

Políticas de Monitoramento da Informação por Compressão Semântica dos seus Estoques[1][2]

Information monitoring policies through stock semantic compression

por [Aldo de Albuquerque Barreto](#)

Resumo: Este artigo se orienta para o estudo da estrutura do texto escrito e sua análise morfológica com a finalidade de extrair informações para uso na gestão estratégica da informação, localizada em estoques específicos [3] [4]. Visa, ainda, fornecer subsídios para um processo de monitoração de conteúdos informacionais em língua portuguesa e a realização de outros estudos de administração da informação. Procura indicar subsídios técnicos e teóricos para construção de softwares para o estudo de contextos de informação utilizando o instrumental da ciência da informação e do processamento computacional do português em linguagem natural. Ambiciona ser um instrumento estratégico para localizar e caracterizar através de palavras-chave conteúdos de famílias de textos visando a gestão e o controle de um estoque específico de informação.

Palavras-chave: Compressão Semântica, Monitoramento da Informação, Estoques de Informação, Palavras-chave

Abstract: This research is guided for the study of the structure of the written text and its morphologic analysis with the purpose of to extract information for strategic administration of the information, located in specific stocks as to supply and give subsidies for construction of intelligent apparatus of monitoring contents in portuguese language; other studies where this type of software agent is useful is also analysed. The intention of this research tries to supply technical and theoretical subsidies for construction of indispensable software to the current scenery in several information context using methodologies of the information science. The aggregation of the methodology and the software agent is intended to be a tool for management and control of information.

Keywords: Semantic Compression, Information Monitoring, Information Stocks, Key-words

Uma explicação teórica inicial

O objetivo do presente artigo é apresentar resultados de uma pesquisa sobre “A Compressão Semântica[5] nos Estoques de Informação” e os resultados que já podemos apresentar após alguns meses de trabalho.

O estudo se orienta para a análise da estrutura do texto escrito e sua análise morfológica com a finalidade de extrair informações tanto para uso na gestão estratégica da informação, localizada em estoques[6] específicos, como para fornecer subsídios para construção de instrumental de monitoração de textos em língua portuguesa e outros estudos onde um agente de software desta natureza seja necessário. Em sua intenção paralela procura fornecer subsídios técnicos e teóricos para construção de agentes indispensáveis ao controle do cenário atual em diferentes contextos informacionais.

Uma análise da eficiência econômica e da viabilidade dos produtos e serviços de informação nos remete a uma reflexão da apreciação da manifestação do fenômeno da informação, aqui entendido, como a sensibilidade na percepção do conteúdo semântico das estruturas[7] de informação pelos sentidos e pela consciência. A essência deste fenômeno, sempre raro e surpreendente, se mostra pela transformação de conjuntos simbólicos, uma peça de informação, em realizações de uma consciência individual ou coletiva, na criação ou na assimilação[8] desta informação.

É neste sentido que, a informação sintoniza o mundo, pois referencia o homem ao seu passado histórico, às suas cognições prévias e ao seu espaço de com(vivência), colocando-o em um ponto imaginário do presente, com uma memória do passado e uma perspectiva de futuro. Coloca o indivíduo em um presente contínuo espaço de apropriação da informação.

Assim, qualquer reflexão sobre as condições políticas, econômicas ou sociais de um produto de informação está condicionada a uma premissa básica da existência de uma relação da informação com uma geração do conhecimento[9].

Desta forma , para demarcar nosso trabalho usaremos o conceito de informação como sendo:

“conjuntos simbolicamente significantes com a competência e a intenção de gerar conhecimento no indivíduo, em seu grupo e na sociedade. ”

Especificado o conceito, a informação fica qualificada como um instrumento modificador da consciência do homem e de seu grupo social. Deixa de ser, somente uma medida de organização, por redução de incerteza, para ser a própria organização em si ao relacionar o homem aos seus passado e ao seu futuro e a um lugar do presente onde habita com outros. Fica, ainda, estabelecida uma relação entre informação, a sua produção pelo emissor e o conhecimento, este só se realizando se a informação for *percebida e aceita*, colocando o indivíduo sensível em um estágio melhor, consciente consigo mesmo e dentro de um mundo onde se realiza a sua odisséia individual.

Como agente mediador da produção de conhecimento, introduz-se o conceito de assimilação[8] da informação, como sendo um processo de interação entre o indivíduo e uma determinada estrutura de informação, interação com apropriação que, vem gerar uma modificação em seu estado cognitivo inicial, produzindo conhecimento[9], que se relaciona corretamente com a informação recebida. É um estágio qualitativamente superior ao simples acesso à informação. Não se pretende aqui debater as questões filosóficas sobre a teoria do conhecimento. Aceitamos, que conhecimento é uma modificação provocada no estado cognitivo do indivíduo. Em nossa argumentação conhecimento é um processo, um fluxo de informação que se potencializa.

Distinguimos, no nosso estudo, fluxos e estoques quando lidamos com a informação. Temos o um estoque de fatos idéias para percepção, estoques institucionalizados ou não, estes se transformam em outro estoque, quando a informação é inscrita ou registrada.

