da Informação
- GESTÃO DA INFORMAÇÃO
  - Gestão de Unidades e Recursos de Informação
    - Gestão de coleções
  - Serviços de Informação e de Biblioteca
- INFORMAÇÃO ESTRATÉGICA NAS ORGANIZAÇÕES
  - Inteligência Competitiva
  - Gestão do Conhecimento
- TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO
  - Aplicações do computador
    - Automação de Bibliotecas
    - Comunicações Mediadas por Computador
    - Inteligência Artificial
    - Mineração de Dados
  - Bases de Dados
  - Bibliotecas Digitais e Virtuais
  - Redes de Informação
    - Internet/Web
  - Preservação de Objetos Digitais
  - Equipamento e Programas de Computador
  - Atributos de Sistemas e de Equipamento
- INFORMAÇÃO COMO ENTIDADE: suportes, tipos e aplicações
  - Tipos de Documentos
  - Conteúdo da Informação
- TRANSFERÊNCIA, ACESSO E USO DA INFORMAÇÃO
  - Comunicação Científica e Tecnológica
  - Avaliação e Editoração de Periódicos Científicos
  - Aspectos Legais, Políticos, Socioeconômicos e Culturais

Nesta forma final da estrutura do tesauro, algumas observações se tornam necessárias, que melhor explicam as suas mutações. Por exemplo, a primeira categoria, que reúne as disciplinas ou subáreas da Ciência da informação denominadas originalmente estruturais, na tese de Pinheiro (1997), passaram a ser reunidas em fundamentos da Ciência da Informação e, finalmente, Epistemologia da Ciência da Informação. Esta categoria engloba as

qualidades científicas da área, especialmente as teorias, métodos e leis, com destaque para a Bibliometria, até chegar à prática, centrada nos aspectos profissionais.

As disciplinas inicialmente de "representação e instrumentais" foram reunidas em "organização e recuperação da informação", e incluem a informação e organização do conhecimento, na sua condição de sistema.

A gestão da informação representa as "disciplinas gerenciais" e, com o surgimento de novas subáreas, como as da Inteligência Competitiva e Gestão do Conhecimento, foi criada mais outra categoria para congregá-las, intitulada informação estratégia nas organizações.

As disciplinas tecnológicas, assim chamadas por Pinheiro (1997) foram ampliadas, em função dos diferentes instrumentos da Internet/Web, para tecnologias de informação e comunicação - TICs e abrangem desde aplicações do computador às bibliotecas digitais e virtuais, e a preservação de objetos digitais.

Uma nova categoria se impôs, cujo conteúdo antes era enquadrado na primeira, de Epistemologia, pelo enfoque diferenciado e necessário, reunindo os suportes, tipos e aplicações da informação como entidade.

As antigas disciplinas "sócio-culturais ou de transferência da informação" (Pinheiro, 1997) receberam a denominação de "transferência, acesso e uso da informação", com ênfase na Comunicação Científica e nos periódicos, inclusive eletrônicos. Finalmente, uma subcategoria de aspectos legais, políticos, socioeconômicos e culturais foi necessária para agrupar aspectos mais amplos, macros, de questões mais relacionadas ao tema transferência, uso e acesso à informação.

5. DAS PRIMEIRAS REFLEXÕES SOBRE O DESAFIO DA CONSTRUÇÃO DE UM TESAURO BASILEIRO DE CIÊNCIA DA INFORMAÇÃO

A construção de um tesauro, por sua complexidade e múltiplas perspectivas combinatórias ou associativas de termos, chega a ter o mesmo fascínio de um caleidoscópio. A cada olhar distinto, novos conjuntos de termos podem ser formados, ainda que tenha sido estabelecida como norteadora uma determinada abordagem, no caso do Tesauro Brasileiro de Ciência da informação, a epistemológica, que representa uma classificação de área e, portanto, um reflexo de seus fundamentos científicos. O método para configuração do tesauro foi, como explicitamos, o mapeamento temático com base na análise da literatura e aplicação da Bibliometria.

A existência de princípios e normas para a elaboração de tesauros, como as relações semânticas entre termos genéricos, específicos e termos relacionados, na sua objetividade, não afasta a subjetividade e, consequentemente, a vulnerabilidade de interpretação ou apreensão de idéias. A subjetividade que envolve uma tarefa desta natureza, sobretudo por lidar com linguagens, por si só é uma difícil empreitada, porque trata de signos, que já carregam tanto o processo cognitivo como a abstração, o que é melhor compreendido pela Filosofia, especialmente Aristóteles (383 a.C.-322 a.C.).

No caso de linguagem documentária de uma área, a Ciência da Informação, representativa do discurso de sua comunidade científica, com todas as implicações de conceitos, metodologias e teorias, nas suas distintas correntes e influências, esta linguagem faz transparecer, ainda, a interdisciplinaridade que a caracteriza e os problemas conceituais e de tradução, particularmente por ser um instrumento trilingue.

A complexidade aparece especialmente na relação todo e parte. A hierarquia é clara quando se trata de gênero e espécie (tipo de), mas na feitura de um tesauro é mais difícil estabelecê-la como hierarquia, na esquematização de coisas ou objetos correlatos.

Tomemos como exemplo a gestão de coleção no tesauro da ASIST, referente a tudo que está relacionado a desenvolvimento de coleção, descarte, seleção e aquisição de documentos, o que não seria exatamente uma hierarquia "tipo de" e sim "parte de". Na estrutura do Tesauro Brasileiro de Ciência da Informação estão presentes tanto as relações partitivas (parte de) quanto as de correlação (tipo de), o que exigiu conhecimento da área como um todo para a sua compreensão, também necessário ao indexador que o utilizará como instrumento.

Outra especificidade relativa a tesauros é a possibilidade da inserção de termos em duas diferentes categorias, o que dependerá de sua abordagem. Uma demonstração deste caso são os termos precisão, revocação e relevância, que tanto aparecem em Epistemologia da Ciência da informação, na qualidade de conceitos e como tal são considerados, como na condição de medidas, assim inseridos na recuperação da informação.

Enfim, reconhecemos que existem ainda múltiplas questões a serem enfocadas, mas o objetivo deste trabalho é direcionado à concepção e fundamentos do Tesauro Brasileiro de Ciência da Informação, suficientemente rico nos seus encadeamentos e articulações infindáveis, tanto quanto as discussões, formulação e reformulações que se impõem. Afinal, um tesauro expressa a vida de um campo do conhecimento, com sua árvore genealógica e frutos, em múltiplas composições, próprias de sua história e caráter epistêmico, desde a sua gênese, ao desenvolvimento e avanços até os dias de hoje, dinâmicos pelas influências advindas de diferentes línguas, discursos e culturas.

6. REFERÊNCIAS DO CAPÍTULO 11

ARIST- Annual Review of Information Science and Technology**.** Disponível em: http://www.asis.org/Publications/ARIST/.

GOMES, Hagar Espanha. Classificação, tesauro e terminologia: fundamentos comuns. Rio de Janeiro, 1996. Palestra preparada para as Tertúlias do Departamento de Biblioteconomia da UNIRIO, apresentada em julho de 1996.Disponível em: http://www.conexaorio.com/biti/tertulia/tertulia.htm.

JAPIASSU, Hilton. *Introdução ao pensamento epistemológico*. 2. ed. Rio de Janeiro: F. Alves, 1977. 202 p.

PINHEIRO, Lena Vania Ribeiro. Campo interdisciplinar da Ciência da Informação: fronteiras remotas e recentes. *Investigación Bibliotecológica*, México, v. 12, n. 25, p.132-163, 1998. Disponível em: <http://biblioteca.ibict.br/phl8/anexos/Mexicolena.pdf>. Publicado também no Brasil em coletânea do IBICT.

PINHEIRO, Lena Vania R. Ciência da Informação: desdobramentos disciplinares, interdisciplinaridade e transdisciplinaridade. In: GONZÁLEZ DE GÓMEZ, Maria Nélida; ORRICO, Evelyn Goyannes Dill. (Org.). *Políticas de memória e informação*: reflexos na organização do conhecimento. Natal: Editora Universitária da UFRN/EDUFRN, 2006, p. 111-141.

PINHEIRO, Lena Vania Ribeiro. Ciência da Informação entre sombra e luz: domínio epistemológico e campo interdisciplinar. Rio de Janeiro, 1997. 278 p. Tese (Comunicação e Cultura) UFRJ/ECO. Orientadora: Gilda Braga. Disponível em: http://biblioteca.ibict.br/phl8/anexos/lenavaniapinheiro1997.pdf.

PINHEIRO, Lena Vania Ribeiro. Ciência da Informação: questões sobre formação, ensino e pesquisa (coluna). *DatagramaZero*, Rio de Janeiro, v. 3, n. 5, out. 2002. Disponível em: www.dgzero.org/out02/F_I_com.htm

PINHEIRO, Lena Vania Ribeiro. Configurações disciplinares e interdisciplinares da Ciência da Informação no ensino e pesquisa no Brasil. In: BORGES, Maria Manuel; CASADO, Elias Sanz (Orgs.). *A Ciência da Informação criadora de conhecimento*. Coimbra: Imprensa da Universidade de Coimbra, University Press, 2009. p. 99-111. ISBN: 978-989-26-0014-7.

PINHEIRO, Lena Vania Ribeiro. Inteligência competitiva como disciplina da Ciência da Informação e sua trajetória e evolução no Brasil. In: STAREC, Cláudio; GOMES, Elizabeth B. P. Gestão; CHAVES, Jorge B. L. (Orgs.). *Gestão estratégica da informação e inteligência competitiva*. São Paulo: Saraiva, 2005. p.17-32.

PINHEIRO, Lena Vania Ribeiro. Pilares conceituais para mapeamento do território epistemológico da ciência da informação: disciplinaridade, interdisciplinaridade, transdisciplinaridade e aplicações. IN: BENTES PINTO, V.; CAVALCANTE, L. E.; SILVA NETO, C. (Orgs.). *Abordagens transdisciplinares da ciência da informação*: gêneses e aplicações. Fortaleza: Edições UFC, 2007. p.. 71-104 ISBN 978-85-7282-239-8.

PINHEIRO, Lena Vania Ribeiro.b Processo evolutivo e tendências contemporâneas da Ciência da Informação. *Informação & Sociedade*: estudos, v. 15, n. 1, 2005. Disponível em: http://www.informacaoesociedade.ufpb.br/IS1510501.htm.

PINHEIRO, Lena Vania Ribeiro c; BRASCHER, Marisa; BURNIER, Sonia. Ciência da Informação: 32 anos (1972-2004) no caminho da história e horizontes de um periódico científico brasileiro. *Ciência da Informação*, Brasília, v. 34, n. 3, p. 25-77, set./dez. 2005. Número especial: IBICT 50 anos. Disponível em: http://www.ibict.br/cienciadainformacao.

PINHEIRO, Lena Vania Ribeiro; LOUREIRO, José Mauro Mattheus. Traçados e limites da Ciência da Informação. *Ciência da Informação*, Brasília, v. 24, n. 1, p. 42-53, jan./jul.1995. Disponível em: www.ibict.br/cienciadainformacao.

PRICE, Derek de Solla.a *A ciência desde a Babilônica*. Belo Horizonte, São Paulo: Itatiaia, Ed.Universidade de São Paulo, 1976.189 p. (Coleção O homem e a ciência, 2).

PRICE, Derek de Solla. b *O desenvolvimento da ciência*. Rio de Janeiro: Livros Técnicos e Científicos, 1976. 96 p.

PRITCHARD, A. Statistical bibliography or bibliometrics? *Journal of Documentation*, Bradford, UK, v. 25, n. 4, p. 348-349, Dec. 1969

SANTOS, Boaventura de Sousa. *Um discurso sobre as ciência*s. 13. ed. Porto: Edições Afrontamento, 1987. 59 p.

SARACEVIC, Tefko. Information Science: origin, evolution and relations. In: VAKKARI, Pertti; CRONIN, Blaise (Ed.). Conceptions of Library and Information Science; historical, empirical and theoretical perspectives. In: *Proceedings of the International Conference for the Celebration of 20th anniversary of the Department of Information Studies, University of Tampere, Finland, 1991*. London, Los Angeles: Taylor Graham, 1992. p. 5-27

SARACEVIC, T. Information Science. *JASIS* -Journal of The American Society for Information Science, v. 50, n. 12: p. 1051-1063, 1999.

ZINS, C. a Classification schemes of Information Science: 28 Scholars Map the Field. *Journal of the American Society for Information Science (JASIST)*, 58 (5): p.645-672, March 2007. Disponível em: http://www.success.co.il/is/zins_28schemes.pdf.

ZINS, C.b Conceptions of Information Science *Journal of the American Society for Information Science (JASIST*), 58 (3), p. 335-350 ,2007;

ZINS, Chaim. Knowledge map of information science: issues, principles, implications. Jerusalem, 2005. Disponível em: http://hw.haifa.ac.il/human/hebrew/ISMapWeb.htm.

V. APLICAÇÕES E DESDOBRAMENTOS DA REPRESENTAÇÃO E ORGANIZAÇÃO DA INFORMAÇÃO E DO CONHECIMENTO

Capítulo 12. – Descoberta de Conhecimento em Textos: Conceitos e Convergências – Marcelo SCHIESSL, Hélia de Sousa CHAVES e Marisa BRÄSCHER, *p. 278-293*

Capítulo 13. – A Contemporaneidade do Tema Governo Eletrônico e Perspectivas de Pesquisas – Rinalda Francesca RIECKEN, *p. 294-319*

Capítulo 12
Descoberta de conhecimento em textos: Conceitos e convergências

Marcelo Schiessl[1], **Hélia Chaves**[2] e **Marisa Bräscher**[3]

SUMÁRIO DO CAPÍTULO 12

Resumo do Capítulo 12, *p. 279*
Como citar o Capítulo 12, *p. 279*

1. INTRODUÇÃO DO CAPÍTULO 12, *p.280*
2. ACERCA DA DENOMINAÇÃO, *p.280*
2.1 Text Mining, *p.281*
2.2 Knowledge Discovery in Text, *p.281*
3. DCD, DCT E RECUPERAÇÃO DA NFORMAÇÃO, *p.282*
4. A DESCOBERTA DE CONHECIMENTO EM DADOS (DCD), *p.283*
4.1 O processo da DCD, *p.284*
5. A DESCOBERTA DE CONHECIMENTO EM TEXTOS (DCT), *p.285*
6. CONSIDERAÇÕES SOBRE A DCT, *p.287*
7. APLICAÇÕES DA DCT, *p.288*
8. ANÁLISE DE AGRUPAMENTOS (*CLUSTERING*) – UMA DESCOBERTA INUSITADA, *p.289*
9. CONSIDERAÇÕES FINAIS DO CAPÍTULO 12, *p.292*
10. RFERÊNCIAS DO CAPÍTULO 12, *p.292*

[1] Mestre em Ciência da Informação. Programa de Pós-Graduação em Ciência da Informação (PPGCInf); Faculdade de Ciência da Informação da Universidade de Brasília (UnB/FCI). CV Lattes: http://lattes.cnpq.br/4402964117780625. E-mail: marcelo.schiessl@gmail.com.
[2] Mestre em Ciência da Informação. Programa de Pós-Graduação em Ciência da Informação (PPGCInf); Faculdade de Ciência da Informação da Universidade de Brasília (UnB/FCI). CV Lattes: http://lattes.cnpq.br/6566830393005282. E-mail: helia.chaves@gmail.com.
[3] Doutora em Ciência da Informação. Professora Adjunta. Programa de Pós-Graduação em Ciência da Informação (PPGCInf) da Faculdade de Ciência da Informação da Universidade de Brasília (UnB/FCI). CV Lattes: http://lattes.cnpq.br/8951909489273046. E-mail: marisab@unb.br.

Resumo do Capítulo 12

A Descoberta de Conhecimento em Texto (DCT) ainda é um processo muito pouco utilizado no Brasil para extração de informações ocultas de documentos textuais, as quais não podem ser visualizadas a partir de recursos tradicionais de recuperação da informação. Conseqüência disso é a escassa literatura em língua portuguesa sobre o tema. Este capítulo compila conceitos de diferentes métodos de análise de informação presentes na literatura estrangeira. Considera questões terminológicas que envolvem a adoção de nomenclaturas em língua portuguesa, assim como as relações entre a diversidade de termos e conceitos utilizados para definir a prática da DCT. Aborda, ainda, o desenvolvimento do tema no campo da Ciência da Informação e apresenta algumas experiências de pesquisa, com vistas a despertar o interesse do público da área nessa nova abordagem no trato da informação textual não estruturada.

Palavras-chave: Descoberta de conhecimento em texto; DCT; Mmineração de textos; Mineração de dados; Descoberta de conhecimento em dados.

Abstract

Knowledge discovery in text: concepts and convergences

Knowledge Discovery in Text (KDT) is a process still rarely used in Brazil to extract hidden knowledge – that cannot be found by means of the traditional information retrieval systems – from textual documents. As a consequence, literature in Portuguese language about this subject is scarce. This chapter compiles from literature in other languages some concepts on different information analysis methods. It considers terminological issues involving the adoption of nomenclatures in Portuguese as well as the relations among the diversity of terms and concepts used to define DCT practice. It also addresses the development of this theme in the scope of Information Science and presents some research results in view of catching the interest of information community towards this new approach to deal with unstructured textual information.

Keywords: *Knowledge discovery in text; KDT; Text Mining; Data mining, Knowledge discovery in data*

Como citar o Capítulo 12

SCHIESSL, Marcelo; CHAVES, Hélia; BRÄSCHER, Marisa. Descoberta de conhecimento em textos: Conceitos e convergências. In: Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios pelo Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento.* Brasília DF: IBICT, 2010. 335 p. ISBN: 978-85-7013-072-3. Capíitulo 12, p. 278-293. Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

1. INTRODUÇÃO DO CAPÍTULO 12

A proliferação de documentos textuais nas organizações e a percepção de que trazem em sua essência o conhecimento levaram à compreensão de que esse tipo de documento merece ser explorado, seja para apoio à tomada de decisão, seja para a geração de novos conhecimentos. Assim, ferramentas e técnicas, antes criadas para estudar dados estruturados provenientes de bases de dados, foram aperfeiçoadas de forma a criar condições para a exploração de textos escritos em linguagem natural. Da mesma forma, tratamentos estatísticos foram aprimorados para expandir a capacidade humana de leitura e registro dessa crescente massa de conteúdos textuais, assim como para revelar novos conhecimentos a partir das relações entre esses conteúdos. Assim é a Descoberta de Conhecimento em Textos (DCT), uma garimpagem de tesouros ocultos nos textos escritos em linguagem natural.

Diante da escassez de literatura em língua portuguesa que trata da aplicação da DCT em busca de conhecimento oculto e da constatação de que essa ainda é uma prática muito pouco explorada no Brasil, decidimos compilar alguns conceitos encontrados na literatura e estabelecer as relações entre a diversidade de termos utilizados na língua portuguesa para referir a essa técnica. Igualmente, são apontadas as convergências de idéias e a falta de consenso no tocante à nomenclatura entre os estudiosos do assunto.

Muito mais do que uma orientação teórica sobre novos conceitos e técnicas no campo da Ciência da Informação, este capítulo pretende instigar a curiosidade do público da área no que diz respeito a uma nova abordagem para a lida com a informação, onde a tônica é posta na proximidade com a expressão humana na sua forma mais natural: o texto.

2. ACERCA DA DENOMINAÇÃO

Há na literatura uma diversidade de termos que denominam as técnicas utilizadas para extração de informações relevantes de grandes volumes de textos armazenados eletronicamente. Trazemos à tona essa discussão considerando aspectos como a origem dos termos associada à finalidade das técnicas que descrevem. Entre convergências e divergências de conceitos, buscamos elucidar as dúvidas do leitor que procura informações sobre essas técnicas ainda pouco documentadas em língua portuguesa.

Quando trazemos para a nossa realidade aplicações técnicas já praticadas em outras culturas, ficamos sujeitos à adoção da terminologia consolidada pela literatura em outros idiomas. Ao optarem pela proposição de termos em língua portuguesa, autores adotam diferentes neologismos. Ficamos, dessa forma, expostos a uma saraivada de termos que está longe de se estabilizar.

Exemplo desse problema terminológico pode ser ilustrado numa busca na Internet pelo termo “Mineração de Textos”, onde recuperaremos diversos documentos com uma miscelânea de nomes. Provavelmente, virão nomes como Mineração de Dados Textuais, Garimpagem de Textos, Descoberta de Conhecimento em Texto, Destilação de Textos e outros que serão criados entre o momento em que escrevemos até a busca do leitor. E então, qual é o termo?

A escolha de determinados termos é influenciada por fatores sociais, culturais e econômicos. Um povo estabelece códigos que são utilizados para comunicação, seja pela criação, pela transformação ou pela incorporação de termos estrangeiros. Esses códigos traduzem o significado de objetos, ações ou qualquer coisa que precise ser representada por meio da linguagem. A influência estrangeira na evolução de uma língua sempre esteve presente na história. O Latim foi dominante por séculos. Alguém se lembra da letra da música “Língua”, de Caetano Veloso, que afirma “Está provado que só é possível filosofar em alemão"? Vimos a pujança do francês em várias culturas, inclusive na brasileira. De onde acham que vem o termo “vitrine", “lingerie" ou “garçon"? Recentemente, a dominação do inglês é inegável. Incorporamos várias palavras no nosso dia-a-dia que dão sentido às coisas.

Algumas delas por não estarem disponíveis no nosso léxico. Há, ainda, traduções de qualidade duvidosa que criam aberrações como o atual gerundismo dos nossos *Call Centers*[4].

Em relação ao aspecto terminológico concernente ao tema deste capítulo, discutiremos o significado das denominações mais usuais e apresentaremos a terminologia em português adotada nesta obra. O termo *Text Mining* é o mais freqüente na literatura. Seguido pela sigla KDT, que significa *Knowledge Discovery in Text*. Há outras variações do termo, mas não serão abordadas aqui.

2.1Text Mining

A expressão "mining", de acordo com o dicionário em linha The Free Dictionary (TFD)[5], significa:

1. Ação, processo ou atividade de extração de minério do solo;
2. Colocação de minas (engenho de guerra).

À primeira vista, podemos concluir que a segunda definição não possui relação com o nosso tema, enquanto que a primeira remete à extração de algo valioso oculto no solo. Logo, a expressão *text mining* pode ser entendida como uma metáfora que traduz a atividade de extração de valores ocultos em textos.

Em nosso idioma, o termo "mineração", conforme definido por Houaiss (2000), significa:

1. Ação ou efeito de minerar; trabalho de extração do minério;
2. Depuração do minério extraído das minas.

Essas definições mantêm o mesmo sentido da expressão inglesa, que é extrair valor de coisa que se subentende oculta no solo ou no mar. Assim, a expressão *text mining* é traduzido literalmente para "mineração de textos". Entendemos que a variante "mineração de dados textuais" encontrada na literatura não adiciona significado especial à expressão. O fato de o texto ser considerado "dado textual" não altera o significado da expressão que também considera a extração de valores da fonte textual.

Adicionalmente, outra expressão encontrada na literatura, "garimpagem de texto", nos remete à definição do vocábulo "garimpagem" encontrada em Houaiss (2000):

1. Atividade artesanal, desenvolvida pelo garimpeiro em aluviões, eluviões, leitos e margens de rios e grupiaras, e que consiste na exploração de minerais preciosos, esp. o diamante, com instrumentos rudimentares; faiscação, cata;
2. Pesquisa minuciosa de palavras, expressões, textos etc.

A primeira definição possui significado similar ao da mineração e a segunda está diretamente relacionada à matéria-prima de que trata este trabalho, o texto. Por esse prisma, a expressão "garimpagem de textos" seria mais adequada pela conexão direta com a atividade de pesquisar textos, expressões, palavras.

A dúvida não acaba aqui, continuemos com o próximo termo para verificar se acrescenta novo significado.

2.2 Knowledge Discovery in Text

A expressão *Knowledge Discovery in Text (KDT)* traduzida para o português significa Descoberta de Conhecimento em Texto (DCT). Em princípio, uma adequada denominação para a atividade de se adquirir coisas de valor embutidas em documentos textuais, como o conhecimento. Se verificarmos com um pouco mais de cuidado a palavra *Knowledge* veremos que sua definição abre um leque de possibilidades.

---

[4] Centros de Atendimento, para os puristas.

[5] The Free Dictionary: http://www.thefreedictionary.com

O termo é fonte de calorosas discussões no meio acadêmico. Na filosofia, não há acordo sobre sua precisa definição. A aquisição do conhecimento se dá pela experiência. Isso envolve a cognição humana, que compreende ou conhece algo através do raciocínio ou da razão. Se o conhecimento só é possível quando utilizamos o nosso aparelho cognitivo, então não há que se falar em descoberta de conhecimento automática, pois a máquina não dispõe de tal mecanismo. Nas áreas relacionadas à gestão, entende-se conhecimento como meio de compreensão segura de assunto com a possibilidade de usá-lo em propósito específico. Essa abordagem nos dá espaço para utilização do termo, pois a compreensão de fenômeno para utilização em atividade específica é exatamente o que procuramos. Continuemos analisando o termo *knowledge*. A definição no dicionário TFD mostra:

1. Estado ou fato de conhecer;
2. Familiaridade, consciência ou entendimento obtido por meio da experiência ou estudo;
3. Soma ou extensão do que foi percebido, descoberto ou aprendido;
4. Saber; erudição: mestres de grande saber;
5. Informação específica sobre algo;
6. Conhecimento carnal.