Entendemos que o conhecimento é um fluxo, isto é uma sucessão de eventos, que se realiza fora dos estoques, em um espaço social e na mente de um ser pensante. É um fluxo subjetivo e diferenciado em relação ao estímulo, e ao receptor. Mesmo que o estímulo seja o mesmo a subjetividade da apropriação o diferencia.

Quando falamos da inteligência pensamos na introdução dinâmica do conhecimento que foi assimilado, para a realidade vivencial do receptor; caracterizada como uma ação: seja ela social, política , econômica ou técnica; representa um conjunto de atos voluntários pelo qual o indivíduo se re-elabora e tenta modificar o seu mundo. *Trata-se de um inicio, do que não iniciou antes e que só se realiza na pluralidade política e vai resultar sempre em uma modificação como resultado da ação inicial.* Ainda que ocorra uma volta, para uma permanência ao estado inicial de partida a ação em si não deixou se ser percorreu seu processo , gerou efeitos. Consideramos , assim, toda inovação como uma ação de inteligência social.

Falo em saber, quando penso no conhecimento que aceitei, apropriei e acumulei nas partições de minha mente. É um estoque que posso evocar para re-elaborar novamente. Pode acrescido, sedimentado ou modificado por novo conhecimento. É meu e de mais ninguém; mas soma-se implicitamente para dar sinais do estado de aprimoramento ou desenvolvimento social e cultural de uma sociedade.

A estrutura de informação

É considerada como qualquer base que aceite uma inscrição de informação; a estrutura é pensada como sendo um conjunto de elementos que formam um todo ordenado e com princípios lógicos, com coerência de raciocínio, de idéias. Este conjunto pode ser linear ou não.

Trabalhamos com a suposição de que, qualquer estrutura de informação textual, um texto de informação, possui características estruturais de linguagem que admitem uma análise morfológica, onde partes podem representar o todo para análise deste todo.

Os fluxos de informação de primeiro nível

Consideramos, também, que os fluxos de informação se movem em dois níveis: em um primeiro nível os fluxos internos de informação se movimentam entre os elementos de um sistema de agregação, armazenamento e recuperação da informação, e se orienta para sua organização e controle.

Estes fluxos internos se agregam, por uma premissa de razão prática e produtivista, em um conjunto de ações pautado por decisões e um agir baseado em princípios. Este é o mundo do gerenciamento e controle da informação. O sistema de armazenamento e recuperação da informação.

Os fluxos de informação de segundo nível

São aqueles que acontecem nas extremidades do fluxo interno de primeiro nível, indicados acima. Os fluxos extremos são aqueles que por sua atuação exibe a Essência do fenômeno de transformação, entre a linguagem do pensamento de um emissor e a linguagem de inscrição da informação.



Assim, no fluxo extremo a direita, a premissa se transforma na promessa, uma promessa ou a esperança de que a informação gerada pelo autor possa ser assimilada como um conhecimento pelo receptor.

No extremo esquerdo do fluxo se realiza um fenômeno de transferência para a informação cuja Essência está na passagem de uma experiência, um fato ou uma idéia, que se encontra em uma linguagem de pensamento do seu criador, para um texto de informação editado por este mesmo gerador; aqui a narrativa mental do autor se transforma em um texto expresso em uma linguagem de edição. Um único fluxo, duas linguagens.

No fluxo do extremo direito temos um processo de cognição[11] que transforma a informação em conhecimento pela correta apropriação desta informação pelo sujeito receptor. Uma apropriação da informação pública para um subjetivismo que se quer privado. Um desfalecer para um renascimento.

No fluxo extremo a direita traz uma desapropriação cognitiva, quando o pensamento se arranja em informação em uma linguagem de inscrição própria, intencionalmente dirigida para tornar-se pública. Aqui a passagem ocorre na direção, dos labirintos do pensar privado para um espaço de vivência pública.

A passagem é da linguagem privada do autor para uma linguagem que ambiciona ser, intencionalmente, para o entendimento do público. Para a aceitação do usuário. Uma pulsão de criação.

Duas pulsões operando em sentido contrário. Uma pulsão de vida na criação de uma nova informação e um desfalecer tenso, intentado para o processo dinâmico, que faz o a informação cumprir uma meta de conhecimento.

Particularmente nos preocupa o estudo do texto, enquanto uma estrutura de informação; o texto[12] e sua estrutura, em um enfoque quantitativo e qualitativa do conteúdo da escritura de informação.

A literatura sustenta nossa pretensão de ser o texto um transmutar do pensar para o escrever: “Escrever é, pois, mostrar-se, dar-se a ver, fazer aparecer o rosto próprio junto ao outro...”. “O que quer dizer que a escrita é um jogo ordenado de signos que se deve menos ao seu conteúdo significativo do que à própria natureza do significante” (Foucault, 1992)[13]

Em nosso trabalho não nos interessa discutir a presença ou a ausência do autor no texto, mas o texto em si, como uma estrutura livre e com características próprias de existência:

“Escrever é retirar-se não para a sua tenda para escrever, mas da sua própria escritura. Cair longe de sua linguagem emancipá-la ou desampará-la, deixá-la caminhar sozinha e desmunida. Deixá-la falar sozinha o que ela só pode fazer escrevendo” (Derrida, 1967)[14].

Existindo uma linguagem do texto, com configurações estruturais delineadas e talvez gerais, seria possível utilizar maquinismos de análise, para obter dados de seu conteúdo, dentro de uma tática de compressão semântica do do seu conteúdo:

“o texto realiza, se não a transparência das relações sociais, pelo menos a das relações de linguagem; ele é o espaço em que nenhuma linguagem comanda a outra, em que as linguagens circulam.... a teoria do texto não pode coincidir senão com uma prática da escrita. (Barthes, 1984) ”.

Estudos da Universidade de Toronto, Canadá, apresentam evidências neurocognitivas de que existem duas diferentes linguagens; uma sendo a do pensar, que antecede a linguagem do editar, de formatar o texto. Estas linguagens teriam características diferenciadas. Nos estudos do grupo canadense que edita a revista "Texte e Informatique" pode-se encontrar indicações de que a linguagem do pensamento se processa em sentenças pequenas; usando, freqüentemente, de cinco até sete palavras, e as palavras com um número pequeno de letras. Assim existirá, em uma fase *a* de edição, uma *nova linguagem*, com características mais formais na estrutura e no estilo, Esta segunda linguagem aparece como que encobrendo a linguagem do pensamento.(Lancashire, 1993)[15].

Nesta mesma linha Walter Ong discute a maneira como a escrita distancia o autor do seu pensar. Ong indica as características da linguagem do pensamento de um ator. *Elaboramos* sobre o texto de Walter Ong (Ong, 1988)[16] para indicar o que entendemos serem **as características básicas da linguagem do pensamento** onde o autor organiza sua narrativa antes de sua transposição para a escrita:

- * as expressões são aditivas em sua narrativa, não se subordinam; é uma linguagem agregativa não é analítica;
- * possui uma tendência para ser redundante ou a re-utilizar conceitos constantemente;
- * possui uma organização conservadora em sua estrutura e simples em sua forma; elabora com

frases pequenas de cinco a sete palavras e com palavras pequenas de quatro a sete letras;

* as expressões têm quase sempre um enunciado de verdade;

* é uma linguagem enfática e direcionada, mantendo um distanciamento objetivo;

* é uma linguagem homeostática[17] ; possui uma tendência à estabilidade interna com um retorno constante aos conceitos já usados;

* é uma linguagem situacional mais que abstrata; tende assim a conceitualizar experiências e memórias adquiridas e então expressa-las com uma relativa proximidade das vivências do cotidiano.

A linguagem **de edição ou da escrita** tem também características próprias:

* explícita, formal, de padrões normativos e de procedimentos formalizados; coloca em prática desígnios ou intentos;

* procura eliminar as repetições das expressões e conceitos;

* procura eliminar a redundância e palavras indeterminadas;

* utiliza figuras de linguagem para facilitar a agilidade das expressões;

* opera com qualidades e relações, e não só com a realidade sensível; é portanto rica em metáforas; possui uma grande fluência de palavras, usa termos peculiares de uma área de atuação, possui extrema liberdade semântica;

* utiliza estruturas sintáticas complexas, mas determináveis e possivelmente generalizáveis e passíveis de padronização;

* é uma linguagem sem controle ou inibição de suas expressões e conceitos; utiliza excesso de sinonímia e de conectores entre conceitos;

* utiliza as palavras sem preocupação com seu tamanho, em frases de construção livre, simples ou complexas e com grande liberdade de elaborar significados.

* é uma linguagem morfológicamente coerente e passível de ter alguma definição de padrões e procedimentos.

É a partir da linguagem de processamento, usada pelo autor no texto, e dentro desta ambiência teórica conceitual que é realizada nossa observação; analisamos a linguagem de edição do autor para concluir resultados de relevância e prioridade de um conjunto de textos, patentes, e-mails, URLs, etc.

Política de compressão da Informação

Aqui se insere o conceito **compressão da Informação**[5], para análise, estudo e monitoramento da informação. O *quantum* de informação disponível é abundante; cabe pois, uma articulação estratégica para conhecer, avaliar e filtrar a informação relevante e prioritária passível de ser transformada em conhecimento, de ser controlada e monitorada para atingir um fim específico.

Uma política de monitorar a informação, por compressão semântica, tem por base a seguinte

articulação teórica-operacional:

* No processo de gestão da informação, armazenada em estoques específicos, existem configurações conjuntas de informação que é relevante e prioritária pela sua singularidade; estes conjuntos podem ser organizados em núcleos. Tais configurações minimizam a tensão cognitiva do usuário ao interagir com um estoque de informação de contextualização intensa;

* A qualidade de informação **relevante**[18] está diretamente relacionada com a quantidade das **palavras mais frequentes**, que aparecem em um texto de informação;

* A qualidade da informação que é **prioritária**[19] pelo seu viés, está diretamente relacionada com as palavras de **freqüência igual a um (1)** no texto de análise. São as palavras que aparecem somente uma vez e mostram, quando devidamente examinadas a singularidade de um texto;