Vimos que as definições oscilam do ponto de vista filosófico ao mercadológico. Segundo Frawley *et al*. (1992), que cunharam o termo *Knowledge Discovery in Data* (KDD)[6], o termo "conhecimento" não deve ser interpretado nos moldes da filosofia[7], mas como padrão oculto nos dados. Analogamente, em KDT, padrão oculto no texto. Nessa visão, o termo *Text Mining* é uma etapa de todo processo de descoberta.

Na tradução do termo para o português, herdam-se os mesmos problemas de definição da língua inglesa. Como exemplo, o termo "conhecimento", segundo Houaiss (2000), possui várias definições, que variam nas diversas áreas do saber. Portanto, não há consenso. Para fins práticos, neste trabalho, estabelece-se que a DCT é o processo que envolve vários passos, os quais estão descritos no item 1.7. Eles vão da coleta de documentos à aplicação do conhecimento descoberto.

Então, qual é o termo? O que queremos é descobrir novos relacionamentos e significados em determinada coleção de documentos. Respondendo à pergunta: poderia ser qualquer um deles. O conceito é o mesmo. Em todo caso, vamos adotar a terminologia que melhor denota o que buscamos. Isto é, extrair valor do texto bruto de forma automatizada. Poeticamente falando: extrair pedras preciosas escondidas no leito dos textos.

Em suma, consideramos que o termo "descoberta de conhecimento em texto" é mais abrangente que "mineração" ou "garimpagem de texto". Entendemos que ele traduz o trabalho envolvido da coleta de textos à aplicação de padrões descobertos. Assim, o termo Descoberta de Conhecimento em Texto e o acrônimo DCT serão utilizados neste capítulo daqui por diante.

3. DCD, DCT E RECUPERAÇÃO DA INFORMAÇÃO

Embora encontremos na literatura diferentes abordagens sobre DCD e DCT, nas quais percebemos a distinção entre essas duas técnicas, há alguns autores que consideram a DCT como uma variação da DCD, enquanto outros defendem a idéia de que se trata de uma "evolução natural" da mineração.

Ambas as técnicas almejam encontrar, de forma automática, padrões valiosos em grandes bases de dados. A principal distinção entre elas reside no fato de que, na DCD, esses padrões são extraídos em formato numérico, enquanto que na DCT ocorre a análise de coleções de documentos nas quais os conteúdos são textuais. Não há dúvidas sobre essa distinção de formas, número *versus* texto. Apesar disso, os conceitos não são distintos, pois para viabilizar o processo de DCT, o texto será processado e transformado em representação numérica. (WEISS *et al.*, 2005).

[6] Descoberta de Conhecimento em Dados (DCD)
[7] Para o leitor interessado na visão filosófica, sugere-se a leitura de Hessen, J. Teoria do conhecimento: tradução João Vergílio Gallerani Cuter, 2ª ed., São Paulo, Martins Fontes, 2003.

Tan (1999) defende a idéia de que a DCT, refere-se geralmente ao processo de extração de padrões interessantes e não triviais de documentos textuais. A técnica pode ser vista também como uma extensão da DCD. Destaca-se, ainda, a complexidade da DCT em relação à DCD, dada a sua característica inerentemente não estruturada e difusa.

A DCD é vista por Trybula (1999) como "processo básico empregado para analisar padrões em dados e extrair informações", e a DCT como "processo básico para analisar padrões em textos e apresentar informações".

Hearst (1999) chama atenção para o uso errôneo do termo *data mining* e alega que há poucas pesquisas sobre *text mining*. Considera a DCT como uma ferramenta de suporte e valorização do conhecimento gerado, já que proporciona sua exploração e reutilização, e a define como:

> Descoberta, por computador, de novas informações, previamente desconhecidas, pela extração automática de informações de diferentes recursos escritos, nos quais o elemento chave é a interligação das informações extraídas para formar novos fatos e novas hipóteses a serem posteriormente exploradas pelos meios de experimentação mais convencionais. (HEARST, 2003)

Há autores que abordam a questão da diferenciação entre a DCT e a recuperação da informação. Hearst (2003) afirma que DCT, ao contrário do que muitos pensam, não é uma forma de facilitar a busca de informação na *Web*. Portanto, não deve ser confundida com o processo de Recuperação da Informação (RI). Segundo a autora, na RI ocorre descoberta de informação já conhecida, o usuário seleciona, em uma coleção de documentos, aqueles que lhe interessam e despreza os demais. Já na DCT, a meta é descobrir informação desconhecida, que não está escrita em nenhum documento individualmente; derivar novas informações a partir dos dados analisados; encontrar padrões em conjuntos de dados, ou, ainda, separar o signo do ruído.

Embora a conceituação das técnicas de tratamento automático de grandes volumes de dados ainda se encontre de certa forma não pacificada, identificamos pontos convergentes fundamentais para a compreensão do seu funcionamento. Assim, fica evidenciada a grande evolução dessas técnicas com a possibilidade de tratamento de conteúdos em linguagem natural, os quais representam, de modo geral, a maioria dos conteúdos gerados por uma organização.

4. A DESCOBERTA DE CONHECIMENTO EM DADOS (DCD)

O admirável mundo novo de documentos digitais e redes de computadores globais abriu os horizontes, mas também mostrou grandes desafios. O volume de informação aumenta além da capacidade humana de apreensão. Então, o lapso crescente entre a criação de dados e a sua compreensão mostra a face da sobrecarga de informação.

A DCD[8] se apresenta como uma importante opção para atender a essa necessidade. Os grandes volumes de dados já não comportam métodos manuais. Eles são custosos, lentos e subjetivos. A meta, então, seria desenvolver métodos e técnicas que dão sentido aos dados. O processo básico é traduzir a forma elementar, o dado, em outra mais compacta e útil, a informação.

1. Frawley *et al.* (1992) afirmam que a Descoberta de Conhecimento é a extração não-trivial da informação implícita, nos dados, previamente desconhecida e potencialmente útil. Fayyad *et al* (1997) complementam que a informação deve ser compreensível.

Não se deixe enganar com a objetividade dessas afirmações. Analisemos alguns dos termos para que se revele a complexidade. Entendemos que dado é um conjunto de fatos. Padrão é a estrutura implícita que será encontrada. Processo envolve a preparação dos dados, a busca por padrões, a avaliação do conhecimento descoberto e os refinamentos necessários em repetidas iterações. Pelo termo não-trivial depreendemos que a busca ou inferência não seja uma operação direta. Se assim o fosse, não se justificaria o trabalho. A validade dos novos padrões descobertos é óbvia, caso contrário, não se

[8] Do inglês *Knowledge Discovery in Database* (KDD). Existe uma discussão entre diversos autores da área a respeito da abrangência do termo KDD e *Data Mining* (DM), porém os termos são referidos em vários trabalhos, indistintamente.

atribuiria confiabilidade. A novidade é inerente à descoberta. A utilidade agrega benefício ao usuário. Por fim, que seja compreensível, mesmo que necessite de pós-processamento.

2. Segundo Berry e Linoff (1997), a DCD é a análise e exploração automáticas ou semi-automáticas de grandes quantidades de dados com o objetivo de descobrir regras e padrões significativos.

As definições 1 e 2 acima exprimem visões com nuances sobre o mesmo tema. A primeira trata do processo de descoberta como um todo. Isto é, desde a aquisição do dado, o armazenamento, a mineração, que retira a informação codificada, até a sua apresentação ao usuário final. A segunda privilegia os aspectos computacionais, os algoritmos dedicados à DCD e à exeqüibilidade da máquina em tarefas de manipulação de grandes volumes de dados e a transformação deles em informação útil ao homem.

Reforçando, o objetivo da DCD é encontrar padrões interessantes ocultos em grandes quantidades de dados e fornecer informações como insumo para aquisição do conhecimento. O fundamento está na oferta de novas capacidades que se traduzem em habilidade para otimizar a tomada de decisão utilizando métodos automáticos para aprender com ações passadas.

O processo da DCD é uma atividade multidisciplinar que se apropria de técnicas de várias áreas, tais como Estatística, Aprendizado de Máquinas, Visualização de Dados, Banco de Dados, entre outras. A DCD, portanto, não está sob a tutela exclusiva de uma área, mas da conjunção de várias.

O desenvolvimento da DCD está intrinsecamente relacionado à evolução da tecnologia. A DCD vem se consolidando como um poderoso ferramental para auxiliar o homem na exploração da grande quantidade de informação disponível em formato eletrônico, dadas as limitações humanas no manuseio e interpretação dessa informação.

4.1 O processo da DCD

Então, existirá uma forma ideal de se fazer a DCD? Infelizmente, não, mas existe luz no fim do túnel! Há realmente diferentes pontos de vista expressos em vários trabalhos que enfocam a DCD e fica claro que ainda não existe consenso quanto a sua definição, nem quanto na sua constituição. Vários autores sugerem processos para a implementação do ciclo da DCD e, nesse sentido, não há a melhor abordagem. Nossa contribuição aqui é a de sintetizar as visões de autores reconhecidos como Fayyad *et al.* (1997), Berry e Linoff (1997), Trybula (1997), Han e Kamber (2000) e Kantardzic (2003).

Identificamos variações na ordem em que esses autores apresentam as etapas da DCD, mas foi possível sintetizá-las nos seguintes passos (contudo podem ser resumidas nos passos seguintes):

- **Compreensão e definição do problema**: estabelecimento do objetivo e levantamento de obstáculos para o desenvolvimento da atividade desejada, além das áreas ou pessoas que poderão ser beneficiadas com o valor agregado da informação extraída dos dados.
- **Seleção de fontes de dados**: adaptação da base de dados por meio da coleta, combinação e integração de fontes, com a finalidade de atender aos objetivos da DCD.
- **Processo de limpeza e adequação dos dados**: remoção de erros ou inadequações de campos e transformação de dados, com o objetivo de ajustá-los à DCD. Esse passo é freqüentemente citado como "pré-processamento".
- **Análise exploratória**: exame dos dados em busca de estruturas que expressem alguma relação entre variáveis ou registros.
- **Redução de variáveis**: redução da dimensionalidade em função da quantidade de variáveis. Em muitos casos, o número de variáveis é tão grande que inviabiliza a análise.

- **Relacionamento de objetivos**: escolha do algoritmo para DCD apropriado à intenção do usuário ou do analista.
- **DCD**: uso específico da ferramenta de mineração que tem o propósito de extrair padrões dos dados.
- **Interpretação de resultados**: análise dos padrões para verificação de sua utilidade e realimentação de informação.
- **Transformação do conhecimento adquirido em ação**: utilização do conhecimento extraído ou sua incorporação à base de conhecimento acumulado.

A *Figura 1* **Erro! Fonte de referência não encontrada.**apresenta um esquema gráfico ilustrando essas etapas. Na fase de avaliação ou interpretação, o processo pode ser retomado em qualquer um dos passos anteriores, de acordo com o julgamento do especialista.

Figura 1 – Passo a passo que compõe o processo de DCD
Fonte: Fayyad et al, 1997, pág. 41

A DCD é uma combinação interdisciplinar de técnicas e métodos concebida para resolver problemas, na sua grande maioria, em dados numéricos e estruturados, isto é, armazenados em bancos de dados. Contudo, apenas uma pequena parte dos dados está adequada ao tratamento proposto pela DCD e, nesse sentido, os dados expressos em linguagem natural não são contemplados. Portanto, existe uma necessidade de adaptação da DCD para que a linguagem natural seja passível de processamento automático visando à extração de conhecimento.

5. A DESCOBERTA DE CONHECIMENTO EM TEXTOS (DCT)

O que não falta é discussão em torno de definições das técnicas de extração automática de conhecimento em grandes volumes de informação. Conceitos e termos se misturam, por vezes sendo utilizados indistintamente, como se sinônimos fossem. Deste modo, utiliza-se o termo "descoberta de conhecimento", passando-se a qualificá-lo a partir do conteúdo a ser analisado: se este foi previamente organizado e estruturado (descoberta de conhecimento em dados) ou se consta em documentos textuais, qualquer que seja o formato ou tamanho (descoberta de conhecimento em textos).

Essa observação encontra sustentação nos estudos de Wives (2004), para quem a descoberta de conhecimento é um "processo de aquisição de novos conhecimentos auxiliado por computador", que pode ser classificado em Descoberta de Conhecimento em Dados Estruturados (DCD, ou KDD em inglês) e Descoberta de Conhecimento em Textos (DCT, ou KDT em inglês).

Tan (1999) e Feldman *et al* (2001) declaram que 80% da informação de empresas está armazenada em documentos textuais e, de acordo com Dörre *et al.* (1999), a informação textual não está prontamente acessível para ser usada por computadores. Ela é apropriada para que pessoas, através da leitura e dos processos cognitivos característicos dos humanos, a manipulem e apreendam as informações nela contidas.

É relativamente simples chegarmos a essa conclusão. O texto é abundante nas organizações. Ele pode ter formato de memorandos, ofícios, normas, e-mail, nas páginas da intranet e Internet, entre outros tipos. Vivemos a era da instantaneidade da informação digital. Textos eletrônicos aparecem a todo instante nas telas de nossos computadores e nos visores de nossos celulares.

Agora, tente fazer com que o computador faça inferências em conteúdos textuais. A não ser que você informe para ele, por exemplo, que casa, lar e morada tratam do mesmo tema, esses termos não serão tratados como semelhantes. Entretanto, qualquer pessoa de instrução mediana conseguirá fazer tais associações sem grandes esforços.

Parece razoável supor que a DCT assemelha-se à DCD, desde que os dados sejam estruturados ou adaptados para a máquina. Que tipo de adaptação seria essa? À primeira vista, transformamos os textos em planilhas. No entanto, ao olharmos com um pouco mais de profundidade, estamos transformando textos em números. Mais especificamente, em 0s e 1s! Essa é a língua que computadores entendem.

Importante clarificar que vários autores afirmam que as bases textuais se apresentam de forma não-estruturada. Porém, possuem uma estrutura implícita que necessita de técnicas especializadas para ser reconhecida por sistemas automatizados. O processamento de linguagem natural (PLN) trata exatamente da descoberta destas estruturas implícitas, como por exemplo, a estrutura sintática (RAJMAN; BESANÇON, 1997).

Como mencionado, o texto, apesar de conter informações valiosas, está codificado de maneira que é difícil de ser decifrado automaticamente. Assim, a ciência vem buscando soluções para simular a cognição humana, que é capaz de processar o texto e de apreendê-lo de maneira satisfatória.

Neste sentido, temos o ferramental da Lingüística, que busca caracterizar e explicar a diversidade de fenômenos lingüísticos que nos cercam, seja em diálogos, seja na escrita, seja em qualquer outro meio. Uma parte se preocupa com o lado cognitivo de como o homem adquire, produz e entende a linguagem; outra parte, a compreensão da relação entre o discurso lingüístico e o mundo; e a terceira, com a compreensão das estruturas lingüísticas por meio das quais o homem se comunica.

Paralelamente, o desenvolvimento da informática tem possibilitado grandes avanços no estudo das línguas naturais. A área que examina as relações entre a Lingüística e a Informática é a Lingüística Computacional, que orienta a construção de sistemas especialistas em reconhecer e produzir informação em linguagem natural. Nesse contexto, o PLN estuda a interpretação e geração de informação nos diferentes aspectos da língua: sons, palavras, sentenças e discurso nos níveis estruturais, de significado e de uso.

Não faz parte do escopo deste capítulo o aprofundamento na discussão do PLN, seus métodos e suas técnicas, mas somente a contextualização da relação entre a DCD e a DCT. Durante várias décadas, inúmeras pesquisas têm provocado avanços no PNL e, atualmente, encontramos procedimentos disponíveis capazes de realizar o tratamento do dado textual de maneira a possibilitar sua transformação e sua estruturação na forma adequada ao uso pela DCD.

A integração de técnicas de PLN e DCD constitui a DCT, que objetiva automatizar o processo de transformação de dados textuais em informação para possibilitar a aquisição do conhecimento.

A *Figura 2* parece bem familiar, não? Ela apresenta o ciclo do processo da DCT e equivale a uma adaptação ao modelo proposto por Fayyad *et al.* (1997). Nela, observamos que o processo de DCT abrange a seleção do *corpus*, o pré-processamento, que envolve sua adequação aos algoritmos, a efetiva DCD textuais, a validação dos resultados e, finalmente, a análise e interpretação dos resultados para a aquisição do conhecimento. Comparando esse esquema ao apresentado na **Erro! Fonte de referência não encontrada.**, verificamos que as etapas da DCT seriam idênticas às da DCD, exceto pela utilização das técnicas de PLN para transformar o texto em formato adequado à mineração dos dados.

Figura 2 – Processo de DCT. Adaptação de Fayyad et al. (1997), pág. 41.

Vale destacar que, embora a aderência entre o PLN e DCT seja muito forte, seus objetivos não se confundem, pois o PLN trata dos aspectos lingüísticos relacionados a um texto específico, enquanto que a DCT busca as relações contidas entre os textos de uma coleção com o objetivo de apresentar a informação relacionada a um grupo ou grupos de textos. Ainda, o PLN analisa o conteúdo dos textos, enquanto que a DCT se utiliza dessas análises, em uma fase de pré-processamento, para transformá-los em dados apropriados à descoberta de padrões e conhecimento entre os textos da coleção (KODRATOFF, 1999).

A DCT possibilita, então, o reconhecimento e a produção da informação apresentada em linguagem natural e, nesse sentido, vem contribuir enormemente com a Ciência da Informação no que tange ao tratamento e recuperação da informação.

6. CONSIDERAÇÕES SOBRE A DCT E A CIÊNCIA DA INFORMAÇÃO

A DCT é a conjunção de várias metodologias e conceitos. Evidentemente, este trabalho apresenta uma reprodução estática de seu desenvolvimento e implementações até o momento. O desenvolvimento e o aperfeiçoamento do processo são constantes, dada a natureza da língua e das ferramentas tecnológicas.

Konchady (2006) declara que a DCT é uma prática relativamente nova derivada da Recuperação da Informação (RI) e da PLN e essa afirmação estabelece um vínculo significativo, visto que a RI é uma das áreas principais de pesquisa da CI.

Segundo Bräscher (1999), os avanços tecnológicos influenciam a CI e favorecem o surgimento de novas técnicas de representação e recuperação de assunto, considerando os aspectos cognitivos envolvidos no processo de comunicação homem-máquina, os quais exigem modelos de representação do conhecimento capazes de contextualizar os significados expressos nos textos armazenados.

Lima (2003), em uma releitura de Saracevic (1995), expõe que a CI é uma área interdisciplinar que reúne a Biblioteconomia, a Ciência Cognitiva, a Ciência da Computação (CC) e a Comunicação, com forte associação dos processos humanos da comunicação e da tecnologia no seu contexto contemporâneo. De fato, a CC trata de algoritmos relacionados à informação, enquanto a CI se dedica à compreensão da natureza da informação e de seu uso pelos humanos. A CI e a CC são áreas complementares que conduzem a aplicações diversas (SARACEVIC, 1995).

Robredo (2003) reafirma a interdisciplinaridade da CI e orienta que não se pode restringir o escopo e a abrangência da informação ao campo exclu-

sivo da Biblioteconomia e da CI, pois variados estudiosos, pesquisadores e especialistas lidam com a informação de um ponto de vista cientifico e nas mais variadas abordagens e aplicações. Ainda, ensina que ela pode ser dividida, para fins de estudo e delimitação do(s) objeto(s), mas sem perder de vista o interesse comum de todos os seus domínios, a entidade "informação".

Ainda, Lima (2003) aponta as possibilidades de interseção entre a CI e a CC que se concentram nos processos de categorização, indexação, recuperação da informação e interação homem-computador.

Nesse sentido, a DCT pode ser vista como a interposição da Estatística, que utiliza métodos quantitativos para transformar dados em informação, da CC, que fornece suporte tecnológico para manipulação da informação, e da CI, que concentra o foco de atuação na gestão da informação.

Essas áreas são mutuamente beneficiadas pelo aporte teórico de cada uma, o qual favorece o desenvolvimento conceitual interdisciplinar. De tal modo, fica caracterizada a evidente contribuição do estudo da DCT no âmbito da CI, contrariando a tradicional visão da DCT como campo de estudo tradicionalmente explorado pela Lingüística, Computação, Estatística e Inteligência Artificial.

7. APLICAÇÕES DA DCT

A aplicação da DCT se dá por meio da utilização de ferramentas tecnológicas desenvolvidas especialmente para revelar padrões e relações ocultas em dados com vistas a propiciar o entendimento de tendências históricas e prever oportunidades futuras. Essas ferramentas incorporam tecnologias de integração de dados voltadas à limpeza e organização de dados brutos e sistemas de inteligência, que convertem em conhecimento as informações ocultas em textos não estruturados, para apoio à tomada de decisão.

Conforme amplamente discutido na literatura, para que seja possível realizar qualquer tratamento automático de uma coleção de documentos escritos em linguagem natural em busca do conhecimento nela embutido, torna-se necessária a limpeza e padronização do texto. Sob a ótica computacional, desconstroem-se textos seqüenciais – esses que conhecemos quando lemos um livro – e os transformam em números – esses que máquinas lêem milhões por segundo. Pragmaticamente, adequamos o texto e o traduzimos para a linguagem das máquinas, a numérica. Aquela que é calculável.

Há uma diversidade de técnicas que propiciam a aplicação da DCT, sendo as mais conhecidas: análise de conglomerados ou agrupamentos (*clustering*), classificação, extração de informações, sumarização, análise qualitativa e quantitativa e identificação de regras de associação. Todas elas são interativas, pois requerem o bom senso humano para intervir e interagir no processo e nas análises. São também iterativas, pois podem existir diversas análises e recomeços até que se descubra algo desejável.

As armadilhas da linguagem estão em todo lugar. A experiência demonstra que se gasta muito tempo na remoção de ruídos com o intuito de padronizar os dados, de forma a possibilitar maior precisão no processo de mineração. Para a máquina, há diferença entre os termos "casa", "casas", "casar". Em um processo de contagem, haveria três termos distintos. Para uma pessoa, nem tanto assim! Dependendo do contexto, eles poderiam ser convertidos para um só, "casar". Todas as conjugações desse verbo passariam para a sua forma infinitiva. O termo seria contado três vezes. Em outra situação, dois termos. O verbo "casar" contado uma vez e o substantivo "casa", duas vezes.

A eliminação de ruídos pode facilmente consumir dezenas de horas no pré-processamento. Há variantes de toda ordem. É preciso tratar sinonímia, homonímia, identificar nomes próprios, instituições e características que o analista julgar importantes. Por exemplo, num *corpus* relacionado ao tema Informática, o nome Bill Gates pode ter relevância maior do que Gilberto Freyre. Já em um tema relacionado à Sociologia, é provável que não. As escolhas são determinantes para o resultado. Parece trivial estabelecer relações orientadas pelo conteúdo semântico quando lemos um texto. Entretanto, torna-se penoso para construção das bases de conhecimento nas quais o cérebro eletrônico extrai inferências. Lembre-se, a máquina só executa o que você pede, responde somente ao que você pergunta. Não há exceções. Pelo menos ainda.

Para ilustrar essa idéia, segue a descrição de um exemplo prático da aplicação da DCT por meio da técnica de análise de agrupamentos em uma base textual.

8. ANÁLISE DE AGRUPAMENTOS (*CLUSTERING*) – UMA DESCOBERTA INUSITADA

A experiência foi realizada com documentos extraídos de uma base de conteúdos textuais, o Serviço Brasileiro de Respostas Técnicas (SBRT)[9]. Foram minerados 2.401 documentos, nos quais constam as seguintes informações: título do documento (RT), resumo, data de publicação, palavras-chave, assunto, título da demanda ou a pergunta feita pelo cliente e instituição respondente ou responsável pela elaboração da RT.

Após a extração dos textos originais da base, o conjunto de documentos foi submetido a vários processamentos para limpeza e padronização dos dados, até que pudesse ser utilizado para a mineração do texto propriamente dita, utilizando-se ferramenta específica.