* A informação que é relevante para um determinado receptor é razoavelmente inelástica[20] em relação ao tempo. Não demonstra variações radicais em um determinado espaço de tempo[21];

* A informação que é prioritária, por seu viés de singularidade, é razoavelmente elástica em relação ao tempo, para um determinado receptor. A condição de prioridade por especificidade, modifica-se com maior rapidez em função do tempo;

* As palavras que indicam a relevância de um texto, as de maior de maior freqüência situam-se dentro de um conjunto (clusters) de palavras com uma freqüência limite. A partir deste conjunto a de palavras a relação freqüência de ocorrência e relevância da palavra apresenta rendimentos decrescentes de escala, em relação à relevância. A relevância das palavras nestes clusters, existem em uma condição de rendimentos decrescentes em escala. Quando menor a freqüência menor também, a relevância no cluster considerado.

Assim o corte da freqüência para explicar a relevância fica determinado pelo coeficiente de relevância[22] abaixo indicado, que são calculados para cada conjunto de palavras de uma mesma freqüência.

A fórmula permite indicar que na freqüência 10, por exemplo, 80 % das palavras são relevantes, e a partir daí haverá um continua perda na taxa de relevância, sendo, assim, está a freqüência limite para estudo daquele conjunto específico de documentos;

$$\text{Coeficiente} = \frac{S \text{ Palavras relevantes}}{S \text{ Palavras cluster}} \times 100$$

* As palavras de freqüência igual a um se agregam em *clusters* de significância. Estes clusters são delimitados pelo tamanho das palavras, (em número de letras) contidas naquele cluster. Existe um rendimento decrescente em escala[23] do valor (das palavras) destes clusters em função do tamanho das palavra. O coeficiente de relevância acima indicado é uma opção de trabalho, para o analista, para determinar o tamanho do cluster desejado;

* Existe uma relação positiva entre o tamanho de uma palavra de um texto e a sua qualidade de relevância e prioridade. Isto é palavras maiores são mais significantes; esta parece ser um diferencial de uma linguagem de pensamento e uma linguagem de edição;

* As configurações de zonas de qualidade intensa, com relevância e prioridade podem ser

determinadas automaticamente por agentes de software, sem interferência humana;

Os preceitos operacionais básicos, validados em pesquisa realizada, agregados ao agente de software[*] formam um instrumental de monitoramento, gestão e controle da informação.

Procedimentos operacionais

Para proceder a nossa análise era preciso um corpus de documentos em português digitalizados e preferencialmente disponíveis on-line na Internet. Neste momento foi utilizada a base do projeto Scielo da Fapesp/Bireme. A base *Scientific Electronic Library Online* - SciELO é uma biblioteca virtual que abrange uma coleção selecionada de periódicos científicos brasileiros, em formato digital e disponível na Internet.

Nossa análise mostrou a validade de todo o raciocínio operacional anteriormente citado. As mais importantes: as palavras de frequência 1 indicam a individualidade de um texto, seus métodos, insumos, processos reações, modelos etc. São diferentes das palavras de frequência alta que mostram a relevância, utilidade, da escrita em uma condição mais geral.

Existe uma relação entre o tamanho da palavra, medido pelo número de letras, e sua importância para a relevância e prioridade do texto considerado; A inscrição da informação em uma estrutura utiliza duas linguagens, uma linguagem que se processa no pensamento do autor e outra na qual os textos são editados para o público

O comportamento das palavras em um texto em relação a sua frequência ser maior que 1 ou igual a 1 mostram as seguintes configurações, para três campos distintos:

Tabela 1 - PALAVRAS COM FREQUÊNCIA MAIOR QUE 1 (F>1)

CAMPO DE CONHECIMENTO	CONFIGURAÇÃO GRÁFICA	NÚMERO PALAVRAS NO NÚCLEO	PALAVRAS RELEVANTES NO NUCLEO		NÚMERO DE LETRAS DAS PALAVRAS RELEVANTES		linguagem de edição: pensamento de processamento da informação
			MÉDIA	%	WORDS	a	
CIÊNCIA DA INFORMAÇÃO	hipérbole decrescente, convexa à origem e tendendo ao ¥ no eixo do x	28	12	40	9	3	pensamento convergente/divergente; próprio das sociais aplicadas?
CIÊNCIAS SOCIAIS		80	20	20	9	2	pensamento divergente, fluência das idéias e palavras
CIÊNCIAS EXATAS E DA SAÚDE		16	9	60	9	2	pensamento convergente; vocabulário exato

Núcleo = das palavras de maior frequência até a frequência limite escolhida

a = desvio padrão

Em algumas áreas como a sociologia, comunicação, história a fluência de palavras é tão grande que é preciso operar em um limite bem mais baixo de relevância para evitar perda de conteúdo, aumentando assim o tamanho do núcleo. Pelo contrário em áreas mais formalizadas como a física química, etc um núcleo de 16 palavras tem alta percentagem de palavras relevantes.