Quadro 1 – Agrupamentos dos Documentos Textuais. Fonte: RAMOS, 2008 (pág. 73)

Agrupa-mento	Peso (Presença na Base)	Freqüência (Quantidade Documentos)	Variabilidade dos termos no agrupamento	Termos de Agrupamento
1	5%	120	0,1032064562	+ óleo, + químico, + usar, + processar, + utilizar
2	19%	464	0,1139042878	+ produto, + alimentar, + processar, + dever, + comer
3	5%	122	0,1075986813	+ químico, + produto, + produzir, + técnico, + resumir
4	20%	485	0,1144369151	+ material, + utilizar, + fonte, + usar, + processar
5	5%	115	0,1017080841	+ animal, + alimentar, + agricultura, + dever, + produção
6	7%	172	0,0941117077	+ cultivar, + solar, + plantar, + dever, + apresentar
7	8%	193	0,1034777388	+ máquina, + fornecedor, + utilizar, + elaborar, + usar
8	6%	136	0,099722373	+ identificação, responsável, + necessário, + dever, + informação
9	2%	37	0,093348348	+ espécie, + criação, + animal, + alimentar, + formar
10	3%	84	0,0876160473	+ material, + reciclar, + resíduo, + plástico, + processar
11	10%	231	0,0976318382	fax, + fornecedor, + equipamento, + máquina, + indústria
12	10%	242	0,1193337586	+ químico, + utilizar, + usar, + resumir, + processar
Total de documentos:		2.401		

[9] SBRT - Serviço de informação gratuito na *Web* destinado a micro e pequenas empresas, criado pelo Ministério da Ciência e Tecnologia (MCT), em parceria com universidades, iniciativa privada e governo: CDT/UnB, Disque-Tecnologia da USP (CECAE/USP), CETEC/MG, REDETEC/RJ, TECPAR/PR, IEL/BA, SENAI/RS, IBICT e Sebrae Nacional. Alimenta uma base de dados contendo soluções tecnológicas aos problemas postados pelos clientes, elaboradas por especialistas, sob medida, e disponibilizadas em forma de Respostas Técnicas (RTs).

Nesse processo, o algoritmo especializado, baseado na freqüência e no peso associado aos termos, elegeu os cinco mais representativos de cada documento, formando os 12 agrupamentos apresentados no *Quadro 1*.

É intrigante a presença do termo "químico" em três agrupamentos. À primeira vista isso pode ser observado com certa desconfiança. Ora, se na mineração de textos os documentos são agrupados tendo como elos os cinco termos representativos que apontam algum tipo de afinidade entre eles, o que estaria provocando a "separação" de três grupos de documentos (1, 3 e 12), os quais com certeza tratam de assuntos relacionados a produtos químicos? Não deveriam esses documentos estar dispostos em um único agrupamento?

Com esses questionamentos em mente, podemos olhar "mais de perto" cada um desses agrupamentos, em busca de sinais que comprovem a pertinência de esses documentos estarem separados.

A intenção é investigar as peculiaridades de cada agrupamento que teriam provocado o "distanciamento" entre os documentos, a despeito da similaridade entre seus termos.

Observamos, à primeira vista, uma forte intersecção entre os agrupamentos 1 e 12, a julgar pela presença de 4 termos comuns dos 5 que os caracterizam (químico, usar, processar, utilizar). Isso provoca a necessidade de um aprofundamento na análise, adicionando-se a esta o agrupamento 3, que apresenta a presença marcante do termo comum a esses dois: "químico".

A *Figura 3* busca representar, em forma de esquema, as interações existentes entre os três agrupamentos, a partir dos termos comuns. Para maior facilidade de visualização, adotamos a seguinte convenção:

i) Os termos que trazem a mesma cor do nome do agrupamento aos quais estão conectados são aqueles que ocorrem exclusivamente naquele agrupamento;
ii) Os termos em preto são aqueles comuns a dois agrupamentos;
iii) Destaque maior é dado ao termo "químico", por ocorrer nos três agrupamentos.

Figura 3 – Os agrupamentos e suas interações. Fonte: RAMOS, 2008 (pág. 77).

Ao analisarmos atentamente os termos dos três agrupamentos em discussão e os seus respectivos conteúdos, observamos que, embora todos estejam relatando ações que empregam processos químicos, há claros indícios de elementos que os diferenciam.

Ao contrário da primeira impressão citada anteriormente, quando inferimos haver forte relação entre os agrupamentos 1 e 12, dada a intersecção de 4 termos em 5, a conclusão que chegamos, após olharmos mais de perto os demais termos extraídos, é exatamente oposta. A semelhança maior se mostra entre os agrupamentos 1 e 3, embora haja apenas um único termo comum entre eles, principalmente do ponto de vista das finalidades do uso de processos químicos.

Ambos os agrupamentos 1 e 3 detêm seu maior conteúdo voltado para a fabricação de produtos cosméticos, material de limpeza e aromatizantes em geral, e aditivos de uso industrial. Entretanto, nota-se a presença marcante, no agrupamento 1, de produtos derivados de óleos, extraídos de produtos naturais, de perfumes a lubrificantes, abarcando também o cultivo de plantas oleaginosas e produção de sementes, óleos e essências de uso culinário. Ainda, observamos um destaque na produção de energia e combustíveis alternativos. Esses são seus principais diferenciais em relação ao agrupamento 3. Já este último apresenta uma diversidade de produtos cosméticos e de higiene pessoal, entretanto, destaca-se na linha de produtos de limpeza e aromatizantes, como desengraxantes, solução desinfetante fungicida, detergente, fumaça artificial, limpa-pedras, entre outros, além de produtos de uso veterinário. Já o agrupamento 12 apresenta aplicações químicas de naturezas diferentes dos outros dois agrupamentos, com conteúdo mais voltado para a indústria de metalurgia, usinagem, couro e calçados, trazendo abordagens específicas, tais como tratamentos de água, pinturas, fundição, edição, gravação, impressão, estamparia, impermeabilização, defensivos agrícolas, imunização e controle de pragas. O *Quadro 2* retrata de forma sucinta o resultado dessa análise.

Quadro 2: Análise comparativa dos agrupamentos 1, 3, 12

Agrupamento 1	Agrupamento 3	Agrupamento 12
Produtos derivados de óleos, extraídos de produtos naturais, de perfumes a lubrificantes.	Produtos cosméticos e de higiene pessoal	Indústria de metalurgia, usinagem, couro e calçados.
Cultivo de plantas oleaginosas e produção de sementes, óleos e essências de uso culinário.	Produtos de limpeza e aromatizantes: solução desinfetante, desengraxantes, fungicida, detergente, fumaça artificial, limpa-pedras.	Pinturas, fundição, edição, gravação, impressão, estamparia, impermeabilização, defensivos agrícolas, imunização e controle de pragas.
Produção de energia e combustíveis alternativos	Produtos de uso veterinário	Tratamentos de água

Dessa forma, fica comprovado que os três agrupamentos tratam de fato de produtos químicos, mas sob aspectos diversos. Essa curiosa descoberta comprova a eficácia da DCT para a geração de agrupamentos de documentos textuais, em busca de conhecimento oculto, o que não é possível detectar em uma busca por palavras-chave. As relações anteriormente desconhecidas permitem o aprofundamento das análises e a conseqüente construção de novos conhecimentos.

9. CONSIDERAÇÕES FINAIS DO CAPÍTULO 12

Neste capítulo, buscamos trazer à tona algumas discussões acerca da Descoberta de Conhecimento em Textos, suas aplicações e contribuições para a Ciência da Informação. Procuramos esclarecer conceitos, a partir da diversidade deles encontrada na literatura, descrever as etapas de sua aplicação e as diversas técnicas utilizadas para a busca de conhecimento oculto em documentos textuais eletrônicos. Da mesma forma, enfatizamos as diferenças entre dois termos amplamente utilizados na literatura e que ainda causam incompreensões: mineração de textos e mineração de dados.

Ressaltamos as diferenças entre a DCT e a recuperação da informação ao mostrar que a DCT toma como base padrões existentes nos textos para criar agrupamentos de documentos que tratam de temas aparentemente díspares e, assim, proporciona um mergulho diferenciado no conteúdo existente. Ou seja, a identificação de documentos semelhantes em uma base de dados vai muito além da utilização da busca e recuperação de termos indexados de forma semelhante.

Há que se destacar a pertinência e importância de se aplicar a DCT em grandes repositórios de informação digital em linguagem natural, para a descoberta de informações até então desconhecidas (ocultas), que podem se transformar em preciosas fontes de novos conhecimentos. No entanto, aqueles que querem se aventurar na aplicação da DCT devem atentar para as dificuldades inerentes aos novos métodos que chegam até nós, tais como: localização de ferramentas de mineração adequadas e pessoal qualificado para manuseá-las; necessidade de várias habilidades profissionais para a realização de todas as etapas da DCT e análises estatísticas (equipe multidisciplinar); e treinamento específico para a otimização do uso dos recursos das ferramentas disponíveis para realizar esse tipo de tratamento textual.

Outro ponto importante é a dependência da língua. Apesar de haver estudos de casos e bibliografia disponíveis principalmente para a língua inglesa, estamos em fase de gestação na língua portuguesa. É necessária a adaptação de recursos para o português brasileiro[10] para que tiremos proveito da DCT. Há carência de bases de conhecimentos, tesauros e dicionários especializados no nosso idioma. Além disso, padecemos com a falta de investimento em programas de computadores preparados para o português.

Esperamos ter motivado os leitores a conhecerem mais sobre a DCT e, quem sabe, até, tê-los incentivado a contribuir para ampliação do uso dessa técnica em diferentes domínios. Certamente novos estudos levarão ao aperfeiçoamento das técnicas e ferramentas disponíveis e conduzirão a novas aventuras no mundo da descoberta de conhecimentos em textos. Finalmente, sabemos que há muito chão para percorrer. Em ciência, nem sempre o caminho é suave, mas a expectativa é sempre de uma grande chegada.

10. REFERÊNCIAS DO CAPÍTULO 12

BERRY, M. J. A.; LINOFF, G. *Data mining techniques - for marketing, sales, and customer support*. New York:John Wiley & Sons, 1997.

BRÄSCHER, M. *Tratamento automático de ambigüidades na recuperação da informação*. Brasília DF: Universidade de Brasília / Programa de Pós-graduação em Ciência da Informação (UnB/PGCInf), 1999. 290 p. Tese de Doutorado em Ciência da Informação.

DÖRRE, J. *et al*. Text Mining: Finding Nuggets in Mountains of Textual Data. *In* : Fifth International Conference on Knowledge Discovery and Data Mining (KDD-99). *Proceedings*, 1999

FAYYAD, U. *et al*. *From Data Mining to Knowledge Discovery in Databases*. AAAI/MIT Press, 1997.

FELDMAN, R. *et al*. A domain independent environment for creating information extraction modules. *In* : Tenth international conference on Information and knowledge management. *Proceedings*, ACM Press, 2001, p.586-588. Disponível em: <http://doi.acm.org/10.1145/502585.502699>. Acesso em 20 ago 2009

[10] Importante citar o português brasileiro explicitamente, devido à diferença entre o nosso idioma e aquele falado na Europa.

FRAWLEY, W. J.; PIATETSKY-SHAPIRO, G.; MATHEUS, C. J. Knowledge discovery in databases: An overview. *AI Magazine*, v. 13, p. 57-70, 1992. Gte Labs Inc, Distributed Cooperating Learning Syst Project, Waltham, Ma, 02254 Disponível em: <citeseer.ist.psu.edu/frawley92knowledge.html>. Acesso em 12 jul 2009.

HAN J.; KAMBER M. *Data Mining: Concepts and Techniques*, Simon Fraser University, Morgan Kaufmann Publishers, 2000.

HEARST, Marti. Untangling Text Data Mining. *In* : The 37th Annual Meeting of the Association for Computational Linguistics - ACL'99, University of Maryland, June 20-26, 1999 (invited paper). *Proceedings*. Disponível em: <http://www.sims.berkeley.edu/~hearst/papers/ acl99/acl99-tdm.html>. Acesso em 20 ago. 2009.

HEARST, Marti. What is Text Mining? SIMS, UC Berkeley. October 17, 2003. Disponível em: <http://www.ischool.berkeley.edu/~hearst/text-mining.html>. Acesso em: 20 ago. 2009.

HOUAISS. *Dicionário Houaiss da Língua Portuguesa*. Universo On Line, v. 2000. Disponível em: <http://houaiss.uol.com.br>. Acesso em: 20 ago. 2007.

KANTARDZIC, M. *Data Mining: Concepts, Models, Methods, and Algorithms*, John Wiley & Sons, 2003, 343 p.

KODRATOFF Y. Knowledge Discovery in Texts: A Definition, and Applications. in Foundation of Intelligent Systems, Ras & Skowron (Eds.) LNAI 1609, Springer 1999.

KONCHADY, M. *Text Mining Aplication Programming*. Boston, Massachussets: Charles River Media, 2006.

LIMA, G. Â. B.; Interfaces entre a ciência da informação e a ciência cognitiva. *Ciência da Informação*. Brasília DF, v. 32, n. 1, p. 77-87, jan./abr. 2003.

RAJMAN, M.; BESANÇON, R. *Text Mining: Natural Language techniques and Text Mining applications*. Chapman & Hall, 1997.

RAMOS, Hélia de Sousa Chaves. *Análise do conteúdo de um sistema de informação destinado à microempresa brasileira por meio de aplicação da Descoberta de Conhecimento em Textos*. Brasília DF: Universidade de Brasília / Programa de Pós-graduação em Ciência da Informação (UnB/PGCInf), 2008. Dissertação de Mestrado em Ciência da Informação. (Orientador: Marisa Bräscher). Disponível em: <http://bdtd.bce.unb.br/tedesimplificado/tde_busca/arquivo.php?codArquivo=3420>. Acesso em 20 ago 2009.

ROBREDO, J. *Da Ciência da informação revisitada aos sistemas humanos de informação*. Brasília DF: Thesaurus; SSRR Informações, 2003, 262 p.

SARACEVIC, T. Interdisciplinary nature of information science. *Ciência da informação*, v. 24, n. 1, 1995.

SCHIESSL, Marcelo. *Descoberta de Conhecimento em Texto aplicada a um sistema de atendimento ao consumidor*. Brasília DF: Universidade de Brasília / Programa de Pós-graduação em Ciência da Informação (UnB/PGCInf), 2007. Dissertação de Mestrado em Ciência da Informação. (Orientador: Marisa Bräscher). Disponível em: <http://bdtd.bce.unb.br/tedesimplificado/tde_busca/arquivo.php?codArquivo=1538> Acesso em 20 ago 2009.

TAN, A.-H. Text mining: The state of the art and the challenges. *In* : The Pacific Asia Conference on Knowledge Discovery and Data Mining – PAKDD'99 Workshop on Knowledge Discovery from Advanced Databases, Beijing, p. 65–70, 1999. *Proceedings*. Disponível em: <http://www.ntu.edu.sg/home/asahtan/Papers/tm_pakdd99.pdf>. Acesso em 19 ago 2009.

TRYBULA, W. J. Text mining. *Annual Review of Information Science and Technology*, v. 34, 1999, p. 385-419.

WEISS, Sholom, INDURKHYA, Nitin, ZHANG, Tong e DAMERAU, Fred. *Text Mining: Predictive Methods for Analyzing Unstructured Information*. New York: Springer, 2005. 237 p.

WIVES, Leandro Krug. *Utilizando conceitos como descritores de textos para o processo de identificação de conglomerados (clustering) de documentos*. Universidade Federal do Rio Grande do Sul. Porto Alegre RS / Programa de Pós-Graduação em Computação, 2004. 126 p. Tese de Doutorado (Orientador: José Palazzo Moreira de Oliveira). Disponível em: <http://www.leandro.wives.nom.br/>. Acesso em 20 ago 2009.

Capítulo 13
A contemporaneidade do tema governo eletrônico e perspectivas de pesquisas

Rinalda Francesca Riecken[1]

SUMÁRIO DO CAPÍTULO 13

Resumo do Capítulo 13, p. 295
Como citar o Capítulo 13, p. 295

1. INTRODUÇÃO DO CAPÍTULO 13, *p. 296*
1.1 Introduz e conceitua o tema "governo eletrônico", *p. 296*
1.2 Benefícios do e-gov, *p. 299*
1.3 Obstaculos do e-gov, *p. 299*
1.4 Literatura, pesquisas e projetos sobre e-governo, *p. 299*
2. A CONTEMPORANEIDADE DO TEMA GOVERNO ELETRÔNICO, *p. 301*
2.1 Estágios do governo eletrônico, *p. 301*
2.2 Governo eletrônico no mundo, *p. 302*
2.3 Governo eletrônico nas grandes cidades do mundo, *p. 305*
2.4 Governo eletrônico no Brasil, *p.306*
2.4.1 Marco inicial do governo eletrônico no Brasil, *p. 306*
2.4.2 A situação do Brasil na classificação (ranking) internacional, *p. 306*
2.4.3 Situação geral do e-gov nos governos e na iniciativa privada, *p. 307*
3. RESULTADOS DE PESQUISA ACADÊMICA EM GOVERNO ELETRÔNICO, *p. 308*
4. O POTENCIAL DE PESQUISAS SOBRE GOVERNO ELETRÔNICO, *p. 310*
5. REFERÊNCIAS DO CAPÍTULO 13, *p. 315*

[1] Rinalda Riecken, Doutora em Ciência da Informação. Programa de Pós-Graduação em Ciência da Informação (PPGCInf), Faculdade de Ciência da Informação Da Universidade de Brasília (UnB/FCI). Pesquisadora e consultora para assuntos de ciência e tecnologia da informação. CV Lattes: http://lattes.cnpq.br/2095646610175565. E-mail: bbzac269@terra.com.br

Resumo do Capítulo 13

O presente capítulo objetiva despertar acadêmicos e pesquisadores da área da Ciência da Informação e suas correlatas para os temas relacionados ao acesso e uso da Tecnologia da Informação (TI) por governos e pela sociedade, bem como apresentar informações introdutórias acerca do potencial de realização de pesquisas sobre o tema governo eletrônico ou e-gov. O capítulo está dividido em quatro tópicos: *1)* introdução e conceituação do tema; *2)* análise da contemporaneidade do tema governo eletrônico no contexto da sociedade da informação; *3)* apresentação de resultados de pesquisa oriundos da tese de doutorado da autora, para fins de exemplificação das preocupações típicas das pesquisas sobre o acesso e uso das tecnologias por governos e pela sociedade; e *4)* rol exemplificativo de temas potenciais de pesquisas sobre e-gov.

Palavras-chave: Governo eletrônico; e-gov; technologia da informação; tópicos de pesquisa, sociedade da infotmação; âmbito governamental

Abstract

The contemporaneousness of electronic government: issues and research prospects
This chapter intends to awake students and researches in the field of Information Science and related areas to the issues concerning the access and use of information technology by governments and the society, as well as to provide introductory information about the potencial of research in the subject electronic government or e-government. The chapter covers four topics: 1) introduction and conceptualization of the topic; 2) analysis of the conteporaneousness of the subject electronic government in the government and information society environment; 3) presentation of research outcomes originating from the author's doctorate thesis, to exemplify typical concerns of research on access an use of technology by government and society; and 4) a series of examples on potential research issues on e-government.

Keywords: Electronic government; e-gov; information technology; research issues; information society; government environment

Como citar o Capítulo 13

RIECKEN, Rinalda. A contemporaneidade do tema governo eletrônico e perspectivas de pesquisas. *In* : Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento – EROIC.* Brasília DF: IBICT, 2010. 335 p. Capíitulo 13, p. 294-319. Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

1. INTRODUÇÃO DO CAPÍTULO 13

O presente capítulo está dividido em quatro seções, a saber:

- a Seção 1.1, contendo os elementos introdutórios com o objetivo de conceituar e introduzir o tema;
- a Seção 1.2, que contextualiza a contemporaneidade do tema governo eletrônico, inserido na temática mais abrangente da sociedade da informação;
- a Seção 1.3, que apresenta um resumo dos resultados de pesquisa acadêmica (RIECKEN, 2008a)[2], com o objetivo de exemplificar para os estudantes e pesquisadores quanto às preocupações típicas de pesquisas sobre o acesso e uso da tecnologia da informação por governos e pela sociedade; e
- a Seção 1.4, que exemplifica uma lista de possíveis pesquisas sobre governo eletrônico.

1.1 Introduz e conceitua o tema "governo eletrônico"

Governo eletrônico é um tema atual, inserido nos governos, em especial por meio de portais e serviços eletrônicos, e na sociedade, como usuária desses serviços e partícipe desse novo processo de comunicação e interação que tanto organiza, quanto esclarece, facilita e agiliza a solução de inúmeros interesses e necessidades. Mesmo que a satisfação com os serviços e informações ainda não ocorra de forma plena, comparativamente ao modelo anterior baseado preponderantemente em atendimento de balcão e por meio de telefones convencionais, há sem dúvida um significativo avanço.

O termo aparece das mais diferenciadas formas na literatura e na *World Wide Web*[3] (*WWW*): e-governo (*e-government, egovernment, eGovernment, Egovernment, E-government, E-Government, EGovernment*), e-gov (e-Gov, egov, EGOV, E-GOV), governo digital (*digital government*), governo on-line (*online government*) e governo transformacional (*transformational government*). Com relação às tecnologias, o e-governo é frequentemente ligado ao entendimento de "governo on-line" ou "governo baseado na *WWW*", mas muitos temas "não-baseados na *WWW*" se inserem na agenda do governo eletrônico. Analogamente, nem todas as aplicações on-line[4] de governo, mesmo baseadas na *WWW*, referem-se a governo eletrônico.

A expressão "e-governo" representa vários tópicos relacionados que se enquadram na noção de atividades governamentais apoiadas ou difundidas por meio de tecnologia eletrônica, alcançando os governos federal, estadual, prefeituras e entidades diversas, sejam elas da administração direta ou indireta, com todas as suas ramificações.

De acordo com Pardo (2000 *apud* HOLZER; KIM, 2008), o governo eletrônico é muito mais que um site na Internet, devendo ser compreendido como um potencial para apoiar a entrega de serviços governamentais por meio do provimento de: acesso ininterrupto às informações e serviços; mecanismos para disponibilizar aos cidadãos os meios de cumprir suas obrigações com as disposições estaduais ou federal, no que se refere a licenças para dirigir, para abertura de negócios, entre outras; uma rede através das várias agências governamentais a fim de disponibilizar a retaguarda colaborativa para servir aos cidadãos; e de vários canais para favorecer e reforçar a democracia digital e a participação cidadã.

---

[2] Pesquisa de doutorado intitulada "Governo eletrônico em administrações locais brasileiras: avaliação de progresso, fatores intervenientes e critérios de priorização de iniciativas", realizada no escopo do programa de Pós-graduação em Ciência da Informação do Departamento de Ciência da Informação e Documentação da Universidade de Brasília (CID/UnB), transformado, em abril de 2010, na Faculdade de Ciência da Informação (FCI).

[3] A Internet, a rede mundial de computadores, é o meio de comunicação e não de publicação de conteúdos, que é a *World Wide Web (WWW).* Entretanto, o termo "Internet" tem se popularizado como entendimento da rede e seus conteúdos.

[4] O estrangeirismo ou aportuguesamento "on-line" vem sendo largamente utilizado na língua portuguesa, sendo preferido ao termo "em linha" ou mesmo ao termo *online* (PRIBERAM INFORMÁTICA, 2010).

A relação de atividades e serviços que poderiam ser estabelecidos seria longa, podendo-se agrupar nas seguintes categorias (ROBREDO, 2005):

- ações políticas e estratégicas referentes à infra-estrutura para consolidar a implementação de e-governo – planejamento de serviços; tecnologia banda larga; acesso por pessoas portadoras de deficiências; integração de serviços; interoperabilidade; e-governo em nível internacional e mundial; federal, estadual e municipal; legislação e regulamentação sobre Internet; privacidade, segurança e e-serviços em geral;
- serviços de informação através de portais, sítios e *links* diversos – estrutura da administração pública e de dirigentes; acesso à legislação; serviços locais; horários de atendimento, requisitos, custos; impostos e tributação; eventos de interesse público; licitações; orçamento; andamento de obras e aplicação de recursos públicos; e
- serviços interativos, inclusive em tempo real – serviços de atendimento ao cidadão (*call centers*); fornecimento de documentos via Internet; andamento de processos; pagamento de impostos, taxas, etc.; inscrições em concursos públicos; educação à distância; acesso a bibliotecas virtuais e centros de referência; participação em concorrências e licitações via Internet.

As interações possíveis, isto é, os atores que potencialmente podem se comunicar na via eletrônica, segundo a classificação adotada por Fang (2002 *apud* Riecken, 2008a), são:

- Governo a Governo (Government-to-Government - G2G);
- Governo a Negócios (Government to Business - G2B) e/ou seu inverso (Business-to-Government - B2G);
- Governo a Cidadãos / Clientes (Government-to-Citizen - G2C) e/ou seu inverso (Citizen-to-Government - C2G);
- Governo a Entidades sem Fins Lucrativos (Government-to-Nonprofit - G2N) e/ou seu inverso (Nonprofit-to-Government - N2G);
- Governo a Servidores Públicos (Government-to-Employee - G2E) e/ou seu inverso (Employee-to-Government - E2G);
- Cidadão a Negócios (Citizen-to-Business - C2B) e/ou seu inverso (Business-to-Business - B2C); e
- Cidadão a Cidadão (Citizen-to-Citizen - C2C).

Embora o e-gov seja frequentemente associado ao uso das denominada "tecnologia da informação" (TI)[5] e da *WWW*/Internet[6] para a entrega de serviços públicos, os conceitos ideológicos que o sustentam estão muito além da prestação de serviços on-line na Internet. A dimensão dos serviços eletrônicos foi a primeira a ser visível pela sociedade dentro do conceito de governo eletrônico.

Ao longo dos últimos dez anos, praticamente nesta primeira década do século XXI, o e-gov abrange conceitos que incluem inúmeras aplicações de tecnologia da informação: a melhoria nos processos da administração pública, o aumento da eficiência, a melhor governança[7] em TI, as consultas públicas para a elaboração e o monitoramento das políticas públicas, a integração entre governos e a democracia eletrônica, e a adequação dos conteúdos e da linguagem para veiculação na Internet.

O estabelecimento de estruturas de governança é imperativo para o sucesso dos projetos de mudança nas organizações públicas, as quais devem estabelecer claramente os direitos e processos decisórios, bem como a forma de operacionalização deve identificar e organizar as fontes e os fluxos de informações que deverão subsidiar as decisões (CAPUANO, 2007).

---

[5] Ou, como são também denominadas, "tecnologias da informação e das telecomunicações" (TIC). Doravante, será utilizada a expressão "tecnologia da informação", ou simplesmente "TI".

[6] A Internet, a rede mundial de computadores, é o meio de comunicação e não de publicação de conteúdos, que é a *World Wide Web (WWW)*. Entretanto, o termo "Internet" tem se popularizado como entendimento da rede e seus conteúdos.

[7] "Governança" pode ser entendida como "capacidade de decidir sobre as prioridades e determinar a sua realização". Vem sendo relacionado ao conhecido conjunto de mecanismos de mercado (*Control Objectives for Information* – COBIT), sendo entendido como o repasse para as áreas de negócio sobre a decisão quanto às prioridades no que se relaciona à tecnologia de informação, em especial as prioridades sobre o desenvolvimento dos sistemas de informação, contrastando com o modelo anterior em que a área de tecnologia da informação decidia sobre as prioridades.

A democracia eletrônica refere-se principalmente ao aumento da capacidade de transparência, da participação eletrônica e de *accountability*[8]. Bobbio (2004) define democracia como um método de governo, "um conjunto de regras (primárias ou fundamentais) que estabelece quem está autorizado a tomar as decisões coletivas e com quais procedimentos", o que envolve um processo decisório com a participação ampla dos envolvidos. Desse modo, a democracia eletrônica pode ser entendida como uma extensão desse conceito, em que a participação ampla é impulsionada com o uso da tecnologia da informação, opondo-se a formas de governo autocrático.

Os registros, com seus dados estruturados e organizados, precisam ser codificados numa linguagem adequada para a sua difusão na Internet e de forma a permitir sua visualização nos computadores, o que se consegue utilizando as linguagens HTML, a XML (*eXtensible Markup Language*) e outras variantes (ROBREDO, 2005). A reengenharia da Ciência da Informação – ou "revisitação", usando o termo de Robredo (2003) – ampliou as antigas classificações (taxonomias, tesauros) e tornam-se cada vez mais importantes no mundo da busca "caótica" via "engenhos de busca".

Paralelamente a esse cenário, instituições buscaram organizar o caos e estabelecer padrões, embora nem sempre observados. Aos atributos originalmente idealizados para representar um objeto informacional agregaram-se novos elementos e qualificadores (*qualifiers*), a exemplo da DCMI (*Dublin Core Metadata Initiative*) e do RDF (*Resource Description Framework*), que introduziu, por exemplo, atributos como o endereço eletrônico (*Uniform Resource Locator* – URL) e a disponibilidade em várias línguas. A URL passou a ser um atributo essencial na estrutura do hipertexto e hoje faz parte do cotidiano das pessoas ao referenciar quaisquer documentos em meio eletrônico localizado na rede mundial de computadores.

Além desses temas, tão variados quanto amplos, as ações com o objetivo de reduzir a exclusão digital (por meio de políticas públicas que visem à promoção da chamada inclusão digital) são abordadas como uma importante dimensão dos programas de e-gov (CHAHIN; CUNHA; PINTO, 2004; FERRER; SANTOS, 2004).

Desse modo, o governo eletrônico configura-se como um veículo dinâmico de informações que permite operacionalizar o ciclo de prestação de informações e serviços com o uso da tecnologia da informação nos diversos níveis da administração pública, como coadjuvante no processo de organização e disponibilização de informações e de prestação de serviços, transparência pública e participação dos cidadãos nos planos e projetos governamentais.

Remete ao desafio de melhor compreender a sociedade em rede, a estrutura de gestão, os meios para captação, organização, operacionalização, transferência e entrega dos conteúdos informacionais envolvidos, novas formas de prestação de serviços não presenciais e novos canais de relacionamento com a população, dentre outros. As transformações tecnológicas e administrativas do trabalho e das relações produtivas passam a centrar-se em empresas em rede, que evoluíram rapidamente para dentro de nossas casas (CASTELLS, 2001).

Do lado da sociedade, reflete o interesse pela satisfação de necessidades através desse amplo veículo de comunicação e cooperação em rede, entre os atores que nela interagem (governos, servidores, cidadãos, empresas, órgãos e entidades, em âmbito nacional e internacional nos diversos níveis de governo), bem como pela abertura de uma comunicação bidirecional entre governo e sociedade.

O governo eletrônico não reside somente na presença na Internet, através de um sítio (página eletrônica) ou portal, mas em um amplo programa de inserção da tecnologia da informação na ação pública e nos intra e inter-relacionamentos entre os entes públicos, privados e as pessoas da sociedade, possibilitando uma efetiva troca de informações e de participação, tornando implícito para sua viabilização um posicionamento aberto dos governos para a transparência e para a democracia.

[8] *Accountability* refere-se à transparência pública dos gestores na prestação de contas e responsabilização por seus atos.

1.2 Benefícios do e-gov

Os potenciais benefícios do e-gov são (DIAS, 2006):

- a prestação de melhores serviços – conveniência, economia de tempo, qualidade, confiabilidade, usabilidade e variedade de canais de atendimento;
- administrações públicas mais eficientes – pela informatização e possibilidade de entrega, com menor intervenção humana, de informações e serviços que podem ser automatizados, bem como pela concentração dos pontos de acesso e da integração dos serviços;
- novas relações entre o governo e a sociedade – pela integração horizontal dos serviços de forma que a estrutura organizacional do governo seja irrelevante ao cidadão, bem como na maior transparência das ações e gastos do governo; e
- outros benefícios – a exemplo de habilitar a população a ingressar na sociedade da informação.

Pode-se dizer que os benefícios esperados para o governo eletrônico no Brasil são o fim das filas, a redução da redundância de informações e dos formulários mal planejados. Embora o governo eletrônico tenha o potencial de beneficiar os processos administrativos e operacionais, reduzindo custos e promovendo a transparência, uma significativa parcela dos serviços para empresas e cidadãos não evoluíram muito. Haveria ainda um conjunto expressivo de funções do Estado que poderiam se beneficiar do potencial da Internet, a exemplo de compras (onde o Estado responde por mais da metade das compras realizadas no Brasil), prestação de serviços (matrícula na escola pública, marcação de consultas e exames médicos na rede pública, etc.), regulação de práticas, consultas democráticas, arrecadação, gestão de funcionários e gestão administrativa (COSTA, 2007).

Com relação à gestão do conhecimento, tão necessária na economia baseada na informação e no conhecimento, os benefícios ocorrem tanto para os cidadãos, quanto para as organizações. Para os cidadãos, os benefícios incluem melhores serviços, melhores escolhas, maior personalização e *accountability* de como o dinheiro do cidadão será gasto. Para a organização, a gestão do conhecimento provê o benefício maior de impulsionar o desempenho organizacional com o objetivo de incrementar a eficiência e a inovação. Mas, para que esses benefícios ocorram, o processo de retaguarda deve estar devidamente estabelecido (NAÇÕES UNIDAS, 2008).

1.3 Obstáculos do e-gov

No Brasil, o progresso do governo eletrônico depara-se com significativas dificuldades e avança gradativamente, uma vez que se apresenta como um instrumento potencial de melhoria da visibilidade da ação e dos serviços públicos, mas esbarra em inúmeros obstáculos, dentre eles: recursos escassos para uma infinidade de prioridades; dificuldades de infraestrutura; contrastes culturais e socioeconômicos dos extratos da população; e a dimensão geográfica do país.

O avanço observado no governo federal e na maioria dos governos estaduais tem, de fato, se refletido em oportunidades ao alcance da população nos grandes centros urbanos. Foram reduzidas drasticamente as idas das pessoas aos bancos, pode-se resolver boa parte dos compromissos ou agilizá-los com o uso do telefone celular; um computador ligado à Internet tornou-se indispensável para as pessoas em atividades administrativas e intelectuais nas grandes cidades, mas, mesmo assim, ainda não se constitui em um recurso ao alcance da maioria dos cidadãos fora dos grandes centros urbanos do país. Mesmo nestes conglomerados urbanos, a exclusão digital é significativa e caminha ao lado da exclusão social e dos graves problemas envolvidos na questão.

1.4 Literatura, pesquisas e projetos sobre e-gov

A literatura nacional começa a sedimentar autores dedicados ao tema como Chahin, Cunha e Pinto (2004), Ferrer e Santos (2004) e Knight, Fernandes e Cunha (2007a). Os temas de uso da tecnologia da informação em geral e do e-gov em particular têm despertado também o interesse de

inúmeros pesquisadores da academia, tais como (SIMÃO, 2004), (MEDEIROS, 2004), (DIAS, 2006), (FREITAS, 2006), (OLIVEIRA, 2006) e (GONÇALVES, 2006), (AGNER, 2007), (BRAGA, 2008), (ARAKAKI, 2008), (RIECKEN, 2008a), (LIMA, 2008) bem como de governos e consultorias. Iniciativas em nível nacional, como o projeto denominado "e-Brasil" (KNIGHT; FERNANDES; CUNHA, 2007a) (KNIGHT; FERNANDES; CUNHA, 2007b), pretendem contribuir para a questão social no Brasil e, simultaneamente, aprimorar a competitividade da economia com o uso intensivo da tecnologia da informação.

Do ponto de vista da academia, a preocupação teórica com a questão da sociedade da informação e do governo eletrônico tem por objeto investigar o porquê das coisas, a consciência política dos envolvidos, para onde se vai ou onde se poderia chegar com determinadas ações que impactam diretamente tanto o governo quanto a sociedade, remetendo – desse modo – à tentativa de se explorar a realidade para, analisando os pontos comuns de diversas experiências, desenvolver metodologias que propiciem o alargamento do tema, bem como fortalecer o referencial teórico e o corpo de autores que se dedicam ao tema.

Iniciativas no nacional, como o projeto denominado "e-Brasil" (KNIGHT; FERNANDES; CUNHA, 2007a) (KNIGHT; FERNANDES; CUNHA, 2007b), pretendem contribuir para a questão social no Brasil e, simultaneamente, aprimorar a competitividade da economia com o uso intensivo da TI. O projeto e-Brasil visa à implementação do governo eletrônico no contexto de outras políticas e ações. O horizonte temporal inicial proposto é de um mandato nas esferas federal e estadual (podendo-se projetar para dois mandatos), que terminaria em 2014, e o início de um hipotético terceiro mandato subsequente, em 2015, o ano limite para alcance dos Objetivos do Milênio (ODM) das Nações Unidas (*United Nations* – UN).

Do ponto de vista da academia, a preocupação teórica com a questão da sociedade da informação e do governo eletrônico tem por objeto investigar o porquê das coisas, a consciência política dos envolvidos, para onde se vai ou onde se poderia chegar com determinadas ações que impactam diretamente tanto o governo quanto a sociedade, remetendo – desse modo – à tentativa de se explorar a realidade para, analisando os pontos comuns de diversas experiências, desenvolver metodologias que propiciem o alargamento do tema, bem como fortalecer o referencial teórico e o corpo de autores que se dedicam ao tema.

Para a Ciência da Informação, o governo eletrônico e a sociedade da informação onde o tema se insere tornam-se objetos contemporâneos para o desenvolvimento de pesquisas na área, sob as diversas possibilidades de estudo: do gerenciamento, representação e organização da informação e do conhecimento, entre outros aspectos.

Adicionalmente, uma pesquisa acadêmica pode eventualmente gerar um produto de utilidade prática e métodos. Cite-se, por exemplo, o modelo de decisão baseado em critérios objetivos com possibilidade de apoiar os governos estaduais a progressivamente incentivarem o avanço de governo eletrônico nos municípios (RIECKEN, 2008a). O modelo não tem o propósito de utilização imediata nem de ser um modelo fechado, mas sim, o ponto de partida para se refletir sobre o motivo e oportunidade desse avanço.

Nos países em que somente uma pequena parcela da população consegue galgar os níveis mais avançados do estudo e da pesquisa em ciência e tecnologia, é importante que os resultados de estudos acadêmicos (por exemplo, oriundos de dissertações, teses e grupos de pesquisas científicas) possam cumprir não apenas o rito da ciência acadêmica, mas também um enfoque prático. Desta forma, produtos práticos podem, de certo modo, favorecer a redução da desconexão entre a ciência e sua aplicação por meio da tecnologia.

A discussão sobre governo eletrônico não pode ser separada de temas como sociedade da informação, transparência, cidadania, democracia, inclusão, poder político, entre outros, necessitando – para tanto – se apropriar dos métodos de inúmeras ciências e/ou disciplinas para o desenvolvimento de estudos na ótica da Ciência da Informação. "Mais importante do que os resultados em si (posto que dinâmicos e limitados ao contexto), são os métodos a maior contribuição de uma pesquisa científica. A característica distintiva do método é a de ajudar a compreender, no sentido mais amplo, não os resultados da investigação científica, mas o próprio processo de investigação" (KAPLAN apud LAKATOS; MARCONI, 1992).

A pesquisa científica tradicional baseia-se em estruturas e seu detalhamento, em categorias cada vez mais pormenorizadas. Entretanto, autores modernos[9] vêm se afastando das visões fragmentadas de perceber o mundo, caminhando para a constatação de que "a soma das partes não faz o todo", uma vez que, reunidos os vários sistemas (estudados isoladamente), a resultante possui um comportamento com potencial próprio que extrapola em muito a mera soma das partes.

Mas, não só os sistemas têm potencialmente uma dinâmica distinta, como formam de por si só um todo, um instrumento, que também é dinâmico. Então, mesmo os aspectos considerados válidos atualmente vão se tornar obsoletos amanhã em decorrência natural de um percurso contínuo. A melhor maneira parece ser a de se considerar a informação e o conhecimento (bem como a Ciência da Informação que os têm como objeto de estudo), como "quadros" onde os limites não são mais fixos.

Essa é a razão pela qual em algumas universidades os alunos da pós-graduação são incentivados a desenvolverem suas dissertações e teses baseadas no pensar sistêmico.

2. A CONTEMPORANEIDADE DO TEMA GOVERNO ELETRÔNICO

A Seção 2 tem por objetivo contextualizar a contemporaneidade do tema governo eletrônico, dando uma visão geral dos níveis evolutivos ou estágios em que o mesmo se desenvolve. Também apresenta o tema e-gov no contexto mundial e o modo como é diferentemente percebido e medido nos vários países.

2.1 Estágios do governo eletrônico

Governo eletrônico é um tema contemporâneo, tanto do ponto de vista do seu referencial teórico e dos métodos, quanto em relação aos aspectos operacionais, em especial, quanto à tecnologia da informação.

Em 2004[10], poder-se-ía dizer que a literatura internacional era esparsa, concentrada em alguns países desenvolvidos, enquanto que a literatura nacional era praticamente inexistente. Ao final da primeira década de governo eletrônico no Brasil, o cenário mudou significativamente. Não se poderia mais chamar governo eletrônico de fenômeno ou mecanismo, tão pouco utilizar denominações similares, egressas da engenharia, ou mesmo visões mecanicistas.

Governo eletrônico só pode ser compreendido sob um ponto de vista sistêmico. O tema está definitivamente incrustado nas pesquisas acadêmicas, de modo transversal em diversas ciências e disciplinas, tanto no que diz respeito à quantidade de pesquisas, quanto qualitativamente pela diversidade de óticas dos estudos. Isso torna o trabalho de pesquisa especialmente árduo, posto que tratar de um tema tão atual e dinâmico obriga o pesquisador a buscar, periodicamente atualizar o embasamento teórico e desenvolver (ou adequar) os métodos.

Consoante pesquisa realizada na literatura pelo comparativo das principais listas de classificação *(ranking)* internacionais do e-gov, os estágios ou níveis de progresso em governo eletrônico diferem entre três ou mais níveis progressivos. O *Quadro 1* permite visualizar esses estágios, possibilitando também cotejar a equivalência aproximada entre as diferentes visões, sendo que a classificação proposta na primeira coluna deriva da análise das diferentes categorias observadas nas demais.

No seu estágio mais avançado, estima-se que o governo eletrônico tende a uma interação/integração plena, de modo transversal abrangendo a área pública em seus vários níveis e esferas, e à antecipação e personalização das necessidades dos usuários, tornando-se (por assim dizer) "transparente"

[9] Como Morin (2003), que desenvolveu a linha de pensamento sobre a complexidade na formação social levando em consideração fatores culturais e a história biológica.
[10] Quando a pesquisa de doutorado (RIECKEN, 2008a) teve início.

ao cidadão. A idéia seria a de desonerar os cidadãos de prestarem informações repetitivas (e por vezes exageradas) comparativamente à informação ou benefício resultante para o mesmo. Isso seria possível uma vez que a interação/integração mencionada possibilitaria as informações necessárias e suficientes relativas aos episódios do ciclo de vida relacionados às interações com a esfera pública, e permitiria a padronização dos eventos e a simplificação dos processos, até mesmo com maior precisão e segurança comparativamente ao processo convencional (não eletrônico).

Quadro 1. – Comparativo entre os diferentes níveis de e-governo[11]

	Classificação das Nações Unidas em 5 estágios	Classificação da União Européia / Capgemini em 4 estágios	Classificação Accenture em 3 estágios	Classificação segundo Chahin *et al.* (2004)	Classificação adotada por Riecken (2008a)
Sem e-gov					0. ausência na internet
Presença institucional	Presença emergente (*emerging*)	Informação	Público	Institucional	1. Presença institucional na internet
	Presença expandida ou estendida (*enhanced*)	Informação em uma via (*download* de formulários)		Informativa	2. Presença expandida na internet
Presença interativa	Presença interativa (*interactive*)	Informação em duas vias (formulários eletrônicos)	Interativo	Interativa	3. presença interativa
Presença transacional	Presença transacional (*transactional*)	Transação (processamento eletrônico completo)	Transacional.	Transacional	4a. Presença transacional emergente
					4b. Presença transacional expandida
Presença em rede	Presença em rede (*networked*)			Integrativa	5a. Presença em rede emergente
					5b. Presença em rede expandida

2.2 Governo eletrônico no mundo

Segundo estudos das Nações Unidas, uma tendência de reformas no setor público emergiu em muitos países nos últimos anos, devido principalmente às aspirações dos cidadãos ao redor do mundo que colocam novas demandas para os governos. O sucesso dos governos, cada vez mais, é medido com base nos benefícios oferecidos aos cidadãos, setor privado e comunidades. Estes clientes governamentais demandam um nível alto de eficiência, transparência na prestação de contas públicas e um enfoque renovado na entrega de melhores serviços e resultados (NAÇÕES UNIDAS, 2008).

Vários países em todo o mundo vêm tentando revitalizar a administração pública com o objetivo de torná-la mais pró-ativa, eficiente, transparente e, especialmente, mais orientada a serviços. Para realizar essa transformação, os governos têm introduzido inovações nas estruturas organizacionais, nas

[11] Chahin *et al.* (2004) sugere uma tipologia adaptada da classificação do Gartner Group e da Unidade de Missão Inovação e conhecimento (UMIC), de Portugal. Ver também "Manual de Implementação de Serviços Públicos Eletrônicos", do Governo do Estado de São Paulo, em (SÃO PAULO, 2009a). Tipologia das Nações Unidas consoante (NAÇÕES UNIDAS (2004). Ver também, FERNANDES (2000).

práticas utilizadas, nas competências e no modo como mobilizam, desenvolvem e utilizam o capital humano e a informação, os recursos tecnológicos e financeiros para a entrega de serviços aos cidadãos. O uso apropriado da tecnologia da informação passa a desempenhar um papel crucial para o avanço das metas do setor público, contribuindo para um ambiente que possibilite o crescimento social e econômico (NAÇÕES UNIDAS, 2008).

Estima-se que haja dezenas de pesquisas globais sobre governo eletrônico *(ranking)*, sendo conduzidas pelo mundo afora, a partir de 2000, podendo-se destacar:

- ACCENTURE – Consultoria em Gerenciamento e Tecnologia de Informação (*Management and Information Technology Consultancy – ACCENTURE*): abrange 21 países e tem por foco serviços selecionados voltados aos negócios e aos cidadãos;
- Universidade de Brown – Centro de Políticas Públicas da Universidade de Brown (*Centre for Public Policy of the Brown University – CPP-BU*): abrange 198 (incluindo territórios) países pesquisados e, tem por foco, um conjunto de indicadores envolvendo avaliação de progresso nos diferentes poderes, bem como nas áreas de arrecadação, educação, administração, assuntos estrangeiros, investimentos, transporte, aspectos militares e de turismo;
- *Economic Intelligence Unit* (EIU): envolve 65 países com foco em serviços voltados aos negócios e cidadãos, classificados em seis categorias;
- *Taylor Nelson Sofres* (TNS): envolve 31 países com foco no uso da Internet e de serviços on-line governamentais;
- *UN Global E-government Survey*: pesquisa coordenada pelo Departamento de Assuntos Econômicos e Sociais (*Department of Economic and Social Affairs* – UNDESA) das Nações Unidas: envolve os 192 países membros das Nações Unidas com foco em serviços on-line nas áreas de educação, saúde, previdência social, finanças e trabalho/emprego;
- *World Economic*: o fórum econômico mundial da Universidade de Harvard engloba pesquisa de 104 países e tem por foco verificar a prontidão (*readiness*) da sociedade para operar em rede; e
- Digital governance in municipalities worldwide: pesquisa coordenada por Holzer e Kim (2008) que avalia a cada dois anos 100 grandes cidades, tendo por foco a avaliação da governança digital com base nos portais municipais.

Outros países – a exemplo do Canadá, da Austrália, do Japão e de alguns países asiáticos (em especial os denominados Tigres Asiáticos) – vêm despontando com portais governamentais nacionais, oferecendo, em paralelo, uma quantidade expressiva de materiais sobre suas experiências e práticas.

Analisando a literatura internacional, os estudos teóricos estão concentrados, principalmente, nos Estados Unidos e nos países mais desenvolvidos.

O portal oficial do governo eletrônico americano (ESTADOS UNIDOS, 2009a), a página da Casa Branca (ESTADOS UNIDOS, 2009b) e o portal do Government Accountability Office – GAO (ESTADOS UNIDOS, 2009c) apresentam expressivo material sobre a estratégia de implementação e progresso do governo eletrônico americano, bem como sobre a Arquitetura Empresarial Federal (Federal Enterprise Architecture - FEA) adotada, em uma visão oficial.

Nos Estados Unidos, a ênfase consiste nas questões de portais, de aspectos operacionais relacionados à tecnologia da informação, inclusive nos aspectos da convergência tecnológica, o que remete a questões de inclusão digital por meio de novas formas de prestação de serviços digitais, como a telefonia móvel e a TV digital.

Já na Europa, o tema tem um contorno mais teórico e social (portanto, menos operacional), vinculando-se a questões de cidadania e democracia. Destacam-se os estudos da Organização para a Cooperação e o Desenvolvimento Econômico – OCDE (Organization for Economic Co-operation and Development - OECD) que medem o acesso e uso da tecnologia da informação em residências e por indivíduos, dado o forte interesse político por assun-

tos ligados ao potencial da TI (e, em particular, à Internet), à democratização do acesso e ao impacto para mudar a sociedade. Tal modelo de pesquisa vem sendo estendido a inúmeros países, inclusive o Brasil.

Lançado em 2001, o projeto sobre governo eletrônico da OCDE (OECD E-Government Project), avalia como os governos podem melhor explorar a tecnologia da informação para o alcance dos princípios da boa governança e dos objetivos das políticas públicas. O projeto produz relatórios sobre as melhores práticas e desenvolve estruturas para temas específicos como análise de custo/benefício, de e-serviços e índice de uso efetivo dos serviços eletrônicos pelos clientes (*take-up*) (OCDE, 2009).

Para o plano de ação evolutivo do avanço do acesso e uso da tecnologia da informação e do governo eletrônico na Europa, no denominado i-2010,estão sendo propostos novos indicadores e um modelo holístico de medidas, conforme mostra a *Figura 1.*

Portanto, prontidão ou disponibilidade para o governo eletrônico (isto é, o quão prontos estão os governos, em seus diferentes níveis, para o uso da tecnologia da informação na oferta de serviços informacionais, dentre outros) refere-se aos principais fatores intervenientes para a sua concretização, do lado da oferta: 1) liderança e estratégia (*leadership and strategy*); 2) aporte financeiro (*funding*); 3) pessoas e habilidades (*people and capability*); 4) tecnologia e conectividade (*technology and connectivity*); 5) envolvimento efetivo dos patrocinadores (*stakeholder involvement*); e 6) organização e mudança (*organization and change*).

Figura 1. Novo modelo de medidas holísticas proposto pela União Européia no plano de ação i-2010
Fonte: adaptação e tradução da figura constante à pág. 17 do relatório (CAP GEMINI, 2006).