Tabela 2 - PALAVRAS COM FREQUÊNCIA IGUAL A 1 (F1 = 1)

CAMPO DE CONHECIMENTO	CONFIGURAÇÃO GRÁFICA	A		B	percentagem B/ A
		Numero de palavras até o cluster com 20 palavras		números de palavras relevantes até o cluster de 20 palavras	
		total	média	média	
CIÊNCIA DA INFORMAÇÃO	Clusters decrescentes em relação ao numero de letras das palavras nele contidas.*	30/50	40	15	30 %
CIÊNCIAS SOCIAIS		60/70	50	20	40 %
CIÊNCIAS EXATAS E DA SAÚDE		30/40	43	18	50 %

* Clusters em escada. Palavras relevantes são diretamente proporcionais ao seu tamanho em número de letras, em um mesmo cluster

Aqui, os Clusters evoluem em progressão decrescente ao tamanho das palavras, medido em número de letras. A Relevância das palavras nos clusters tem uma evolução como em uma progressão decrescente.

Os Clusters crescem no tamanho do conjunto por agregados de palavras com tamanho decrescente:

Exemplo: 1: 3: 7 : 12 : 15 - clusters contendo 1 palavra de 20 letras, 3 de 17 letras, 12 de 14 letras, 15 de 9 letras, assim em diante.

As palavras se *clusterizam*, em importância, em uma progressão decrescente ao seu tamanho em letras.

Existe, ainda, uma diferente articulação da morfologia do texto para diferentes áreas do conhecimento. A tabela 3 a seguir, mostra estas diferenciações, fatos que necessitam, ainda, melhor estudo:

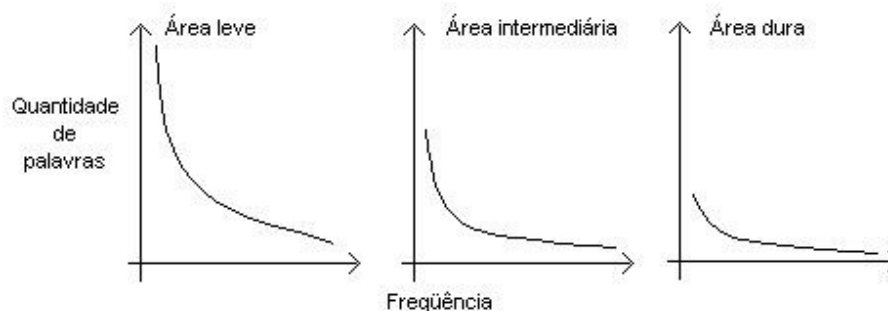
Tabela 3 - COMPORTAMENTO DA PALAVRAS NA ESCRITURA DE INFORMAÇÃO

ÁREA	características do sub-código lingüístico	Léxico de edição do texto	Cadeia de pensamento na edição do texto
LEVES Tipo: ciências humanas e sociais	-difícil delimitação do sub-código -Extrema liberdade semântica. - discurso longo, difuso e informal com conceitos interligados	-estruturas sintáticas complexas -excesso de sinonímia, metáforas, conectores e plurais - linguagem sem controle ou inibição	-fluência de idéias e palavras muita independência em elaborar significados -re-evocação de palavras da memória com grande liberdade conceitual - pensamento divergente na reconhecimento dos conceitos

<p>INTERMEDIÁRIAS</p> <p>Tipo: sociais aplicadas</p> <p>(com um aporte do discurso tecnológico)</p>	<p>-código mais formalizado e visível</p> <p>-condições semânticas controladas</p> <p>- discurso semi-técnico</p>	<p>-estruturas sintáticas elaboradas</p> <p>- vocabulário com alguma inibição das figuras de linguagem</p>	<p>-alguma independência na elaboração dos significados</p> <p>-re-evocação de conceitos da memória com orientação ao assunto</p> <p>- predomina o pensamento orientado a reconhecimento</p> <p>convergente dos conceitos na memória</p>
<p>DURAS</p> <p>Tipo: ciências exatas e da natureza</p>	<p>-código bastante formalizado e delineado</p> <p>- discurso técnico e direcionado</p>	<p>-estruturas sintáticas simples</p> <p>-vocabulário formal e específico</p> <p>- bastante inibição quanto as figuras de linguagem</p>	<p>-pouca fluência de idéias e conceitos</p> <p>-precisão na re-evocação dos conceitos da memória</p> <p>-pensamento altamente convergente na re-evocação das palavras</p>

Conceito[24]; Pensamento Convergente[25]; Pensamento Divergente[26]

Considerando o eixo vertical igual a quantidade de palavras e o horizontal como freqüência das palavras, quanto mais o texto se enquadra na área “leve”, mais a curva se acentua. Quando o texto caminha na direção da área “dura” a curva vai tendendo ao eixo horizontal [18]



Considerações sobre os resultados

A análise precedente indica que, existindo um conjunto de condições teóricas e operacionais coerentes e consistentes e um agente de software apropriado têm-se um poderoso instrumental estratégico para análise, monitoramento e controle de conteúdos informacionais diversos.