Essa disponibilidade (prontidão) propicia a entrega dos serviços informacionais e a disponibilidade e qualidade desses serviços, para que uma mudança mais profunda no governo possa ocorrer. Essa nova fase é denominada "transformação do governo", momento no qual os fatores intervenientes apontam para: a necessidade da integração da retaguarda (o que requer um expressivo esforço de integração dos sistemas legados para que o conjunto de processos administrativos, operacionais e logísticos passe a operar "em linha"), do compartilhamento dos serviços, do gerenciamento dos dados e das relações com os cidadãos.

Nos países desenvolvidos, onde a maioria da população possui um computador com acesso à Internet e onde os serviços e informações oferecidos pelos governos alcançam um avançado nível nas transações (na sofisticação, na usabilidade, etc.), não se fala mais em "prontidão" do governo e da sociedade, mas sim, em transformação do governo (*transformational government*) para um estágio virtuoso de interação, onde o governo eletrônico, em um extremo de evolução, torna-se transparente, antecipando-se aos desejos dos cidadãos, ou até mesmo "desnecessário" (em determinadas situações), pela eliminação de exigências que extrapolem a razoabilidade do benefício, do direito ou do reconhecimento almejados. Mas que grau de requisitos (tais como: integração, preparação da retaguarda do governo e da sociedade, satisfação de necessidades, maturidade de modelo de gestão, etc.), faz-se necessário para atingir esse grau de e-gov virtuoso?

Observação empírica quanto às tentativas de implantação de novos modelos de gestão (com profunda modernização dos sistemas de informações legados nos órgãos do poder executivo brasileiro) indica que não faltam recursos financeiros, em tecnologia e conectividade, restando frágeis outros aspectos.

Já nos demais países, não tão desenvolvidos (afora as "ilhas" nos grandes centros urbanos), pressionados pela escala de demanda e envolvendo as esferas "afastadas" da população (governos federal e estadual), governo eletrônico ainda é – de um modo geral – uma utopia.

2.3 Governo eletrônico nas grandes cidades do mundo

Estudos vêm sendo desenvolvidos voltados para avaliar os grandes municípios no mundo. Um exemplo disso é a pesquisa, realizada em 2007, denominada Governança Digital nas Municipalidades no Mundo (Digital Governance in Municipalities Wordwide) (HOLZER; KIM, 2008).

A referida pesquisa replica trabalhos similares realizados em 2003 e 2005 e consiste na avaliação de *websites* nas municipalidades em termos de governança digital, classificando-as em uma escala global. A governança digital inclui ambos: o governo digital (entrega de serviços públicos) e a democracia digital (participação dos cidadãos na governança). Especificamente, a pesquisa analisa a segurança, a usabilidade e os conteúdos dos portais, o tipo de serviços on-line oferecidos e a participação dos cidadãos (HOLZER; KIM, 2008).

A pesquisa selecionou cidades do mundo com base no tamanho da sua população e no número total de indivíduos que se utilizam da Internet. Os 100 maiores países foram identificados usando os dados da União Internacional de Telecomunicações (International Telecommunication Union - ITU, uma organização da família das Nações Unidas). A maior cidade, segundo a sua população, em cada um desses países, foi então selecionada para o estudo e utilizada como representante para todas as cidades em seu respectivo país.

Das 100 cidades selecionadas, em 86 delas foram identificados os *websites* municipais. Isso significa um aumento, comparativamente à pesquisa de 2005, quando 81 das 100 maiores cidades apresentavam portais. As 20 primeiras classificadas, na ordem decrescente da escala a partir da primeira colocada, foram: Seul (República da Coréia), Hong Kong (Hong Kong), Helsinque (Finlândia), Singapura (Singapura), Madri (Espanha), Londres (Reino Unido), Tóquio (Japão), Bangkok (Tailândia), Nova York (Estados Unidos), Viena (Áustria), Dublin (Irlanda), Toronto (Canadá), Berlim (Alemanha), Zurique (Suíça), Praga (República Checa), Buenos Aires (Argentina), Bratislava (República Eslováquia), Sydney (Austrália), Amsterdã (Países Baixos) e Roma (Itália).

Na América Latina, as maiores cidades identificadas pela pesquisa em grandes municípios foram: Buenos Aires (Argentina), Caracas (Venezuela),

Guaiaquil (Equador), La Paz (Bolívia), Lima (Peru), Montevidéu (Uruguai), Santa Fé de Bogotá (Colômbia), Santiago (Chile) e São Paulo (Brasil). Somente a cidade de Buenos Aires foi classificada entre as 20 primeiras. A cidade de Buenos Aires ficou na décima sexta. posição na classificação geral; como a cidade melhor colocada da América do Sul, Caracas, na quadragésima primeira.; Santiago, na quadragésima sexta, e a cidade de São Paulo, na qinquagésima quarta...

2.4 Governo eletrônico no Brasil

Esta subseção tem por objetivo apresentar de modo sucinto o marco inicial do governo eletrônico brasileiro, a situação do país na classificação *(ranking)* internacional e a situação geral do e-gov nas diversas esferas de governo.

2.4.1 Marco inicial do governo eletrônico no Brasil

No Brasil, o interesse pelo governo eletrônico teve início em meados dos anos 90, quando a Internet surgiu no país (ROSA, 2005). Logo depois, em 1996/1997, o Governo Federal brasileiro ingressou na rede mundial, com o início do recebimento das declarações do imposto de renda pela Internet e com a divulgação de informações por órgãos e entidades da Administração Pública Federal. Em 1998, foi sancionada a Lei Geral de Telecomunicações, que organizou o setor, criou a Agência Nacional de Telecomunicações - ANATEL, dando início à abertura do setor de telecomunicações à iniciativa privada. A referida Lei dividiu o serviço em três regimes jurídicos: público, privado e misto, possibilitando, no ano seguinte, a privatização da telefonia e gerando um significativo avanço no setor das telecomunicações (RODRIGUES, 2008).

No início, a presença dos órgãos públicos federais na Internet ocorria de modo "estático", com páginas institucionais com pouca ou nenhuma interação. Entretanto, esse cenário mudou em um curto espaço de tempo, apresentando um desenvolvimento extraordinário, sobretudo com a abertura das telecomunicações no país, com o objetivo desafiador de universalização dos serviços.

Pode-se afirmar que o marco institucional mais importante e inspirador do tema foi a publicação do "Livro Verde" (TAKAHASHI, 2000), que elevou a questão da tecnologia da informação a um paradigma ao inseri-la em um campo maior de estudo da sociedade da informação. Desse modo, o livro estendeu o tema a inúmeros eixos temáticos, tais como: a sociedade da informação; mercado, trabalho e oportunidades; universalização de serviços para a cidadania; educação na sociedade da informação; conteúdos e identidade cultural; aplicações governamentais; informações e serviços ao cidadão; infra-estrutura de redes para governo; diretrizes tecnológicas; legislação; gestão estratégica da tecnologia da informação; pesquisa e desenvolvimento; e infra-estrutura avançada e novos serviços (TAKAHASHI, 2000).

O tema difundiu-se cada vez mais, passando a estar inserido na pauta de muitos estudiosos na área da Ciência da Informação, na Administração, no Direito, nas Tecnologias de Informação e das Telecomunicações, dentre inúmeras outras áreas, apresentando, progressivamente, numa vasta literatura, em especial nos próprios portais de governo, instituições e estudiosos sobre o tema.

Takahashi coloca: [...] a construção de uma sociedade da informação democrática no Brasil está visceralmente dependente do apoio à pesquisa em tecnologia de produção e comunicação de conteúdos e da criação de condições para a capacitação universal dos cidadãos para o uso das novas tecnologias. *(TAKAHASHI, 2000, p.61)*

2.4.2 A situação do Brasil na classificação (ranking) internacional

Na pesquisa das Nações Unidas (2008), reconhecidamente uma das mais conceituadas, o Brasil aparece em 45º lugar no índice de prontidão e em 23º no índice de participação, mostrando uma relativa piora comparativamente ao *ranking* 2005. A Argentina superou o Brasil e também o Chile, passando a liderar o *ranking* da América do Sul, consequência – principalmente – do aumento do seu índice de infra-estrutura, verificando-se um maior

incremento de serviços contratados de celular, bem como no aumento do número de computadores pessoais. O Equador e o Paraguai melhoraram suas posições no *ranking* 2008 pelo melhor desempenho na entrega de serviços on-line.

2.4.3 Situação geral do e-gov nas diversas esferas de governo

Atualmente, no Brasil, tanto nos governos federal, estaduais e em geral nas cidades mais populosas, quanto na iniciativa privada, são verificados portais de alta qualidade. Todavia, enquanto na iniciativa privada os portais visam, principalmente, às operações de compra e venda (e, portanto, ao lucro, propósito natural das empresas), nos governos – tanto na administração direta quanto indireta, seja nas esferas federal, estadual ou municipal – os propósitos são de outra natureza, a exemplo da melhoria da eficiência (e, com isso, da imagem da sociedade sobre os serviços públicos); do "tornar-se conhecido"; de dar publicidade aos feitos, ações ou projetos; do aumento da arrecadação por meios de melhores cadastros; de oferecer facilidades de acesso, ou mesmo comunicar atrações no que se refere aos aspectos turísticos e culturais (RIECKEN, 2008a).

No âmago da questão, busca-se levar ao usuário o conhecimento da administração pública, a oferta de serviços, a comunicação de realizações (muitas vezes com interesse político) e, de certo modo, o interesse econômico de redução de custos e de melhoria da eficiência do processo de arrecadação e de seus resultados.

No que se refere aos níveis de governo, os governos federal e estadual são reconhecidos como os mais avançados no uso da tecnologia da informação, mas paradoxalmente mais "distantes" fisicamente da população.

Programas de governo eletrônico apóiam a informatização dos processos, antes em papel. Desse modo trazem benefícios práticos de redução de custos de aquisição e impressão de papéis, da guarda de documentos comprobatórios – que podem ficar armazenados em meio eletrônico de mais fácil recuperação – simplificação de inúmeras obrigações e, principalmente, incentivando o relacionamento com cidadãos, empresas e outras entidades que precisam interagir com o governo. Possibilita, ainda, a agilização do cruzamento de informações, antes dispersas em arquivos documentais, permitindo o aumento da base de monitoramento com vistas à fiscalização e encaminhamentos pertinentes.

Nesse sentido, alguns estados conduzem iniciativas exitosas. O Programa Nota Fiscal Paulista, por exemplo, é um programa típico de estímulo à cidadania fiscal no Estado de São Paulo, evolução do programa Nota Fiscal Eletrônica, que se propõe a devolver 30% do Imposto sobre Circulação de Mercadorias e Prestação de Serviços - ICMS efetivamente recolhido pelo estabelecimento a seus consumidores, sendo um incentivo para que os cidadãos que adquiram mercadorias exijam do estabelecimento comercial o documento fiscal (SÃO PAULO, 2009b).

Já os governos municipais, menos desenvolvidos no que se refere à tecnologia da informação, encontram-se mais "próximos" da população, no sentido da prestação de serviços presenciais. Não se trata apenas de características diferenciadas dos portais dessas entidades, mas das relações humanas envolvidas nesse processo, decorrentes da possibilidade de uma maior proximidade física (no caso dos municípios) e da pressão (no caso das esferas federal, estadual, nas capitais e nos grandes municípios) devido ao aumento da escala de demanda por serviços melhores e mais eficientes (RIECKEN, 2008a).

Como decorrência natural disso, as esferas federal e estadual parecem mais desenvolvidas em seus portais, inferindo-se que, nesses níveis, a via virtual estreita a desconexão física. Já no nível municipal, de um modo geral e afora os grandes municípios, a proximidade física favorece a utilização de mecanismos tradicionais presenciais, em detrimento do meio eletrônico.

A reatividade natural às quebras de paradigma (isto é, da cultura presencial de balcão para a cultura virtual) é um processo não somente de capacitação para as tecnologias, mas também geracional, no sentido de estar naturalmente incrustado nas atividades das pessoas (desde jovens), bem como de se criar uma identidade cultural pela disponibilização de conteúdos úteis e de interesse que propiciem o conhecimento e facilitem as diversas atividades humanas, o que resulta na criação dos valores da sociedade por meio desses novos mecanismos informacionais.

Dwivedi e Michael (2008) analisando empiricamente a influência de variáveis demográficas (idade, gênero e nível educacional) na adoção pelos cidadãos da iniciativa do governo do Reino Unido, denominado "Portal Governamental" (Government Gateway), – cujas variáveis foram coletadas através de uma ampla pesquisa nacional realizada por meio postal – sugere que as variáveis idade e educação, bem como o acesso por meio de banda larga nas residências, influencia o maior uso pelos cidadãos do portal governamental, enquanto a variável gênero, não se mostra significativa. Assim, tem-se que as futuras gerações se utilizarão cada vez mais da tecnologia digital, na medida em que estas estiverem mais presentes em suas vidas, em especial, no processo de formação educacional, pela ampla disponibilização de infra-estrutura de acesso e da disseminação como veículo em todas as atividades humanas.

Questiona-se: em que medida o governo eletrônico está irremediavelmente atrelado a temas como a vontade política de seus dirigentes, a transparência (como caminho de consolidação da democracia e de participação da sociedade) e a cidadania (enquanto requisito necessário para que a interação seja possível)?

A sociedade da informação (ou do conhecimento) é um processo dinâmico, ligado ao modelo das sociedades nas quais ela se insere e herdando seus profundos problemas, o que de certo modo pode frustrar as expectativas de que a tecnologia da informação poderia gerar um coletivo mais harmônico e com menos desigualdade com relação aos ganhos potenciais esperados com essa tecnologia e da disponibilização massiva de informações. Uma sociedade com dirigentes eventualmente pouco éticos, naturalmente não patrocinarão o uso potencial da tecnologia da informação como veículo de transparência e de participação democrática.

Os governos vêm entendendo, equivocadamente, que a sociedade da informação consiste na informatização da sociedade ou a distribuição de computadores (SUAIDEN, 2006). Com essa premissa enviesada, as ações têm caminhado no sentido da distribuição de computadores e de *software* livre, em especial para as escolas públicas. Mas, assim como a distribuição de livros isoladamente não faz uma geração de leitores, a distribuição de computadores com *softwares livres* não fará uma geração de jovens preparados para a sociedade da informação. Isso porque inexiste o planejamento adequado para esse processo, com a preparação daqueles que introduzirão as novas tecnologias de educação e informação, nem o monitoramento do progresso e a avaliação do impacto dessas ações.

Analogamente, é equivocado entender que a informatização das pequenas prefeituras em todo o país, isoladamente, tenha a capacidade para revitalizar essas administrações públicas e as populações que com elas interagem, resolver as questões indissociáveis de governo eletrônico como a ética e a transparência pública, a cidadania e democracia, dentre outras, nem para sozinhas reduzirem a desconexão tecnológica e educacional de comunidades no interior do país.

Jambeiro *et al.* (2007) apontam que a transparência estimula a democratização, dando a todos a possibilidade de conhecer, de criticar e de opinar sobre as ações do governo, otimizando-o e evitando um governo autoritário. Nesse sentido, a tecnologia da informação pode viabilizar um canal moderno e oportuno para o exercício da transparência pública, do lado dos governos, e para a mobilização e participação, do lado da sociedade.

3. RESULTADOS DE PESQUISA ACADÊMICA EM GOVERNO ELETRÔNICO

Esta seção tem por objetivo apresentar, de modo conciso e exemplificativo, os resultados de pesquisa no tema de governo eletrônico (RIECKEN, 2008a), realizada no escopo do programa de Pós-graduação em Ciência da Informação do Departamento de Ciência da Informação e Documentação da Universidade de Brasília (CID/UnB).

A pesquisa foi aplicada aos governos eletrônicos (e-govs) estaduais e municipais. Com relação às experiências nos governos eletrônicos estaduais pesquisados, 9 (nove) dos 27 (vinte e sete) gestores dos governos eletrônicos estaduais responderam à pesquisa. Os resultados sinalizam uma descone xão entre a apropriação da tecnologia da informação por e-govs estaduais da Região Sul/Sudeste, comparativamente aos da Região Norte/Nordeste.

Enquanto os respondentes dos governos estaduais das Regiões Sul/Sudeste não verificam, de um modo geral, impedimentos para o avanço da iniciativa de governo eletrônico, os do grupo da Região Norte/Nordeste apontam inúmeros obstáculos, a exemplo da falta de prioridade para o tema e do desconhecimento dos benefícios por parte daqueles que decidem.

Considerando que essa desconexão regional é verificada também pelos índices sócio-econômicos, pode-se concluir que o governo eletrônico é influenciado pelo desenvolvimento sócio-econômico.

Os estudos empreendidos ressaltam haver (de um modo geral) vontade política para o governo eletrônico, entretanto, sugerem que inúmeros impedimentos ainda precisam ser transpostos, haja vista que passam por ações estruturantes que vão desde a infra-estrutura, a preparação dos agentes e a preparação geracional para o acesso e uso da tecnologia da informação com efetividade, até o apoio dos que estão à frente no tema.

Nas prefeituras municipais pesquisadas, as iniciativas de e-gov concentram-se (afora os municípios das capitais) em estágios diferenciados e, de um modo geral, incipientes, aquém dos resultados da avaliação realizada nos e-govs estaduais.

Quanto à verificação de viabilidade de desenvolvimento de iniciativas de e-gov nas prefeituras municipais, onde tais iniciativas não tiveram início (ressalvadas as limitações da pesquisa, cuja análise se restringe às respostas dos 20 respondentes), os resultados apontam para uma percepção concordante com a oportunidade do desenvolvimento de uma iniciativa de governo eletrônico, desde que sejam superados os entraves para a sua consecução.

Tais entraves estão relacionados, em especial, à falta de acesso e preparo por parte da população, bem como à falta de capacidade técnica, de recursos financeiros e, em alguns casos, de vontade política por parte das prefeituras.

Sugestões simples e oportunas foram oferecidas pelos respondentes, em especial, o desejo de implantação de uma iniciativa de e-gov para a promoção do município no que diz respeito aos aspectos turísticos e culturais.

As pesquisas na literatura e de campo, empreendidas, permitem concluir que, sem o apoio de outras esferas de governo, o governo eletrônico (entendido no seu sentido mais amplo, e não apenas como a simples presença na Internet por meio de um portal institucional) é, no presente momento, inviável na maioria das prefeituras brasileiras.

Enquanto os governos estaduais – mesmo com os diferentes avanços regionais verificados e as deficiências apontadas na pesquisa quanto à baixa eficiência dos canais (tipo "Fale Conosco") – vêm desenvolvendo seus portais, buscando oferecer progressivamente mais e melhores informações, serviços e canais de comunicação para a população, as prefeituras municipais estão, de um modo geral, significativamente aquém do nível de desenvolvimento dos portais estaduais.

Uma questão que se apresenta é se haveria uma forma de os governos estaduais capitalizarem no nível municipal os avanços verificados (de um modo geral) em seus programas de governo eletrônico. Neste sentido, é proposto um modelo de decisão baseado em critérios objetivos (modelo estatístico), que se denominou "Critérios de Priorização"[12], cuja finalidade é a elaboração de uma lista progressiva dos municípios que se destacam, segundo propósitos estabelecidos (populacionais, sociais, econômicos, culturais, etc.) a fim de incentivá-los a progressivamente integrar uma rede intra-estadual de informações, serviços e comunicação eletrônicos, capitalizando no nível municipal os avanços já percebidos na esfera estadual e oferecendo, portanto, experiências e práticas que possibilitassem maiores chances de sucesso para as iniciativas municipais de e-gov.

[12] A palavra "priorização", de certo modo um neologismo amplamente utilizado na língua portuguesa, foi preferida à locução verbal equivalente ("dar prioridade"). Tal uso pode ser facilmente constatado em pesquisas nos motores de busca comuns de mercado. Desse modo, as tecnologias de informação propiciam uma verificação rápida da capacidade reprodutiva da língua, que, a partir de uma palavra ou de um radical, possibilitam por adjunção de um sufixo, aumentar a capacidade e a diversidade lexicais da língua. Neste caso, a aparente necessidade de substituir uma locução verbal (dar prioridade) por uma única palavra permitiu a criação dos verbos priorizar e outros derivativos. Outros casos semelhantes podem ser encontrados na língua, sem que haja qualquer desrespeito pelas regras de boa formação do português (PRIBERAM INFORMÁTICA, 2010).

Sendo os resultados da pesquisa de campo (aplicada às prefeituras municipais) limitados ao conjunto de respondentes – embora, segundo Yin (1988), sejam úteis para ilustrar algumas situações –, o modelo de priorização das prefeituras municipais para e-gov (RIECKEN, 2008a) recorreu às bases estatísticas oficiais do Instituto Brasileiro de Geografia e Estatística (IBGE) e do Programa das Nações Unidas para o Desenvolvimento (PNUD) para reunir os dados no nível municipal necessários à estruturação do modelo Critérios de Priorização.

A construção desse modelo foi baseada nos estudos de Saaty (1991), sendo o método de hierarquização conhecido como *Analytic Hierarchy Process* (*AHP*) (em português, Processo Hierárquico Analítico). O método AHP foi aplicado a um conjunto de aproximadamente duas centenas de variáveis técnicas, geodemográficas, culturais, de gestão e relativas ao nível de desenvolvimento humano, extraídas de séries estatísticas oficiais, o que resultou em um total de pontos para cada município, calculado com base na ponderação desse conjunto de variáveis.

O resultado final é o modelo Critérios de Priorização, com o objetivo de que os governos estaduais possam coordenar um processo a que se denominou de "municipalização virtual gradativa" dos estados por meio de redes sistêmicas de informações, serviços e comunicação, presumindo a inclusão de todos os municípios nesse processo. Atribui-se prioridade no *ranking* ao ingresso inicial dos municípios com aspectos mais favoráveis para a implantação da iniciativa de governo eletrônico, de modo que estes municípios possam, em um segundo momento, já com a infra-estrutura sedimentada e com o preparo necessário, cooperarem no apoio ao ingresso de outros municípios, favorecendo afinidades regionais e interesses comuns.

Uma rede intra-estadual de informações, serviços e comunicação eletrônica é proposta, instigando um pensar sistêmico a partir do momento em que se colocam objetivos de longo prazo (como o meio ambiente e as futuras gerações) e não apenas um cliente visível e imediato (isto é, a sociedade atual).

Como ficou evidenciada na referida pesquisa, a adaptação cultural às novas tecnologias é diretamente dependente da vontade política e compreensão dos benefícios por parte dos dirigentes, bem como da formação adequada dos gestores e monitores para esse processo, bem como da crescente produção de serviços e informações eletrônicos com utilidade para essas realidades. Isso requer uma visão de futuro, vontade firme e uma adequada preparação dos dirigentes e demais envolvidos para um processo de inclusão digital, também das administrações públicas do interior do país, de modo social e ambientalmente responsável e com alicerce na educação para a cidadania.

O que se propõe com o modelo sugerido é, principalmente, uma estratégia para integrar virtualmente seu imenso território de diversidades, com repercussões de longo prazo.

As diferentes realidades não permitem concluir sobre uma solução única de modelo de governo eletrônico (institucional, de arquitetura do portal, de conteúdos, de operacionalização, de governança, etc.), mas sim, da necessidade de se adequar os métodos e as práticas a cada caso concreto. Desse modo, os principais intentos e resultados da pesquisa empreendida (como a estrutura de maturidade em governo eletrônico, o modelo de decisão – baseado em critérios objetivos – para apoiar a disseminação de governo eletrônico nos municípios e a proposta da estruturação da rede intra-estadual de informações, serviços e comunicação eletrônicos) não têm o propósito de ser únicos, tão pouco algo acabado.

4. O POTENCIAL DE PESQUISAS SOBRE GOVERNO ELETRÔNICO

A presente seção trata da apresentação, em especial para estudantes de graduação e de pós-graduação, ou mesmo para pesquisadores ou interessados, de temas potenciais para estudos posteriores relacionados ao acesso e uso da tecnologia da informação por governos e pela sociedade, instigando-os e despertando-os para um conjunto expressivo de temas contemporâneos.

É bastante vasto o conjunto de temas para pesquisas, em geral, em Ciência da Informação (RIECKEN, 2006) e sobre o acesso e uso da tecnologia da informação e do governo eletrônico, em particular. Segundo Capuano (2008), existe atualmente uma profusão de ideias sobre as conexões entre ser-

viços públicos, democracia, cidadania, informação, tecnologia da informação e administração pública, temas que têm despertado o interesse da academia como área de conhecimento multidisciplinar.

O objetivo de toda pesquisa é o de avançar (um pouco que seja) na fronteira do conhecimento. A trajetória de sua execução e a consecução dos seus objetivos remete, muitas vezes, à necessidade de superação de novos desafios em decorrência do surgimento de novas questões.

Os temas para futuras pesquisas foram extraídos de Riecken (2008a), cuja tese implicou no desenvolvimento e teste de modelos e instrumentos necessários à consecução de seus objetivos. Entretanto, em pesquisas sociais, não há um ambiente fechado onde o teste dos modelos se encerre em si próprio, abrindo desse modo, cenários para novas indagações e, portanto, novas pesquisas, tanto no sentido do aperfeiçoamento (ou adequação) de modelos e instrumentos elaborados, como para o desenvolvimento de novos estudos com diferentes abordagens.

Com relação aos métodos de avaliação do progresso em governo eletrônico, nos diversos níveis de governo, são identificadas as seguintes possibilidades de pesquisas:

a) sugere-se o aprofundamento de estudos, em especial quanto ao detalhamento dos indicadores propostos para o estudo comparativo das listas de classificação *(rankings)* internacionais de e-gov e das variáveis consideradas;
b) há uma carência na literatura sobre métricas e indicadores relacionados aos planos e projetos para dar impulso às iniciativas do e-gov no interior do país; a verificação de uma maior concentração de pesquisas para medir os avanços nos níveis federal e estadual (em especial, quanto ao Poder Executivo) aponta para a importância de serem empreendidos estudos direcionados à esfera municipal e a outros poderes;
c) praticamente inexistem indicadores para verificar o avanço nas questões de legalidade e transparência em governo eletrônico; onde há lacunas na literatura, apresenta-se um potencial de desenvolvimento de estudos e pesquisas;
d) os resultados sugerem a oportunidade de se confirmar a verificação empírica de correlação entre o e-gov, o acesso e uso da tecnologia da informação pela sociedade e a produção de riqueza (medida pelo Produto Interno Bruto) nos diversos níveis do governo – o progresso do governo eletrônico parece caminhar paralelamente ao avanço do enriquecimento dos países; isso remete a possíveis pesquisas quantitativas que possam relacionar o sucesso do e-gov com a renda média da população dos municípios, grupo de municípios, estados, regiões e países; alguns países (em geral desenvolvidos) destacam-se em várias pesquisas, em posição vantajosa; embora os resultados das pesquisas selecionadas não possam ser comparados, observa-se que os primeiros colocados nas diversas pesquisas estão, também, situados em boas colocações nas demais listas classificatórias *(rankings)*, o que sugere que e-gov e tecnologia da informação estão diretamente relacionados a um melhor desempenho econômico-social das nações;
e) no mérito dos resultados do estudo comparativo entre as pesquisas internacionais de governo eletrônico, o principal achado foi a necessidade de mudança no foco de verificações de progresso para pesquisas mais qualitativas e de resultados, ao invés de se comparar diferentes realidades, o que remete ao aprimoramento da metodologia para essas abordagens; e
f) o estudo prospectivo empreendido sugere a oportunidade de aprofundamento no sentido da seleção de indicadores significativos e objetivos que permitam aprimorar os mecanismos de avaliação dos vários aspectos envolvidos na questão do e-gov (aspectos relacionados à qualidade e à quantidade dos serviços eletrônicos oferecidos, questões de usabilidade de portal, abertura do governo para a participação da sociedade nos planos e projetos governamentais, etc.) e do acesso e uso da tecnologia da informação pela sociedade.

A pesquisa propõe uma estrutura preliminar para verificação do estágio (ou maturidade) em governo eletrônico (RIECKEN, 2008a). Com relação a isso, as seguintes oportunidades de pesquisas se apresentam:

a) a estrutura proposta pode ser aprimorada para (por exemplo) se proceder à sua adequação a situações específicas, ou para reduzir a subjetividade (o que pode ser feito pelo desdobramento do roteiro em questões ainda mais objetivas quanto aos requisitos em cada eixo / estágio);
b) o método desenvolvido propicia que quaisquer órgãos ou entidades possam orientar seu progresso no e-gov por meio de um instrumento que deixe claro os requisitos do que se constituiria cada etapa evolutiva; nesse sentido, cabem estudos futuros para o aprimoramento da estrutura proposta, no escopo dos instrumentos de auto-avaliação de incentivo à melhoria da gestão pública e à qualidade na prestação dos serviços públicos, como o Programa Nacional de Gestão Pública e Desburocratização (GESPÚBLICA) do Ministério do Planejamento, Orçamento e Gestão (BRASIL, 2009;
c) a pesquisa de campo realizada para verificação do progresso apontou, ressalvadas as limitações, que o e-gov das prefeituras municipais estão, de um modo geral, aquém dos avanços verificados nas iniciativas estaduais, o que remete à necessidade de estudos futuros com o objetivo de reduzir essa distância, bem como compreender o porquê da desconexão entre o e-gov estadual e os equivalentes municipais.

Quanto à realidade nacional, as pesquisas de campo aplicadas aos governos estaduais e às prefeituras municipais permitiram a identificação de alguns pontos importantes e experiências que os respondentes poderiam passar relacionados ao e-gov. O resultado foi um levantamento representativo da realidade nacional nas diversas esferas do governo; desse modo apontando para um potencial de pesquisas em todos os temas do nexo temático[13] estruturado na tese, com o propósito de melhor compreender tais achados, a exemplo de:

a) estudos para aprofundar os motivos pelos quais foram verificadas rejeições ao desenvolvimento de e-gov, ou a falta de vontade política, analisando os impedimentos na busca por soluções de contorno;
b) investigar com maior profundidade os canais de interação com a população, suas deficiências, alcance e a participação da sociedade nesses importantes veículos de cidadania;
c) estudos relacionados à questão (da maior relevância) da inclusão digital da sociedade, em confronto com a constatação de que a própria inclusão digital da administração pública encontra-se aquém do esperado; a verificação do grau de inserção da tecnologia da informação na função pública apresenta-se como um importante campo de estudo, em especial nas áreas de educação, saúde, urgência local e riscos ambientais;
d) o perfil do profissional da informação e de tecnologia da informação para o e-gov;
e) o grave problema da escala da prestação de serviços eletrônicos e a necessidade de organização da retaguarda de serviços;
f) estudos para verificar em que grau a modernização das administrações para o uso do e-gov dinamiza as economias locais e impacta em processo eletrônico para o segmento privado;
g) investigar o motivo da pouca interação entre os órgãos dos diferentes poderes e para o e-gov, bem como de formas de parcerias;
h) estudos no sentido do desenvolvimento de uma visão mais sistêmica do e-gov e pesquisas sobre estruturas de governança;
i) pesquisas sobre o posicionamento dos responsáveis por expressivos processos informacionais na estrutura organizacional e sobre a influência deste posicionamento nos resultados da gestão;
j) estudos sobre a visão estratégica da informação e da tecnologia da informação, bem como sobre o grau de envolvimento dos que decidem no processo do e-gov; segurança da informação, padrões; políticas de privacidade;

[13] Como mostrado em detalhe na referida pesquisa (RIECKEN, 2008a).

k) estudos sobre o problema crítico da falta de integração dos sistemas de informação;
l) estudos e pesquisas para o desenvolvimento do planejamento estratégico e operacional para o e-gov de longo prazo, atrelado ao acesso a linhas de incentivo, minimizando desse modo os riscos das descontinuidades administrativas, comuns nas administrações públicas;
m) estudos e pesquisas relacionados à suficiência e adequação do marco regulatório, o impacto de sua aplicação, vulnerabilidades na regulamentação, riscos para a sociedade;
n) estudos sobre estratégias para incentivar o uso de portais;
o) pesquisas sobre os métodos e práticas envolvidos nos serviços típicos de ouvidoria e o potencial de geração de conhecimento desse importante canal de interação com a população; caberia investigar, também, os motivos pelos quais o "Fale Conosco" não funciona adequadamente em uma expressiva quantidade de órgãos públicos, bem como o grau de frustração do usuário que, de um modo geral, não é chamado a opinar sobre o e-gov e os canais de interação;
p) investigar as razões dos impedimentos culturais e da falta de confiabilidade das informações;
q) estudos de modelos referenciais de arquitetura de informação e de arquitetura de tecnologia da informação voltados para e-govs; a acumulação de um expressivo volume de dados e informações em meio eletrônico vem tornando necessário o estudo de modelos de organização desses acertos, em geral armazenados em grandes bases eletrônicas (RIECKEN, 2008b);
r) as profundas diferenças regionais do país remetem à verificação de que não há uma solução única para o modelo de governo eletrônico;
s) estudos sobre as diversas interações com o e-gov;
t) estudo sobre as formas de avaliação da qualidade, resultados e satisfação dos usuários, adequados ao e-gov;
u) estudos sobre os riscos do e-gov e da informação nas redes de telecomunicações (segurança da informação);
v) aprofundar a questão da lista de serviços típicos e do grau de sofisticação verificados e desejados para as administrações; e
w) estudos de caso, em especial voltados para a realidade nacional – como em Riecken e Lanza (2007) –, ou em países/regiões com problemas simulares com relação às expressivas diferenças culturais, sociais, geográficas, entre outras.

A tese não teve por escopo investigar apenas as administrações onde as iniciativas de governo eletrônico já se encontravam estabelecidas, mas também, verificar a viabilidade onde tais iniciativas eram incipientes ou inexistentes. Os resultados dessa investigação preliminar permitiram concluir pela oportunidade de serem aprofundados os importantes estudos sobre a questão da viabilidade do e-gov, em especial quanto aos impedimentos ou obstáculos, bem como para verificar se haveria um "porte" de município onde o e-gov seria mais ou menos recomendado,

Embora as ciências sociais apontem para a oportunidade de pesquisas qualitativas, as pesquisas quantitativas e o desenvolvimento de modelos estatísticos também podem se tornar importantes instrumentos de estudo de situações, onde o volume numérico de entes envolvidos e das variáveis, sinalizem a conveniência da sua utilização.

Nesse sentido, a tese desenvolveu um modelo de decisão (modelo estatístico) denominado "critérios de priorização", baseado em variáveis objetivas, o qual se apresenta como um referencial inicial para direcionar o incentivo progressivo do e-gov nas administrações municipais, em cada estado da federação.

O teste do modelo de decisão na situação da realidade brasileira, realizado na pesquisa, resultou em um *ranking* estadual (segundo o grau de relevância de cada município para fins de adesão a uma presumida rede intra-estadual de informações e serviços eletrônicos, consoante modelo proposto na tese) e sinaliza a oportunidade de refinamento do modelo, em especial pelo pressuposto de que o mesmo deve ser flexível e adaptável às regiões interessadas na sua utilização.

O modelo poderia ser útil, também, por exemplo, para celebração de eventuais acordos e convênios de cooperação técnico-financeira, onde os recursos são muitas vezes limitados e os dirigentes necessitam de um instrumento que apoie e direcione o bom uso do recurso público segundo critérios claros.

Assim, o elenco de dezenas de variáveis certamente pode ser reduzido, com o objetivo de se obter um conjunto mais reduzido que explique de modo similar o comportamento da "população" como um todo. Desse modo, as recomendações para estudos possíveis poderiam ser:

a) reduzir o modelo de decisão baseado em critérios objetivos a um conjunto menor de variáveis, por exemplo, pela verificação da correlação entre as variáveis;
b) estudos para implementação do modelo de decisão para escolha de prefeituras municipais que incentive o progresso do e-gov local com base em valores regionais nos aspectos cultura, turismo, etc.; e
c) estudos sobre a organização das informações e serviços eletrônicos em arquiteturas de informação voltadas para a e-gov; as questões relacionadas à operacionalização dessa organização formam, também, um vasto campo de estudo, mesmo porque "arquitetura" é um termo de significado amplo, o que remete à investigação sobre sua estrutura e seus agregados de componentes; elaborar uma arquitetura de sistema de informação significa desenvolver um modelo que se coadune com a arquitetura proposta, como: uma arquitetura em camadas (camada de dados, camada de componentes onde se localizam as regras de negócio, ou camada do aplicativo) ou arquitetura e padrões de interoperabilidade, de modo a se implementar estruturas preferencialmente flexíveis com alta coesão e baixo acoplamento entre seus componentes; assim, pode-se dizer que cabem estudos de arquitetura da informação, enquanto dados organizados (uma vez que a partir da modelização os dados se transformam em informação), ou enquanto estrutura do sistema.

As estatísticas oficiais, em especial as do IBGE e das Nações Unidas, compõem um material de excelente qualidade para o desdobramento de estudos e pesquisas por parte de pesquisadores interessados em temas envolvendo a tecnologia da informação e na apropriação destas (e principalmente das informações e conhecimentos envolvidos) por governos e pela sociedade.

Verifica-se, também, a oportunidade de se aprofundar alguns achados nessas pesquisas oficiais para melhor compreender os resultados das mesmas, relacionados ao tema dos serviços de portal e da tecnologia da informação. As principais pesquisas do Instituto Brasileiro de Geografia e Estatística utilizadas na tese (IBGE, 2001) (IBGE, 2004) (IBGE, 2006) apontam para uma quantidade expressiva de sítios de prefeituras municipais com presença na Internet, enquanto outros indicativos, por exemplo, no portal Brasil.gov, sinalizam uma quantidade aquém da apurada pelo IBGE. Desse modo, cabem estudos mais aprofundados para avaliar a sistemática de manutenção dos portais nacionais e o modo como esses mecanismos estão sendo estabelecidos.

Poderiam, também, ser desenvolvidos estudos com base nos *websites*, complementados por pesquisas mais aprofundadas em alguns casos selecionados. O portal de maior destaque em um município é aquele no qual são apresentadas, em especial, as informações sobre a administração central e os serviços on-line disponíveis. Muitos municípios possuem portais bastante desenvolvidos quanto à disponibilização de serviços on-line, entretanto governo eletrônico á mais do que simplesmente estar presente na Internet com um site ou portal. O propósito principal do governo eletrônico deve ser o de utilizar a tecnologia da informação para prover serviços governamentais eficientes (bom uso dos recursos), eficazes (cumprindo metas mensuráveis cada vez melhores) e efetivos (que correspondam às necessidades e desejos da população). Estudos mais aprofundados possibilitariam uma melhor compreensão dos diferentes indicativos de presença na Internet das prefeituras municipais com um sítio ou portal, bem como o aprofundamento de outros aspectos (meios de atualização dos endereços eletrônicos, garantindo a integridade dos mesmos; nível de conclusibilidade das denúncias, sugestões e solicitações diversas encaminhadas pelos cidadãos aos portais, entre outros).

A rede intra-estadual de informações e serviços eletrônicos proposta apresenta-se como um vasto campo para futuras pesquisas, a exemplo dos seguintes temas:

a) os tipos e perfis de redes informacionais;
b) a necessária adaptação e preparação para a mudança;
c) a verificação de temas de interesse comum, prioridades e fontes de financiamento;
d) o estabelecimento e acompanhamento dos objetivos da rede;
e) as formas de organização, transmissão, mineração de dados e textos (CAPUANO, 2009) e recuperação dos dados e informações em geral;
f) a gestão do conhecimento (*knowledge management* – KM), apontada pela OCDE como cerne das tarefas governamentais e inseparáveis da estratégia, planejamento, consulta prévia e implementação (NAÇÕES UNIDAS, 2008);
g) as questões motivacionais para o envolvimento e empenho de cada participante;
h) os mecanismos de estímulo à participação social na gestão pública;
i) a exigência da transparência na democracia moderna (para que de fato represente a vontade popular);
j) os mecanismos mais diretos através dos quais o cidadão possa, eventualmente, expressar seu descontentamento; e
k) a municipalização do país pela via virtual.

Finalmente, o tema e-gov, bem como a questão do acesso e uso da tecnologia da informação pela sociedade, é relativamente recente, mas inexoravelmente difundido em todas as atividades humanas. Revela-se, desse modo, como um vasto potencial de pesquisas e oportunidades para o desenvolvimento de métodos e técnicas.

Nesse sentido, as áreas acadêmicas, entre elas a Ciência da Informação, tornam-se espaços oportunos para o empreendimento de pesquisas no tema, devido à transversalidade natural dessa área de conhecimento, por meio da inserção de linhas de pesquisas sobre tão importante temática.

Os modelos propostos na tese podem e devem ser aprimorados com base nas distinções locais, considerando o peso das tradições e da história locais, sendo apenas um referencial inicial. A derivação para novos modelos sai da esfera das ciências exatas para entrar no foco das ciências humanas, com toda a subjetividade que lhe é inerente.

5. REFERÊNCIAS DO CAPÍTULO 13

AGNER, Luiz. *Arquitetura de informação e governo eletrônico: diálogo cidadãos-Estado na World Wide Web: Estudo de caso e avaliação ergonômica de usabilidade de interfaces humano-computador*. 2008. 354 f. Tese (Doutorado em Design) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, 2007.

ARAKAKI, Cristiane. *O governo eletrônico como instrumento de aproximação do governo e o cidadão.* 2008. 164 f. Dissertação (Mestrado em Comunicação) - Universidade de Brasília, Brasília, DF, 2008.

BOBBIO, Norberto. *O futuro da democracia*. 9. ed. Rio de Janeiro: Paz e Terra, 2004. 207 p.

BRAGA, Lamartine Vieira Braga. *Contribuições da certificação digital ao desenvolvimento do governo eletrônico e aperfeiçoamento de políticas públicas e serviços públicos no Brasil*. 2008. 99 f. Dissertação (Mestrado em Gestão do Conhecimento e da Tecnologia da Informação) – Universidade Católica de Brasília, Brasília, DF, 2008.

BRASIL, Ministério do Planejamento, Orçamento e Gestão. Gestão Pública**.** *GESPÚBLICA* (portal). Brasília, DF, 2009. Disponível em: https://www.gespublica.gov.br/. Acesso em: 19 jun. 2009.

CAP GEMINI. *Online availability of public services: How is europe progressing?* [S.l.], 2006. Disponível em: http://ec.europa.eu/information_society/eeurope/i2010/docs/benchmarking/online_availability_2006.pdf >. Acesso em: 19 jun. 2009.

CAPUANO, Ethel Airton. Redesenho de processos nas organizações da administração pública. In: KNIGHT, Peter Titcomb; FERNANDES, Ciro Campos Christo; CUNHA, Maria Alexandra (Org.). *E-desenvolvimento no Brasil e no mundo: Subsídios e Programa e-Brasil*. São Caetano do Sul: Yendis, 2007. p. 585-625

CAPUANO, Ethel Airton. Construtos para modelagem de organizações fundamentadas na informação e no conhecimento no serviço público brasileiro. *Ciência da Informação,* Brasília DF, v. 37, n. 3, p. 18-37, set./dez. 2008.

CAPUANO, Ethel Airton. O poder cognitivo das redes neurais artificiais modelo ART1 na recuperação da informação. *Ciência da Informação,* Brasília, DF, v. 38, n. 1, p. 9-30, jan./abr. 2009.

CASTELLS, Manuel. A sociedade em rede. In: CASTELLS, Manuel. *A era da informação: Economia, sociedade e cultura*. São Paulo: Paz e Terra, 2001. v. 1, 617 p.

CHAHIN, Ali; CUNHA, Maria Alexandra; PINTO, Solon Lemos. *E-gov.br: A próxima revolução brasileira*. São Paulo: Prentice Hall, 2004. 400 p.

DIAS, Cláudia Augusto. *Método de avaliação de programas de governo eletrônico sob a ótica do cidadão-cliente: uma aplicação no contexto brasileiro*. 2006. xvi, 277 f. Tese (Doutorado) – Universidade de Brasília, Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Departamento de Ciência da Informação e Documentação, Brasília DF, 2006.

DWIVEDI, Yogesh K.; WILLIAMS, Michael D. *Demographic influence on UK citizens' e-government adoption.* (Survey). Electronic Government 5.3 (June 21, 2008): 261. Academic OneFile. Gale. Disponível em: < http://find.galegroup.com/itx/start.do?prodId=AONE >. Acesso (via portal de periódicos da CAPES), em 17 jun. 2009.

COSTA, Eduardo. *ESR/RNP promove palestra sobre governo eletrônico*. Palestra proferida na Escola Superior de Redes em 10 jul. 2007. Rio de Janeiro: ESR/RNP, 2007. Disponível em: < http://www.esr.rnp.br/noticias/?noticia=71 >. Acesso em: 19 jun. 2009.

ESTADOS UNIDOS. **USA.gov** (portal do governo eletrônico dos USA) Washington, 2009a. Disponível em: < http://www.usa.gov >. Acesso em: 19 jun. 2009.

ESTADOS UNIDOS. The White House. *Office of e-government & information technology (portal)*. Washington DC, 2009b. Disponível em: http://www.whitehouse.gov/omb/e-gov/. Acesso em: 19 jun. 2009.

ESTADOS UNIDOS. *Government Accountability Office* **-** *GAO* (portal), 2009c. Disponível em: < http://www.gao.gov >. Acesso em: 19 jun. 2009.

FANG, Zhiyuan. E-government in digital era: concept, practice and development. *International Journal of The Computer, The Internet and Management*, Thailand, v. 10, n. 2, p. 1-22, 2002.

FERNANDES, Andréa Gomes. E-governo: o que já fazem estados e municípios. *Informe-se*, [Rio de Janeiro], n. 20, out. 2000. Disponível em: < http://www.bndes.gov.br/conhecimento/informeSF/inf_20.pdf >. Acesso em: 19 mar. 2008.

FERRER, Florência; SANTOS, Paula. E*-government: O governo eletrônico no Brasil*. São Paulo: Saraiva, 2004. 258 p.

FREITAS, Alexandre Vítor Figueira de. *A dimensão comunicativa na ciência da informação: Um suporte teórico à compreensão da construção de sistemas.* 2006. xiv, 144 f. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília Brasília, DF, 2006.

GONÇALVES, Sonia Maria Goulart. *Elementos básicos para a formulação de uma política de gestão do conhecimento para a administração pública federal brasileira*. 2006. xii, 140 f. Dissertação (Mestrado em Gestão do Conhecimento e Tecnologia da Informação) – Universidade Católica de Brasília, Brasília, DF, 2006.

HOLZER, Marc; KIM, Seang-Tae. *Digital governance in municipalities worldwide (2007): A longitudinal assessment of municipal websites throughout the world.* New York: United Nations: The American Society for Public Administration, 2008. Disponível em: http://unpan1.un.org/intradoc/groups/public/documents/aspa/unpan031374.pdf#xml=http://unpan1.un.org/intradoc-cgi/idc_cgi_isapi.dll?IdcService=GET_XML_HIGHLIGHT_INFO&QueryText=dDocType+%3csubstring%3e+%60ASPA%60++%3cAND%3e++%28Digital+Governance%0d%0a%29&SortField=dInDate&SortOrder=Desc&dDocName=UNPAN031374&HighlightType=PdfHighlight >. Acesso em: 4 set. 2009.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Pesquisa de Informações Públicas Municipais. *Perfil dos municípios brasileiros: 2001*. Rio de Janeiro, 2001. 1 CD-ROM.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Pesquisa de Informações Públicas Municipais. *Perfil dos municípios brasileiros: 2004.* Rio de Janeiro, 2004. 1 CD-ROM.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Pesquisa de Informações Públicas Municipais. *Perfil dos municípios brasileiros: 2006.* Rio de Janeiro, 2006. 1 CD-ROM.

JAMBEIRO, Othon. *et al.* Cidades, cidadania e tecnologias avançadas de comunicações. In: II CONGRESSO DA ASSOCIAÇÃO BRASILEIRA DOS PESQUISADORES EM COMUNICAÇÃO E POLÍTICA, 2007, Belo Horizonte. *Anais do II Compolítica*, 2007. Disponível em: http://www.fafich.ufmg.br/compolitica/anais2007/gt_csc-othon. Acesso em: 8 set. 2009.

KAPLAN, Abraham. *A conduta na pesquisa: Metodologia para as ciências do comportamento*. São Paulo: Herder, EDUSP, 1969.

KNIGHT, Peter T.; FERNANDES, Ciro Campos Christo; CUNHA, Maria Alexandra (Org.). *E-Brasil: Um programa para acelerar o desenvolvimento socioeconômico aproveitando a convergência digital*. São Caetano do Sul: Yendis, 2007a. 96 p.

KNIGHT, Peter T.; FERNANDES, Ciro Campos Christo; CUNHA, Maria Alexandra. O projeto e-Brasil e o programa e-Brasil. In: KNIGHT, Peter T.; FERNANDES, Ciro Campos Christo; CUNHA, Maria Alexandra. (Org.). *E-Brasil: Um programa para acelerar o desenvolvimento socioeconômico aproveitando a convergência digital*. São Caetano do Sul: Yendis, 2007b. Disponível em: < http://www.tedbr.com/projetos/e-Brasil/e-brasil_pub/e-desenvolvimento/pdfs/cap_01.pdf >. Acesso em: 19 jun. 2009.

LAKATOS, Eva Maria; MARCONI, Maria de Andrade. *Metodologia científica*. 2. ed. São Paulo: Atlas, 1992. 248 p.

LIMA, João Alberto de Oliveira. *Modelo Genérico de Relacionamentos na Organização da Informação Legislativa e Jurídica*. 2008. 289 f. Tese (Doutorado em Ciência da Informação e Documentação) –– Universidade de Brasília, 2008.

MEDEIROS, Paulo Henrique Ramos. *Governo eletrônico no Brasil: Aspectos institucionais e reflexos na governança*. 2004. 318 f. Dissertação (Mestrado em Administração) – Universidade de Brasília, Brasília, DF, 2004.

MORIN, Edgar. Da necessidade de um pensamento complexo. In: MARTINS, Francisco Menezes; SILVA, Juremir Machado da (Org.). *Para navegar no século XXI*. 3. ed. Porto Alegre: Sulina/Edipucrs, 2003. p. 13-36

NAÇÕES UNIDAS (UNITED NATIONS). Department of Economic and Social Affairs. Division for Public Administration and Development Management. *UN global e-government readiness report 2004*. New York, 2004. Disponível em: < http://www.unpan.org/egovernment4.asp >. Acesso em: 19 jun. 2009.