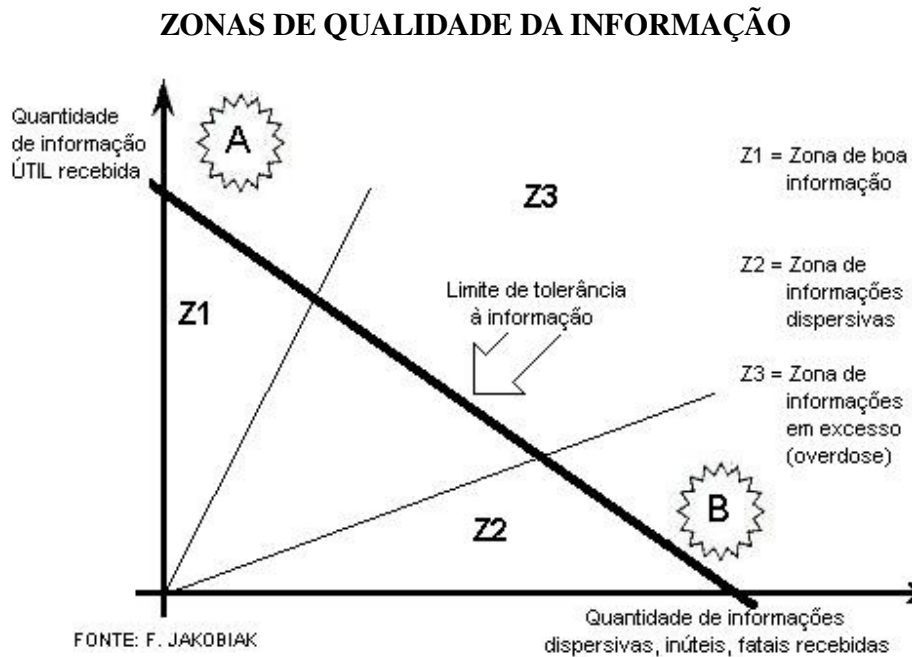
Diferente de uma simples contagem de palavras em um texto ou uma família de textos, sem um direcionamento lógico e uma base teórica e conceitual de análise, propõe-se uma estrutura teórica que fundamente e, direcione esta operação prática e fundamenta com raciocínio adequado uma análise específica.

Trabalha-se com este modelo para verificar desde condições estratégicas dos mais diversos tipos de monitoramento até a gestão da informação para estabelecer zonas de qualidade intensa[27] em um determinado contexto informacional.

Do ponto de vista da gestão e monitoramento da informação em seus estoques a informação relevante e prioritária pode ser zoneada. Na elaboração de uma política de compressão cabe ao administrador

do estoque a responsabilidade de selecionar indicadores representativos desta compactação, que pode ser feita automaticamente.

O Gráfico a seguir, adaptado de Jakobiak[28], mostra o posicionamento das zonas de boa informação, informação dispersiva e informação em excesso em um estoque de informação:



Com a mesma intenção tem-se a **tabela 4**, 2x2, ponderando **relevância** e **prioridade**

		Relevância	
		A	B
Prioridade	A	DOCUMENTOS RELEVANTES E PRIORITÁRIOS	DOCUMENTOS RELEVANTES MAS NÃO PRIORITÁRIOS
	C	DOCUMENTOS NÃO RELEVANTES MAS PRIORITÁRIOS	DOCUMENTOS NÃO RELEVANTES E NÃO PRIORITÁRIOS
		ZONA 1	ZONA 3
		ZONA 3	ZONA 2

A zona de excelência seria a zona 1, que poderia ainda ser fragmentada em diferentes níveis de qualidade. A zona 2 seria descartada para àquele receptor ou grupo de receptores semelhantes. As zonas C3 e B3 necessitam de um trabalho de identificação estratégica junto ao grupo ou a estratégia de monitoramento.

A zona B3 pode indicar, por exemplo, que uma informação sobre turismo pode ser prioritária para um planejamento de férias no mês de dezembro, mas não é relevante no tempo atual. A zona C3 ou pode

indicar que existem informações sem utilidade (não relevantes) naquele tempo e contexto por falta da abrangência ou completeza ou por serem marginais ao foco atual do receptor que lhes atribui prioridade potencial pela possibilidade de uma adição de valor no futuro.

Seguramente, a possibilidade de uma partição dos estoques de informação com a criação de contextos informacionais de excelência para o receptor traria a facilidade de amenizar a tensão cognitiva[29] dos usuários de informação.

Uma boa política, entendida como um conjunto de procedimentos para execução de uma ação de monitoramento da informação com benefícios para sua gestão e controle, com melhoramentos para o receptor e sua qualidade de interatuação com conjuntos de documentos. Interatuação, apropriação, conhecimento.

No caso de uma política, entendida como a arte de controle no governar com, práticas de governo de interesse e favorecimento do Estado, então, um monitoramento desse tipo pode ser tornar um perigoso instrumento de controle das liberdades individuais, da privacidade do sujeito e de seus escritos.

Notas

[1] Nesta notas muitos conceitos foram definidos. Preciso indicar que são conceitos válidos para o contexto deste trabalho. Como diria Derrida, são mais uma aventura de meu olhar que uma maneira de questionar o objeto em si.

[2] O presente artigo é fruto de pesquisa realizado com auxílio da Coordenação de Ciências Humanas e Sociais do CNPq.