NAÇÕES UNIDAS (UNITED NATIONS). Department of Economic and Social Affairs. Division for Public Administration and Development Management. *UN e-government survey 2008: from e-government to connected governance.* New York, 2008. Disponível em: http://unpan1.un.org/intradoc/groups/public/documents/un/unpan028607.pdf. Acesso em: 19 jun. 2009.

OCDE – ORGANIZAÇÃO PARA A COOPERAÇÃO E O DESENVOLVIMENTO ECONÔMICO (OECD - ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT). *Information and Communication Technologies (portal)*. E-government. Disponível em: http://www.oecd.org/topic/0,3373,en_2649_34129_1_1_1_1_37441,00.html . Acesso em: 19 jun. 2009.

OLIVEIRA, João Batista Ferri. *Governo eletrônico: Uma proposta para o provimento de recursos da tecnologia da informação e comunicação em prefeituras*. 2006. 285 f. Dissertação (Mestrado em Administração) – Universidade do Vale do Rio dos Sinos/UNISINOS, São Leopoldo, 2006.

PARDO, T. *Realizing the promise of digital government: It's more than building a web site*. Albany, NY: Center for Technology in Government, 2000.

PRIBERAM INFORMÁTICA. **[Dicionário da] língua portuguesa on-line.** Lisboa, 2010. Disponível em: < http://www.priberam.pt/dlpo/ >. Acesso em: 13 junho 2010.

RIECKEN, Rinalda; LANZA, Beatriz Barreto Brasileiro Lanza. E-Paraná: a rede de informações e serviços eletrônicos do governo do estado do Paraná. *Informação & Informação*, Londrina PR, v.12, n.2, jul./dez. 2007.

RIECKEN, Rinalda. Frame de temas potenciais de pesquisa em Ciências da Informação. *Revista Digital de Biblioteconomia e Ciência da Informação*, Brasília, DF, v. 3, n. 2 (2006), p.43-63.

RIECKEN, Rinalda. *Governo eletrônico em administrações locais brasileiras: Avaliação de progresso, fatores intervenientes e critérios de priorização de iniciativas.* 2008. 2 v. Tese (doutorado) - Universidade de Brasília, Faculdade de Economia, Contabilidade e Ciência da Informação e Documentação, Departamento de Ciência da Informação e Documentação, 2008a. 1074 p. Disponível em: https://consulta.bce.unb.br/pergamum/biblioteca/index.php?resolution2=1024_1. Acesso em: 19 jun. 2009.

RIECKEN, Rinalda. Estudo comparativo entre diferentes modelos de organização de acervos de dados eletrônicos. *Perspectivas em Ciência da Informação*, Belo Horizonte, MG, v.13, n.2, p. 43-61, maio/ago. 2008b.

ROBREDO, Jaime. *Da ciência da informação revisitada aos sistemas humanos de informação*. Brasília, DF: Thesaurus: SSRR Informações, 2003. 245 p.

ROBREDO, Jaime. *Documentação de hoje e de amanhã: Uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas.* 4. ed. rev. e ampl. Brasília, DF: Edição de autor, 2005. 409 p.

RODRIGUES, Lorena. *Entenda o setor de telecomunicações no Brasil***.** Artigo publicado na Folha Online, em Brasília, em 25 abr. 2008. Disponível em: <http://www1.folha.uol.com.br/folha/dinheiro/ult91u395749.shtml>. Acesso em: 19 jun. 2009

ROSA, Maria. *Ruptura construtiva com o passado*. Rio de Janeiro: RNP, Artigo publicado no Jornal Estado de Minas e transcrito na Rede Nacional de Ensino e Pesquisa, em 31 mar. 2005. Disponível em: < http://www.rnp.br/noticias/imprensa/2005/not-imp-050331.html >. Acesso em: 19 jun. 2009.

SÃO PAULO (Estado). Governo do Estado de São Paulo. *Comitê de Qualidade da Gestão Pública* (portal). São Paulo, 2009. Disponível em: http://www.gestaopublica.sp.gov.br/gati/gc_estudos.html. Acesso em: 19 jun. 2009a.

SÃO PAULO (Estado). Governo do Estado de São Paulo. *Secretaria de Fazenda. Programa Nota Fiscal Paulista* (portal). São Paulo, 2009. Disponível em: < http://www.nfp.fazenda.sp.gov.br/ >. Acesso em: 06 set. 2009b.

SAATY, Thomas L. *Método de análise hierárquica*. São Paulo: McGraw-Hill, 1991. 367 p.

SIMÃO, João Batista Simão. *Universalização de serviços públicos na Internet para o exercício da cidadania: Análise crítica das ações do Governo Federal*. 2004. xvii, 174 f. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, Brasília, DF, 2004.

SUAIDEN, Emir José. *O Brasil e o acesso às novas tecnologias*. Texto divulgado no blog do autor, em 25 jul. 2006. Disponível em: http://blogdoemir.blogspot.com/2006/07/o-brasil-e-o-acesso-s-novas.html. Acesso em: 19 jun. 2009.

TAKAHASHI, Tadao (Org.). *Sociedade da informação no Brasil: Livro verde*. Brasília, DF: MCT, 2000. Disponível em: http://www.mct.gov.br/index.php/content/view/18878.html. Acesso em: 19 jun. 2009.

YIN, Robert K. *Case study research: Design and methods*. London: Sage, 1988. 171 p. (Applied social research methods series, v. 5).

VI. OUTROS TEMAS E ATIVIDADES

Capítulo 14 – Competência informacional em tempos de Web – Fábio Augusto Guimarães TEIXEIRA e Greyciane Souza LINS, *p. 321-335*

Capítulo 14

Competência informacional em tempos de Web

Fábio Augusto Guimarães Teixeira [1] e Greyciane Souza Lins [2]

SUMÁRIO DO CAPÍTULO 14

Sumário do Capítulo 14, *p.322*
Como citar o Capítulo 14, *p.322*

1. INTRODUÇÃO DO CAPÍTULO 14, *p. 323*
2. TECNOLOGIAS DE INFORMAÇÃO E COMUNICAÇÃO, *p. 324*
  2.1. Hardware, *p. 325*
  2.2. Software, *p. 325*
  2.3. Comunicação de dados, *p. 327*
3. COMPETÊNCIA INFORMACIONAL, *p. 328*
4. COMPETÊNCIA INFORMACIONAL EM TEMPOS DE WEB, *p. 332*
5. REFERÊNCIAS DO CAPÍTULO 14, *p. 333*

[1] Mestre em Ciência da Informação. Programa de Pós-Graduação em Ciência da Informação (PPGCInf), Faculdade de Ciência da Informação da, Universidade de Brasília (UnB/FCI). E-mail: fabioteixeira@gmail.com.
[2] Mestre em Ciência da Informação. Programa de Pós-Graduação em Ciência da Informação (PPGCInf), Faculdade de Ciência da Informação da, Universidade de Brasília (UnB/FCI). CV Lattes: http://lattes.cnpq.br/3752454481868302. E-mail: greycilins@gmail.com.

Resumo do Capítulo 14

Os modelos disponíveis de competência informacional precisam ser periodicamente revistos para acompanhamento das evoluções tecnológicas que impulsionam o crescente uso de seus instrumentos como fonte de informação. Este capítulo apresenta elementos das tecnologias de informação, *hardware*, *software* e comunicação de dados como estrutura de formação da Web, bem como as habilidades requeridas para utilização eficiente dos recursos tecnológicos. Para isso, revisa conceitos de aplicações de modelos de competência informacional para usuários de perfis diferentes, propondo um modelo de aplicação de habilidades requeridas para usuários competentes informacionais em tempos de Web.

Palavras-chave: Competência Informacional, Tecnologia da Informação e Comunicação

Abstract

Informacion literacy in Web times

The information literacy models available must be periodically revised to garnish the technological evolution that drives the raising use of its instruments as information source. This chapter presents information technologies elements, hardware, software, and data communication as the formation structure of the Web, as well as the required skills for efficient application of technological resources. To achieve this, revises concepts of information literacy application models for users with different profiles, proposing an application model of required skills for users to acquire information literacy in Web times.

Keywords: Information literacy, Information and Communication Technology

Como citar o Capítulo 14

TEIXEIRA, Fábio A. G.; LINS, Greyciane Souza. Competência informacional em tempos de Web*. In* : Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios pelo bosque da informação: Estudos sobre representação e organização da informação e do conhecimento – EROIC*. Brasília DF: IBICT, 2010, 335 p. Capítulo 14, p. 321-334. Edição eletrônica. Disponível em: http://www.ibict.br/publicacoes/eroic.pdf. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

1. INTRODUÇÃO DO CAPÍTULO 14

Na apresentação do Livro Verde do Programa Sociedade da Informação no Brasil, o então ministro brasileiro da Ciência e Tecnologia, Ronaldo Mota Sardenberg, escreveu da seguinte maneira sobre a importância do conhecimento, da educação e do desenvolvimento científico e tecnológico:

> O conhecimento tornou-se, hoje mais do que no passado, um dos principais fatores de superação de desigualdades, de agregação de valor, criação de emprego qualificado e de propagação do bem-estar. A nova situação tem reflexos no sistema econômico e político. A soberania e a autonomia dos países passam mundialmente por uma nova leitura, e sua manutenção - que é essencial - depende nitidamente do conhecimento, da educação e do desenvolvimento científico e tecnológico. (PROGRAMA SOCIEDADE DA INFORMAÇÃO, 2000, p. v)

Situando o conhecimento como recurso necessário ao desenvolvimento social e econômico da sociedade, o ministro o relacionou explicitamente com as necessidades educacionais e profissionais para o uso da informação. A análise dessas necessidades passa pela compreensão do contexto em que o indivíduo, potencial usuário da informação, está inserido. Para que objetivos precisa da informação? Como ele define suas necessidades de informação? Quais são as formas de acesso à informação que ele conhece ou dispõe? De que forma associa a informação a que tem acesso com os seus objetivos iniciais? As respostas a todas essas e ainda outras possíveis perguntas são vinculadas a características do comportamento de cada usuário ou às suas competências informacionais. Além disso, o contexto mencionado se torna cada vez mais influenciado pelo ferramental tecnológico criado para o compartilhamento da informação e do conhecimento, cujo maior exemplo é o advento da Web.

Miranda (2004, p. 118) definiu competência informacional de um profissional da informação como a "expertise em lidar com o ciclo informacional, com as tecnologias da informação e com os contextos informacionais". Por ciclo informacional, a autora entende as etapas de trabalho com a informação, como levantamento de necessidades, coleta, tratamento, emprego e distribuição. Como tecnologias da informação, definiu as tecnologias que influenciaram a arquitetura do conhecimento, citando computadores, estruturas de telecomunicação e sistemas de *software* como exemplos. E, por fim, o contexto organizacional foi definido como o meio em que ocorre o ciclo informacional e que, portanto, influencia o seu fluxo e as suas características.

O desenvolvimento de tecnologias de informação e comunicação (TICs) tem uma relação intrínseca com a necessidade de conhecimento de uma sociedade e, por consequência, com as competências informacionais que cada indivíduo apresenta na busca ou no trato da informação a que tem acesso. Segundo Tarapanoff *et al* (2002): É necessário que o profissional da informação atue como uma mediador [sic] entre o mundo digital e a capacidade real de entendimento do receptor da informação, garantindo a efetiva comunicação e a satisfação da necessidade informacional do usuário dessa tecnologia.

O papel do profissional da informação como mediador entre a informação disponível na sociedade e aqueles que dela necessitam para suprir suas necessidades informacionais, normalmente vinculadas a um contexto e objetivo a ser cumprido, tem como uma das principais características abranger a utilização e o fomento do desenvolvimento das TICs como ferramenta neste processo comunicativo, mas não somente isso. Primeiramente como usuário, o profissional da informação deve conhecer o ferramental tecnológico disponível para acesso à informação, até mesmo para saciar as suas próprias necessidades e também poder criar empatia com o público-alvo de seus esforços. Após esta etapa, o profissional poderá, como conhecedor das características dos indivíduos a quem atende e/ou promotor de oportunidades para melhorias nas técnicas disponíveis, sugerir e trabalhar em prol de novos desenvolvimentos dessas tecnologias.

O objetivo principal deste capítulo é abordar o papel das competências informacionais humanas na sociedade apoiada no preceito de compartilhamento da informação ocasionada pelas TICs e, em especial, pela Web. Para tanto, deverão ser compreendidos: 1) os tipos de ferramentas

que se caracterizam como TICs e que influenciam o nosso cotidiano informacional; 2) o contexto em que surgiu a expressão competência informacional, bem como as características definidas pela história e seu relacionamento com a tecnologia; e, por fim, 3) a importância das competências informacionais como catalisadores do processo, baseado nas TICs, de transformação da informação para o desenvolvimento da sociedade.

2. TECNOLOGIAS DE INFORMAÇÃO E COMUNICAÇÃO

Utilizando uma definição do próprio Livro Verde, TICs são "Tecnologias utilizadas para tratamento, organização e disseminação de informações." (PROGRAMA SOCIEDADE DA INFORMAÇÃO NO BRASIL, 2000, p. 176). Sobre esta descrição, cabe um enfoque um pouco menos abstrato e que se aproxime de nossa percepção cotidiana.

Primeiramente, nos tempos atuais, não há como mencionar o termo TIC sem associá-lo ao formato digital da informação, o emaranhado de zeros e uns que são compreensíveis apenas pelos computadores e por alguns usuários mais fanáticos. Isso sem menosprezar a transmissão analógica existente em sinais televisivos ou de rádio, ainda extremamente importante na comunicação, mas em franco processo de substituição por transmissões digitais.

Apesar de pensarmos que ainda são inalcançáveis para grande parte da sociedade, os computadores se espalharam de forma a não mais ser possível conceber a nossa existência sem a presença deles. Mesmo quando não percebemos. Na origem da palavra, computador é qualquer ferramenta que auxilie o ser humano a realizar cálculos. De fato, o primeiro computador de que se tem notícia é o ábaco. Atualmente aparatos tão distintos quanto um carro, um celular, um sistema de som e, até mesmo, eletrodomésticos são portadores de computadores de tamanho por vezes diminuto. Isso sem contar os próprios *desktops* e *notebooks*, normalmente os primeiros aparatos que nos vêm à cabeça quando falamos de computadores e que invadiram as empresas e domicílios.

Porém, estas mesmas máquinas não fazem as maravilhas que delas se espera sem uma série de comandos ordenados (e que sejam por elas compreensíveis em uma determinada linguagem de programação). Diferentes das máquinas físicas, denominadas de *hardware*, estes conjuntos de comandos, encapsulados em programas de computador, são denomidados de *software*.

Interligando toda essa disposição tecnológica, torna-se necessária também uma infraestrutura de comunicação que suporte a transmissão dos dados em formato digital. Assim, surgiram dispositivos como barramentos (aqueles filamentos em cobre ou outro metal que existem em placas, transmitindo sinais digitais de um lado para outro em celulares, computadores, veículos, relógios de pulso, etc.), e finalmente redes de comunicação de dados como redes corporativas (Ethernet, ATM, DSL, etc.), a cabo, sem fio (Wireless), de comunicação celular (GSM, CDMA) e, principalmente, a Internet.

Em uma era em que a presença destas TICs se torna tão presente no cotidiano, acaba sendo natural verificar a procura por aperfeiçoamento profissional de pessoas interessadas nos campos de estudo atingidos por elas. Parte daí o desenvolvimento de diversos estudos nas áreas de informação, notadamente computação, ciência da informação, ou mesmo os já tão ofertados cursos de informática. Indo mais além, novas exigências profissionais tornaram-se genericamente consolidadas nas vagas ofertadas no mercado de trabalho como o "conhecimento de informática".

Entretanto, propondo que o papel principal destas tecnologias deve ser o de intermediar o conhecimento com os que dele desejam fazer uso, uma questão chave deve ser proposta: qual a real importância do conhecimento sobre as TICs como competência informacional?

Desde já se faz necessário esclarecer que não há resposta correta para esse questionamento em todas as situações. A onipresença das TICs faz com que este tipo de reflexão deva ser realizado caso a caso, no contexto em que está inserido. Inclusive a própria concepção sobre o papel destas tecnologias, o que foi abordado neste texto apenas para facilitar o exercício, pode e deve ser questionado. Com isso em mente, o primeiro passo é o de

aprender o que são (ou podem ser) essas tecnologias. Abordamos isoladamente cada tipo de TIC, classificadas em *hardware*, *software* e redes de comunicação.

2.1. Hardware

O *hardware* é o tipo de TIC formado pelos componentes eletrônicos materiais. Como componentes eletrônicos, eles são caracterizados pela transmissão de informação através de impulso elétrico. A ausência de corrente é simbolizado pelo sinal 0. O sinal 1 significa a presença de corrente. Decorre daí o uso destes dois sinais (chamados de bits) para representar qualquer informação digital.

Um exemplo básico de *hardware* são os computadores (*desktops,* ou computadores de mesa, e *notebooks* ou *laptops,* ou computadores portáteis). Cada computador é composto por vários outros componentes (que também são tipos de *hardware*) como processadores, que são o cérebro do computador, diferenciados entre si pela quantidade de cálculos por unidade de tempo que são capazes de realizar; placas que processam sinais de vídeo (e depois os enviam para dispositivos como os monitores), sinais de áudio (que os enviam para fones de ouvidos, alto falantes, etc.); leitores de mídias de armazenamento como discos rígidos, CDs e DVDs; além dos próprios dispositivos de entrada (como teclados, *mouses*, *scanners*, câmeras, canetas óticas, leitores de impressão digital, sensores de telas *touchscreen* - sensíveis ao toque) e de saída (como monitores de vídeo e impressoras).

O universo de componentes de *hardware* dispostos em um computador, como mencionado no parágrafo anterior, já apresenta uma grande variedade de aplicações, possibilidades e combinações. Além disso, verifica-se uma tendência de difusão destas tecnologias em outros dispositivos presentes em nosso cotidiano, numa escala tão grande quanto imperceptível.

Como exemplo, os carros hoje em dia são controlados por vários computadores. Pode-se mencionar o módulo de controle do motor, da injeção eletrônica, os dispositivos de segurança, a central eletrônica e até os sistemas mais sofisticados de frenagem (como freios antitravamento – ABS) e de proteção contra colisão (bolsas de proteção – *airbags*). Diz-se que para ir de carro até a esquina, mais informações são processadas do que a Apollo 11 para chegar à Lua. (MOURA, 2005)

Um telefone celular é o melhor exemplo desta presença maciça e em miniatura dos componentes, pois se assemelha bastante aos computadores convencionais, só que em versão de bolso. Estes pequenos aparelhos possuem dispositivos de entrada (o microfone, as teclas, a câmera, a tela *touchscreen*, etc.), processador e dispositivos de saída (o alto-falante e a tela). E essa invasão não conhece limites. Aparelhos eletrodomésticos como geladeiras, fogões, lavadoras de roupa e de louça já começam a ser dotadas de pequenos computadores com o intuito de oferecer maiores facilidades a seus donos. Um exemplo à parte são as televisões e monitores.

O advento de tecnologias de LCD (*Liquid Crystal Display* – Monitor de Cristal Líquido), plasma e LED (*Light Emitting Diode* – Diodo Emissor de Luz) permite não apenas a melhoria da qualidade das imagens, como também acompanha novas funcionalidades agregadas aos monitores, tal como a recepção de sinais digitais e até mesmo a gravação de imagens.

Porém, todos os avanços em tecnologias de componentes eletrônicos não serviriam de nada se esses não fossem comandados por programas de computador, ou *softwares*. Eles são abordados logo a seguir.

2.2. *Software*

Os programas de computador têm como função básica usufruir das capacidades de cálculo dos componentes de *hardware*. Os *softwares* são compostos por conjuntos de seqüências ordenadas de instruções, ou comandos, que manipulam dados com um determinado objetivo. Por exemplo, um programa editor de texto simples tem como objetivo receber texto inserido por alguém e apresentá-lo. Para isso, possui instruções que fazem com que o

computador aguarde a digitação de teclas em um teclado e também as envie para apresentação em um monitor ou uma impressora. Outros comandos podem ser utilizados para salvar um arquivo em algum lugar, aumentar o tamanho do texto apresentado no monitor, etc. De qualquer forma, estes comandos são sempre conjuntos ordenados de instruções.

Um tipo tão básico quanto importante de *software* é composto pelos sistemas operacionais. Um sistema operacional pode ser descrito como o programa que controla todos os componentes de *hardware* em uma máquina. Em computadores, os sistemas operacionais mais famosos são o Windows (o mais utilizando em computadores pessoais), o Linux, o MacOS (para computadores da Apple) e o Unix (mais voltado para computadores de grande porte).

Independente do dispositivo, sempre existe um sistema operacional por trás. Os celulares também possuem diferentes tipos de sistemas operacionais, como a versão MacOS X para IPhones, o Symbian (amplamente utilizado por celulares da empresa Nokia), o Windows Mobile, o Android (lançado pela empresa Google) e também versões mais simples do Linux. Dispositivos ainda menores, como os presentes em eletrodomésticos ou carros, também possuem sistemas operacionais normalmente desenvolvidos para cada aparelho específico.

Uma vez dotados de sistemas operacionais que permitem controlar o poder de processamento do *hardware*, os computadores podem ser utilizados para desempenhar tarefas específicas. As mais conhecidas são as executadas por *softwares* de produtividade de escritórios como processadores de texto, planilhas eletrônicas, apresentação eletrônica, navegadores de Internet e leitores de mensagens eletrônicas, entre outros. Os mais famosos pacotes de produtividade são o Microsoft Office, composto por Word, Excel, PowerPoint, entre outros; e o OpenOffice.

Cabe destacar que são esses, juntamente com os sistemas operacionais, os *softwares* cujo conhecimento de uso é normalmente exigido por empregadores, de certa forma independentemente da ocupação. O "conhecimento em informática" tem se tornado cada vez mais uma exigência profissional tal como o correto emprego da língua portuguesa. Pode-se dizer que isso é uma das bases para a preocupação cada vez maior acerca do conhecimento sobre TICs como competência informacional.

Além dos sistemas operacionais e das ferramentas de produtividade de escritório, existem vários tipos de *software* com fins específicos: gerenciadores de bancos de dados, programas de monitoramento, ferramentas de auxílio a *design* (CAD – *Computer Aided Design*), programas de planejamento estratégico, de apoio a decisão, simuladores, etc.

A indústria de *software*, assim denominados os fabricantes de programas de computador, encontra-se atualmente balançada entre dois modelos de construção e de distribuição. O modelo mais antigo é o proprietário. Neste caso, as empresas são responsáveis pelo investimento no desenvolvimento do *software* e cobram um valor pela sua venda, licença de uso ou prestação de serviços a seus usuários, mas sem deixar que eles visualizem a forma (o código) como estes programas operam e sem permitir que eles os alterem. É dessa forma que empresas como a Microsoft (proprietária dos sistemas operacionais Windows e do pacote Office) obtêm o seu faturamento.

O modelo mais recente e que vem apresentando uma revolução na forma de desenvolver, promover e utilizar *software* é o modelo de código aberto (também chamado de *open source* ou código livre). Este modelo tem como filosofia a não obrigatoriedade do retorno financeiro direto. Neste caso, o detentor de um programa permite que outras pessoas tenham acesso ao código (as instruções) do *software*, estimulando que elas o melhorem (seja corrigindo algumas falhas ou mesmo implementando novas funcionalidades) de forma livre com a condição de que os responsáveis por estas modificações deverão torná-las disponíveis para os demais. Este modelo permite não só que mais pessoas tenham acesso ao conhecimento comum sobre o programa, mas que outros possam utilizá-lo sem custos. Exemplos de programas desenvolvidos por iniciativas destas são o sistema operacional Linux e o pacote de produtividade OpenOffice.

2.3. Comunicação de dados

Embora existentes desde o surgimento dos computadores eletrônicos, tecnologias cujo uso tem se destacado recentemente são as que permitem a intercomunicação de dados. São um tipo especial de TIC, porque conjugam *hardware* e *software* em unidades básicas com o intuito de transmitir dados.

A melhor forma de visualizar esta descrição complexa é imaginar a transmissão de dados entre dois computadores ligados por um cabo de rede. A rede de comunicação de dados montada entre esses dois computadores é formada pelo *hardware* (as placas de redes em ambos os computadores, além do próprio cabo) e pelo *software* (programas especiais de transmissão de dados instalados em ambos os computadores, de forma que cada um entenda o que o outro enviou através do *hardware*).

Mais uma vez, embora o exemplo mencionado envolva o emprego de computadores convencionais, as redes de comunicação de dados podem ocorrer em escalas variáveis. Isso permite que dispositivos como um relógio digital, um telefone celular ou mesmo um computador, possam funcionar de forma autônoma. Da mesma forma, permite que redes complexas de comunicação – tais como redes sem fio, de celular ou mesmo a Internet – existam.

Partindo de escalas menores (e até microscópicas), as redes de dados podem se constituir de simples barramentos de transmissão de informações em uma placa eletrônica ou entre várias. Isso possibilita, por exemplo, que os números que você digita no teclado de um telefone sejam efetivamente transmitidos até o interior do aparelho onde são convertidos nos tons de discagem. Num segundo momento, estes mesmos tons são transmitidos via rede telefônica para uma central, onde formam um padrão de endereçamento que permitirá à central saber para quem você quer ligar.