[3] Estoques de informação: conjunto de itens de informação agregados segundo critérios de interesse de uma comunidade informacional.

[4] Estrutura: forma de organização de elementos, onde estes adquirem sentido apenas enquanto fazendo parte de um conjunto com sentido de ordenação lógica e racionalidade.

[5] Compressão semântica: ação ou efeito de uma movimento que comprime um corpo, tendendo a aproximar, uma das outras, as partes que o compõem; pretende que determinadas palavras existindo no texto em condições morfológicas demarcadas e especiais podem possuir condições de qualidade, quando em conjunto, que podem ser explicadas por condições específicas de representação e compreensão do conteúdo de todo o texto. Uma destas condições de compressão semântica ou de substituição representativa seria um grupo de palavras representando a relevância e a prioridade pelo viés temático do conteúdo. Existem ferramentas e modelos, que de modo totalmente automatizada podem exercer esta compressão semântica em um texto, analisando e retirando estas palavras para uso estratégico.

[6] Estoques de informação: conjunto de itens de informação agregados segundo critérios de interesse de uma comunidade informacional.

[7] Estrutura: forma de organização de elementos, onde estes adquirem sentido apenas enquanto fazendo parte de um conjunto com sentido de ordenação lógica e racionalidade.

[8] Assimilação: misto de sensibilidade e percepção e apropriação, na assimilação o receptor aceita a informação reconhecendo-a como tal, e esta atua para alterar o seu estoque de saber por acréscimo, por modificação ou sedimentação de saber já estocado.

[9] Conhecimento: é organizado em estruturas mentais por meio das quais o sujeito assimila o meio (informação). Conhecer é um ato de interpretação, uma assimilação do objeto (informação) pelas estruturas mentais do sujeito. Estruturas mentais não são pré-formatadas no sentido de serem programadas nos genes. As estruturas mentais são construídas pelo sujeito sensível, que percebe o meio. A Produção ou geração de conhecimento é uma reconstrução das estruturas mentais do indivíduo através de sua competência cognitiva, ou seja, uma modificação em seu estoque mental de saber acumulado, resultante de uma interação com uma informação. O conhecimento seria operado através de uma ação de cognição.

[10] Essência: ação com vigor de propósitos; onde o fenômeno desenvolve a força de sua energia. Escreve-se o E em maiúscula para diferenciar de essência como natureza.

[11] Cognição: conjunto dos processos mentais subjetivos usados quando da aquisição de conhecimento; relaciona-se às características funcionais e estruturais da representação de um saber, um objeto, um procedimento, fatos e idéias. A cognição sendo um ato de conhecimento, representa as atividades relacionadas como atenção, criatividade, memória, percepção. Um enfoque cognitivista considera a mente humana como um sistema de estruturado de informação, um estoque de informação organizado. A entrada de nova informação é analisada pela mente do receptor, significados são evocados e comparados na memória, e se processa a confirmação de um conhecimento já existente, a modificação de um conhecimento já existente ou a aceitação da informação como novo conhecimento. Em qualquer caso cumpre-se uma cerimônia da informação operando intencionalmente sobre o conhecimento.

[12] Texto: qualquer inscrição de informação em uma base que a aceite e a mantenha, sendo o papel a base mais comum e também a mais conservadora. Pode estar em diferentes bases e em diversas linguagens de inscrição.

[13] Foucault, M., O que é um autor, 3ª edição, Passagens, Lisboa, 1992.

[14] Derrida, J., A Escritura e a diferença, 2ª edição, Perspectiva, São Paulo, Brasil, 1995.

[15] Lancashire, I., Uttering and Editing: Computational Text Analysis and Cognitive Studies in Authorship, *Texte et Informatique*, n. 13/14, (1993): pp 173-218

[16] Ong, W.J., Orality and Literacy: The Technologizing of the Word, Terence Hawkes, New York, 1988

[17] Tendência à estabilidade do meio interno do organismo. Propriedade auto-reguladora de um sistema ou organismo que permite manter o estado de equilíbrio de suas variáveis essenciais ou de seu meio ambiente.

[18] Relevante: tudo aquilo que possui a condição de utilidade, que é a qualidade das coisas materiais e imateriais em satisfazer nossas necessidades e desejos.

[19] Prioridade: qualidade do que está em primeiro lugar; o que antecede aos outros em tempo, lugar, serie ou classe quando da prática de alguma coisa. Em nossa análise a prioridade se associa a individualidade de um texto, ao seu viés.

[20] Elasticidade tempo (E t): variações relativas entre as condições da relevância e prioridade (neste caso) em decorrência da passagem do tempo. No caso da relevância inelástica, indica sua maior permanência com o passar do tempo. A prioridade elástica indica uma maior variação em relação ao

tempo.