A própria existência das redes de comunicação de dados permite que os computadores funcionem. Ao descrevermos o *hardware*, vários dispositivos foram citados. Porém, para que funcionem em conjunto é necessário existir componentes que transmitam dados entre eles, controlando para onde vão, como vão e como serão tratados por cada componente que os recebe. Por isso, ao digitar as letras que formam este texto em um componente de *hardware* de entrada, elas passam por um longo caminho de transmissão de dados, pelo processador, até aparecerem na tela do monitor. O fluxo de informação, controlado pelo *software* (sistema operacional e pela aplicação de edição de texto), passa pela rede de dados interna do computador.

Uma vez esclarecido o conceito das redes de comunicação de dados, passemos ao que é largamente definido como o papel dessas TICs na sociedade atual. Da mesma forma como podem interligar componentes internos de um dispositivo eletrônico (cujos computadores são apenas um exemplo), elas podem conectar diversos dispositivos eletrônicos. O exemplo da telefonia convencional acima é uma aplicação relativamente antiga. Mas, atualmente, as redes de computadores e de telefonia celular são as vedetes, e cada vez mais confundíveis entre si.

Computadores e celulares possuem várias características em comum. Uma delas é a diversidade de *hardware* e *software* que empregam. Ambos podem possuir processadores, sistemas operacionais e dispositivos de comunicação distintos. Para que o processo de intercâmbio de informação seja possível, é necessária a definição de protocolos de troca de dados. Como em diplomacia, um protocolo é um padrão de códigos que deve ser seguido pelas partes envolvidas. Logo, para que computadores completamente diferentes troquem informação, devem fazê-lo segundo um padrão pré-definido.

Protocolos são a chave das redes de dados. Sem eles, elas não existiriam. E com o avanço no desenvolvimento deles, também elas são impulsionadas. Com o aumento da escala de uso de um protocolo, maiores são as chances de que vários dispositivos possam trocar informação entre si. Assim surgiram as redes de computadores. Primeiro, em ambientes pequenos, com poucas máquinas. Posteriormente, se espalhando por um prédio. Depois, para outros prédios. Finalmente, não respeitando fronteiras, seja pelo tamanho das redes ou pelo grande número de dispositivos interligados que abrangem, e tampouco as barreiras físicas, com as redes sem fio (que utilizam protocolos próprios) e que, por exemplo, permitem a comunicação móvel.

Um conjunto de tecnologias e protocolos que merece destaque é identificado pela sigla TCP/IP (junção das siglas de dois de seus protocolos). Traduzindo para o português, TCP significa Protocolo de Controle de Transmissão (no inglês *Transmission Control Protocol*). Já IP seria Protocolo de Inter-rede (no original *Internet Protocol*). Como dá para se inferir, o termo em português "inter-rede" não prosperou. A palavra original em inglês veio a ser utilizada para descrever algo que muitos já descrevem como a maior invenção da humanidade: a Internet, que nada mais é do que uma rede em que qualquer dispositivo conectado a ela pode enviar e receber dados utilizando o conjunto de protocolos TCP/IP. Simples não? E isso é apenas o começo.

A Internet tem como característica ser uma rede de várias redes de comunicação. De forma resumida (e em níveis genéricos), cada dispositivo a ela conectado inicialmente está ligado a uma rede de um provedor de acesso. Este provedor de acesso, por sua vez, está conectados a outros provedores de larga escala. Finalmente, estes provedores são conectados entre si formando a teia de comunicação a qual estamos todos interligados. A Internet permite que dispositivos de *hardware* distintos, que utilizam diferentes tipos de *softwares*, tenham acesso mundial às informações da mesma maneira, desde que os *hardwares* e *softwares* envolvidos utilizem o mesmo conjunto de protocolos.

Dentre o conjunto de protocolos TCP/IP que caracterizam a Internet existem os padrões para aplicações ou serviços, tais como transmissão de *e-mails* ou mensagens eletrônicas (SMTP - *Simple Mail Transfer Protocol* e POP - *Post Office Protocol*), transferência de arquivos (FTP - *File Transfer Protocol*), comunicação direta entre dois computadores na Internet (TelNet), etc. Porém, o mais importante destes protocolos de aplicação é o Protocolo para Transferência de Hipertexto, ou HTTP (em inglês, *Hypertext Transfer Protocol*). Ele permite a criação de conteúdo textual que seja de leitura não seqüencial e que incorpore visualização de imagens, sons, vídeos e outros tipos de conteúdo num mesmo lugar. A leitura seqüencial é a que temos, por exemplo, em livros. Para atingir a próxima página, temos que ler o conteúdo da página atual até o seu fim. Na leitura não seqüencial, um indivíduo pode simplesmente pular de um determinado ponto do texto para outro ou mesmo para outro texto que seja virtualmente relacionado ao ponto de partida. Um texto neste formato é denominado hipertexto, os pontos de ligação são denominados hiperlinks e o processo envolvido é chamado de navegação.

O conjunto de hipertextos e seus respectivos conteúdos incorporados compõem o que é considerada a aplicação mais famosa da Internet, a World Wide Web (WWW ou simplesmente Web). Os hipertextos passaram a ser denominados páginas Web, com a característica que cada uma delas possui um endereço (ou identificador) único para acesso. De um ponto de vista técnico, a Web nada mais é do que uma aplicação do protocolo HTTP, que por sua vez é um padrão de serviço disponível pelo conjunto de protocolos TCP/IP, que tem o intuito único de permitir o intercâmbio de informações.

Ao permitir o compartilhamento e a colaboração entre dispositivos em escalas cada vez maiores, o advento das redes de comunicação de dados, principalmente simbolizadas na Internet e na Web, possibilita também interligar pessoas com o objetivo de disseminar informação e construir conhecimento. Isso já faz parte do nosso cotidiano. Como profissionais da informação, o nosso papel é não apenas entender os impactos dessa revolução, mas também direcioná-la para o bem da sociedade. Entender o desenvolvimento e o funcionamento das TICs como instrumentos de competência informacional é uma das maneiras de fazê-lo.

3. COMPETÊNCIA INFORMACIONAL

Historicamente, a habilidade especial em usar a informação nasceu em um ambiente onde saber ler, escrever e realizar contas aritméticas não era o suficiente. Quando havia todo o aparato educacional e algum acesso possível à comunicação, era necessário saber quais informações seriam importantes para garantir uma melhoria na qualidade de vida: de uma boa jornada de trabalho à decisão sobre o voto nas eleições.

Olhando a história e comparando com os dias atuais, é possível observar que o ideal de uma sociedade bem informada ainda é almejado, com a facilidade que pode fazer com que o objetivo seja mais melhor alcançado: as TICs. Com esse suporte, os comportamentos em relação à informação ficam mais ágeis, há o crescimento do acesso da população à grande rede de computadores e a abertura dos vários canais emissores de informação característica da Web. No entanto, podemos nos perguntar se este grande acesso também traz a qualidade de vida que foi proposta inicialmente quando se criou a expressão *information literacy* na década de 70 do século XX.

Neste período, três discussões foram a base para o desenvolvimento de uma nova política de educação (BEHRENS, 1994, p. 310):

- **Fontes informacionais aplicadas em situações no trabalho;**
- **Técnicas e habilidades necessárias para usar recursos de informação;**
- **Informação para tomada de decisão.**

Evidentemente, esses pontos foram essenciais para o desenvolvimento de uma nova mentalidade em relação ao próprio conhecimento humano, passando por fases aplicadas a áreas diversas como:

- **Aplicação ao trabalho**: Identificar fontes informacionais para agregar valor ao trabalho de funcionários de uma empresa;
- **Oranização do conhecimento**: Saber localizar e usar a informação;
- **C omunicação**: Discernir informações publicadas nos meios de comunicação, utilizando canais alternativos de informação;
- **Política**: Utilizar informações no processo do cumprimento de responsabilidades civis (por exemplo, eleitores bem informados podem vir a tomar decisões mais inteligentes

Ao longo do desenvolvimento dos estudos sobre o assunto, surgem modelos variados de aplicação em competência informacional. Cada modelo sugere diferentes habilidades, mas que envolvem basicamente as mesmas capacidades individuais no trato da informação. Um dos modelos, criado pelo SCONUL (2009) – Society of Colleges National and University Libraries, contempla habilidades associadas à biblioteca e tecnologia:

- **PILAR 1:** Reconhecer a informação necessária: Ao reconhecer uma necessidade de informação, o usuário precisa identificar qual o seu conhecimento sobre o assunto e qual é exatamente a informação que precisa saber (preencher a lacuna informacional).
- **PILAR 2:** Distinguir formas de preencher as lacunas – Ao identificar a necessidade informacional, o usuário deve conhecer quais fontes de informação estão disponíveis (jornais, livros, pessoas, organizações, Web), e qual poderia satisfazer sua necessidade.
- **PILAR 3:** Construir estratégias para localizar informação – Para os autores, não basta identificar a fonte. É necessário ainda identificar seu funcionamento a fim de facilitar a estratégia de busca, pois cada uma delas possui uma forma peculiar de organização da informação.
- **PILAR 4:** Localizar e acessar a informação – Neste tópico, são associadas habilidades de acessar a fonte e recuperar a informação. Por isso, o uso das TICs e o conhecimento de cada fonte como citado no pilar anterior. A utilização de bases de referências ou resumo compõe parte dessa etapa. A recuperação da informação em formato integral é o que vai distinguir a fase de busca da fase de acesso.
- **PILAR 5:** Comparar e avaliar Infere-se que o usuário deva ter acesso a mais de uma fonte para realizar esta tarefa. Por isso a importância do primeiro pilar, saber exatamente qual pergunta deve ser respondida. A partir disso, faz-se uma relação cuidadosa e crítica com o conteúdo recuperado. Para essa etapa, é importante ter conhecimento sobre processos de avaliação pelos pares em periódicos científicos, coleta e análise de dados e métodos de pesquisa com o intuito de saber como avaliar a relevância e a qualidade da informação recuperada.

- **PILAR 6:** Organizar, aplicar e comunicar – O uso da informação recuperada deverá servir para somar aos conhecimentos prévios do usuário, tomar decisões e compartilhar seus resultados com outras pessoas. Esta etapa compõe as habilidades em comunicar e compartilhar a informação relevante recuperada, utilizando recursos de trabalhos acadêmicos, conversas, e-mail, etc. A comunicação deve ser realizada em um meio apropriado para que atinja o público de interesse. Para a realização dessa atividade, o conhecimento das leis de direitos autorais e propriedade intelectual se faz necessário.
- **PILAR 7:** Sintetizar e criar – A criação de um novo conhecimento ou um produto que seja resultado da pesquisa é parte essencial da competência informacional, apesar de ser uma tarefa que não se completa em casos não acadêmicos.

Observa-se que o modelo não dissocia as habilidades para uso em bibliotecas (neste caso, físicas) e as tecnologias de informação. O modelo proposto foi criado para servir de base para o desenvolvimento de outros modelos para qualquer tipo de usuário.

O modelo criado por Michael Einsenberg e Robert Berkowitz, *The big 6,* aplica-se a estudantes que precisam ser mais eficientes em suas pesquisas, e sugere seis habilidades, que segundo os autores, são vitais para a sobrevivência no século XXI.

Quadro 1 – Modelo *The Big 6* (EISENBER; BERKOWITZ, 2010, tradução dos autores)

Etapa	Ação 1	Ação 2
Definição da tarefa	Definir o problema da informação	Identificar a informação necessária
Estratégia de busca da informação	Determinar todas as possibilidades de fontes	Selecionar as melhores fontes
Localizar e acessar	Localizar fontes (intelectualmente e fisicamente)	Encontrar informações dentro das fontes
Uso da informação	Dedicação (isto é, ler, ouvir, ver, tocar)	Extrair a informação relevante
Síntese	Organizar de múltiplas fontes	Atualizar a informação
Avaliação	Julgar o resultado (efetividade)	Julgar o processo (eficiência)

Etapa 1: Assim como o modelo *The seven pilars*, a primeira etapa é saber qual pergunta deve ser respondida. A definição da necessidade informacional auxilia no método como as outras etapas serão executadas. Perguntas como "que produto final eu espero entregar?", "qual o prazo que eu tenho?", ou ainda "quais assuntos são correlatos?" devem ser feitas.

Etapa 2: Nessa etapa, o usuário deve listar fontes possíveis de pesquisa e selecionar não só as melhores como as que de fato são passíveis de acesso, fator determinado pelo custo e tempo.

Etapa 3: A estratégia de busca, a forma de abordagem da fonte (física e intelectual) são processos que fazem parte da localização e do acesso.

Etapa 4: A dedicação em ler, ouvir e ver significa entender o conteúdo e analisar se está de acordo com a pergunta inicial.

Etapa 5: Organizar o conteúdo de acordo com métodos que possam ser entendidos pelo usuário posteriormente.

Etapa 6: A avaliação do processo se dá tanto pela qualidade do objetivo que foi atingido como pela forma com que foi atingido. Tal comportamento serve de aprendizagem para pesquisas futuras além de modelos para outros usuários.

O modelo *Empowering Eight* foi proposto por pesquisadores asiáticos para disseminar o conceito nos países da região. Este modelo pode ser adaptável a crianças e adultos em fase escolar.

Quadro 2 – Modelo Empowering Eight (SAYERS, 2006, tradução dos autores)

Passos	Componentes	Resultados
1	Identificar	Identificar o assunto Definir o assunto, objeto Identificar palavras-chave Identificar diferentes tipos de recursos onde a informação pode ser encontrada
2	Explorar	Localizar recursos apropriados para o assunto escolhido Encontrar informações apropriadas para o assunto escolhido Fazer entrevistas, pesquisas de campo e outras pesquisas
3	Selecionar	Escolher a informação relevante Determinar qual fonte é mais fácil, mais difícil e a mais certa Registrar a informação relevante através de notas ou organização visual Identificar os estágios no processo Colecionar citações apropriadas
4	Organizar	Classificar a informação Distinguir entre fatos, opiniões e ficção Conferir tendências, linhas de pensamento Seqüenciar a informação em uma ordem lógica Usar a informação visual para compará-las
5	Criar	Preparar a informação em suas próprias palavras Revisar e editar sozinho Finalizar o formato bibliográfico
6	Apresentar	Compartilhar a informação com o público Exibir o produto em um formato adequado
7	Avaliar	Aceitar a opinião de outros estudantes Auto-avaliar a apresentação de acordo com o que o professor Refletir em como deveria ter feito melhor Determinar se novas habilidades foram aprendidas Considerar o que pode ser feito melhor da próxima vez
8	Aplicar	Revisar a opinião e a avaliação fornecida Esforçar-se para aplicar o conhecimento em novas situações e entender Quais habilidades podem ser usadas novamente

Os modelos disponíveis servem para definir o que é competência informacional, bem como para mostrar que sua definição é ampla e depende do perfil de cada indivíduo, quanto ao seu grau acadêmico, finalidade da informação desejada e experiência individual com pesquisas. Por isso, quando falamos em competência informacional podemos englobar toda a história de seu desenvolvimento conceitual, com o diferencial que as fontes ou recursos informacionais dispõem cada vez mais das TICs. No manual da ALA, publicado no ano 2000, a tecnologia era vista como um auxílio para um comportamento de competência em informação:

> Competência informacional está relacionada com habilidades em tecnologia de informação, mas contém implicações mais amplas para o indivíduo, o sistema educacional e a sociedade. Habilidades em tecnologia da informação capacitam o indivíduo a usar computadores, aplicações de software, bancos de dados e outras tecnologias para atingir uma grande variedade de objetivos acadêmicos, profissionais e pessoais. Indivíduos competentes em informação necessariamente desenvolvem algumas habilidades em tecnologia (ALA, 2000, p. 3)

A partir disso tentou-se definir quais habilidades eram necessárias para ser competente em tecnologia da informação.

4. COMPETÊNCIA INFORMACIONAL EM TEMPOS DE WEB

Por meio do desenvolvimento e da criação de serviços da Internet, as pessoas descobriram que compartilhar informações de forma ampla, sem fronteiras e em grande velocidade também é uma grande (se não a maior) oportunidade para compartilhar e construir conhecimento. Só por isso já é possível afirmar que o desenvolvimento das TICs mudou as nossas vidas. Mas, indo mais além, a globalização do conhecimento tem levado ao desenvolvimento cada vez mais rápido de novas tecnologias e também à dificuldade crescente em acompanhá-las. Eis o porquê do questionamento que suscitou a explicação sobre *software*, *hardware* e redes de comunicação de dados. Vamos resgatá-lo: "Qual a real importância do conhecimento sobre as TICs como competência informacional?"

O desenvolvimento dessas tecnologias tem influenciado não só o aprendizado educacional formal, mas também a nossa estrutura de conhecimento cotidiano e, até mesmo, os requisitos para o exercício profissional. A própria concepção de competências informacionais para profissionais de informação (principalmente, mas não apenas) passa pelo domínio destas tecnologias. Isso porque o fluxo informacional conforme visto atualmente é totalmente indissociável destas tecnologias.

Entretanto, o que não se pode esquecer é o fato de que a simples existência dessas tecnologias não é por si só fator responsável pelo preenchimento dos requisitos informacionais necessários para correlacionar o conhecimento e a utilização dele para o bem da sociedade. Aí está uma distinção entre as TICs e as competências informacionais que é de extrema importância para o entendimento dos profissionais de informação.

O objeto de estudo da competência informacional deve estar embasado no comportamento que diferentes públicos têm em relação à informação, podendo levar em consideração princípios de comportamento informacional adotados para a interação com as tecnologias de informação e comunicação. Nesse sentido, as habilidades definidas como competência informacional no contexto das tecnologias de informação estão direcionadas para tarefas que contemplam o uso eficiente dos recursos tecnológicos para satisfazer necessidades informacionais em ações específicas:

a. **Conhecimento:** É necessário compreender os elementos que interferem o uso eficiente de *hardware* e *software* e comunicações de dados. Neste caso, algumas habilidades como identificar tipos diferentes de *hardware* e conhecer diversos sistemas operacionais dão ao usuário a base de um entendimento computacional básico que não é isolado de outras habilidades informacionais.
b. **Comunicação:** O ponto principal das redes, a transmissão de dados, passa pelo entendimento e uso destas, o que significa comunicar no seu sentido mais real. Aqui, o sujeito é receptor e emissor de dados. Por isso, conhecer as ferramentas disponíveis para disseminar e compartilhar dados é essencial para o usuário em tempos de web. O uso de *wikis*, *blogs*, fóruns, folksonomias, redes sociais, conversas em tempo real são tarefas de comunicação onde se encontram espalhados vários pontos emissores de dados.
c. **Recuperação:** A recuperação está submetida a todas as etapas descritas nos modelos anteriormente apresentados, pois a estratégia de busca a ser utilizada pelo usuário depende de como ele vê sua necessidade informacional, ou seja, como ele pergunta: "como eu vou achar isso na rede?". Importante destacar, também, que tal habilidade depende ainda muito da organização disponível das informações na rede e dos métodos disponíveis de busca.

d. Uso: Quando estamos diante de um objeto desconhecido precisamos entender seu funcionamento para ter total habilidade no seu uso. Com as tecnologias, a peculiaridade comum é que as evoluções das suas versões tornam seus modelos muitas vezes intuitivos, deixando para o usuário a manipulação para treinamento e prática. Como as competências informacionais estão ligadas ao comportamento individual, é possível afirmar que o receio ou a substituição por sistemas mais simples podem deixar o usuário com defasagem em suas habilidades de competência informacional por deixar de aproveitar modelos tecnológicos mais apropriados para sua necessidade específica.

e. Avaliação: A avaliação se faz a partir de conhecimentos pré-concebidos a respeito de um assunto determinado. No sentido da competência informacional aplicada às TICs e Web, o usuário deve ter uma visão crítica do que se pode obter de vantagem ou como se pode agregar valor ao seu cotidiano quanto ao uso de *hardware* e *software* (questões financeiras) além de uso dos sítios como fontes informacionais (veracidade, profundidade, autoria). A avaliação deve ser entendida como um processo de comparação de todos os elementos de TI envolvidos quanto à adequação ao estilo de vida do usuário. Reforça-se, neste caso, a idéia de que para cada usuário existem necessidades informacionais diferentes, implicando em avaliações cada vez mais criteriosas quanto às fontes informacionais usadas.

Quadro 3 – Competências informacional em tempos de Web

<table>
<tr><td rowspan="7">Competência informacional em tempos de Web</td><td colspan="3">Habilidades</td><td>Componentes</td></tr>
<tr><td colspan="3">• Conhecer conceitos de segurança da informação, tipos de vírus, crimes digitais e direitos autorais;
• Conhecer bases de dados;
• Diferenciar conhecimento científico de conhecimento popular disponível na rede web;
• Entender os mecanismos de busca avançada nos motores de busca;
• Utilizar fóruns para discutir conhecimento de interesse;
• Utilizar recursos de redes sociais;
• Utilizar recursos de ensino à distância para fins acadêmicos</td><td rowspan="5">Avaliação

Comunicação

Conhecimento

Recuperação

Uso</td></tr>
<tr><td>Estar atento às tendências de mercado em hardware</td><td>Uso de softwares básicos convencionais ou livres</td><td>Entender como se dá o processo de transmissão de dados entre dispositivos eletrônicos</td></tr>
<tr><td>Saber das capacidades e limitações de cada máquina, incluindo seu processamento, custo e benefício</td><td>Entendimento dos sistemas operacionais, suas diferenças, vantagens e desvantagens</td><td>Compreender a vantagem da “inteligência coletiva”</td></tr>
<tr><td>Hardware</td><td>Software</td><td>Comunicação de Dados</td></tr>
<tr><td colspan="3">Tecnologias de Informação e Comunicação</td></tr>
</table>

O modelo acima indica habilidades e componentes que sugerem o melhor uso de recursos disponíveis na rede a partir dos elementos de TICs (*hardware*, *software* e comunicação de dados). No modelo, temos a lista dos componentes como elementos cognitivos que interferem em todas as habilidades no uso das tecnologias. Entretanto, a lista de habilidades descritas ainda está longe de ser completa, pois tecnologias são respostas às necessidades humanas sociais, culturais e organizacionais e todas em constantes transformações seguindo o percurso da evolução natural do ser humano em contato com a informação.

5. REFERÈNCIAS DO CAPÍTULO 14

ALA. American Library Association. *The information literacy competency standards for higher education*. Chicago: Association of College and Research Libraries, 2000.

BEHRENS, Shirley J. *A conceptual analysis and historical overview of information literacy*. College and Research Libraries, p. 310-323, July 1994.

EISENBERG, Mike; BERKOWITZ, Bob. *The Big 6 Model.* Disponível em: <http://www.big6.com/>. Acesso em 8 maio 2010.

MIRANDA, S. Identificando competências informacionais. *Ciência da Informação.* v. 33, n. 2, p. 112-122, maio/ago. 2004. Disponível em: <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/83/76>. Acesso em 16 jul. 2009.

MOURA, M. Inteligência embarcada. *Quatro Rodas,* São Paulo, set. 2005. Seção Novas Tecnologias. Disponível em: <http://quatrorodas.abril.com.br/reportagens/novastecnologias/conteudo_151380.shtml>. Acesso em: 15 jul. 2009.

SAYERS, R. *Principles of awareness-raising for information literacy: a case study*. Bangkok: UNESCO, 2006.

SCONUL. *Society of College, National and University Libraries*. [S.l.], 2009. Disponível em: http:///www.sconul.ac.uk. Acesso em 9 maio de 2010.

SOCIEDADE DA INFORMAÇÃO NO BRASIL. *Livro verde*. Organizado por Tadao Takahashi. Brasília: Ministério da Ciência e Tecnologia, 2000. 195p. Disponível em: <http://www.mct.gov.br/index.php/content/view/18878.html>. Acesso em 10 ago. 2009.

TARAPANOFF, K *et al*. Funções Sociais e Oportunidades para Profissionais da Informação. *DataGramaZero – Revista de Ciência da Informação*, v. 3, n. 5, out. 2002. Disponível em <http://www.datagramazero.org.br/out02/F_I_art.htm>. Acesso em 1 jul. 2009.

WIJETUNGE, P.; ALAHAKOON, U.P. Empowering 8: the Information Literacy model developed in Sri Lanka to underpin changing education paradigms of Sri Lanka. *Sri Lanka Journal of Librarianship & Information Management*. Vol. 1, n. 1, pág. 33-41, 2005.

Sítios recomendados:

http://big6tools.pbworks.com/
http://www.big6.com/blog/
http://www.whitehouse.gov/the_press_office/Presidential-Proclamation-National-Information-Literacy-Awareness-Month/
http://ictnz.com/infolitmodels.htm (modelos de competência informacional)
http://sibi-usp-agentes-informacao.blogspot.com/
http://competencia-informacional.blogspot.com/

2010

Instituto Brasileiro de Informação em Ciência eTecnologia (IBICT)
e
Grupo de Pesquisa Estudos sobre a Representação e Organização da Informação e do Conhecimento – EROIC
Programa de Pós-Graduação em Ciência da Informação (PPGCInf) da Faculdade de Ciência da Informação da Universidade de Brasília (UnB/FCI)

E A CIÊNCIA DA INFORMAÇÃO