[21] Definido aqui como a período de tempo em que a informação permanece relevante para um determinado usuário, considerando sua permanência na área, contexto informacional e contexto reflexivo e profissional; pode variar no entorno de (1-8) anos;

[22] No Curso de Ciência da Informação, no Rio de Janeiro, em 2001, foi oferecida a disciplina "Estrutura e Fluxos de Informação" para alunos de doutorado e mestrado que, durante um semestre, discutiram as bases teóricas e submeteram a prova o modelo de trabalho exposto neste trabalho. Deste curso participaram: Carla Tavares, Cládece Nóbile Diniz, Cláudio Starec, Eugenia Vitória Loureiro, Jaqueline de Almeida Netto, Jose Ricardo Ozório Jardim, Leila Beatriz Ribeiro, Marta Catarina Feghali, Mauro Behring, Michell Xavier da Costa, Mônica Cristina Santiago, Nilton Bahlis dos Santos, Palmira Moriconi, Suzana Tavares Blass, Vera Mangas da Silva. Estes alunos produziram na disciplina excelente material complementar, que aqui aproveitamos.

[23] Rendimento decrescente em escala: termo utilizado em economia; ao se aumentar uma quantidade de produção, em quantidades definidas de determinado elemento, existe uma perda quantitativa ou qualitativa, em quantidades crescentes, na produção total do produto que se está produzindo.

[24] Conceito: a menor unidade de uma estrutura significante, com condições representacionais; para nós o mesmo que uma palavra.

[25] Pensamento convergente: aquele em que a seleção das palavras do texto se direciona a uma cadeia de ligações precisa, determinada, convencional, pontual.

[26] Pensamento divergente: aquele em que a seleção das palavras para o texto caminha em diferentes direções como que pesquisando livremente os meandros das figuras de elaboração do estilo no momento da edição da informação.

[27] Zonas de qualidade intensa: partições de um estoque de informação, onde os textos estejam contextualizados por qualidade de relevância e por condições das propriedades de sua singularidade, seu viés. Zona de relevância com escalonamentos de preferência duma coisa que deve ser posta em primeiro lugar, numa série ou ordem.

[28] JAKOBIAK, F. Maîtriser l'Information Critique, Les Editions Organization, Paris, 1988.

[29] Tensão cognitiva representa o tempo despendido pela informação na procura necessária e o esforço mental de um receptor para apropriar a informação como conhecimento. Devido à alta exposição atual à informação, esta tensão é considerável e todos os instrumentos para sua redução representam uma eficiência no processo cognitivo.

[*]Desenvolvimento de um software específico

Existe em nível internacional uma quantidade razoável de instrumentais para análise computacional da linguagem natural, para as mais diversas finalidades.

Entretanto, existem três problemas com os softwares já desenvolvidos e disponíveis comercialmente, ou do tipo *freeware*: são elaborados para um estudo específico de um determinado pesquisador e não atendem à necessidade de outro estudo específico; b) seu desenvolvimento se direciona para a língua para a qual haviam sido programados, não reconhecendo particularidades do português, tais como acentuação e cedilha; c) grande parte destes instrumentos havia sido projetado para análise literária, como poesia e ficção, e não suportam adequadamente a literatura de ciência e tecnologia.

Foi indispensável construir um software específico para o desenvolvimento de nossa pesquisa. Para tanto, tivemos a colaboração da pesquisa do *Dr. Luiz Carlos Paternostro*, professor da UFRJ, que desenvolveu um instrumento chamado de *Protexto*. O Protexto, ainda em desenvolvimento, foi projetado para as especificidades do nosso trabalho, mas será um software de caráter geral para o estudo do processamento do português como linguagem natural, seguramente um dos poucos existentes no Brasil, com as características de um instrumento orientado para o usuário, fácil de operar e conversacional.

Referências Bibliográficas

1. Arendt, H , A Vida do Espírito- O pensar, o querer, o julgar , Relume-Dumara, Rio,1991.
2. Barreto A. de A. "Padrões de assimilação da informação - a transferência da informação visando a geração do conhecimento", Relatório apresentado ao CNPq em fevereiro de 2000. Publicado como "O Rumor do Conhecimento" pela Revista São Paulo em Perspectiva, v 12, n. 04, Fundação Seade, São Paulo.
3. Barreto A. de A. "A Transferência de Informação, o Desenvolvimento Tecnológico e a Produção de Conhecimento", IBICT/ECO, 1993 (Relatório Apresentado ao CNPq).
4. Barreto, A. de A. "A Informação e o Cotidiano Urbano", IBICT/ECO, 1991 (Relatório apresentado ao CNPq).
5. Barreto, A. de A. "Perspectivas da Ciência da Informação", Revista de Biblioteconomia de Brasília, v. 21, n.2, 1997.
6. Barreto, A. de A. "A Questão da Informação", São Paulo em Perspectiva, v. 8, n. 4 ,1994, p. 3-8, Fundação Seade, São Paulo.
7. Barthes, R, "A morte do autor", em O Rumor da Língua, Edições 70, Lisboa ,1987.
8. Bloor, D , Poppers Mystification of Objective Knowledge, Science Studies , v 4, pp 65-76, 1974.
9. Boulding, K , Knowledge and Life in the Society, University of Michigan Press, USA, 1960.
10. Bourdieu, P , O Poder Simbólico, Bertrand, Rio,1989.
11. Butcher, H J , A Inteligência Humana, Perspectiva, São Paulo, 1968.
12. Farradane, J , Relational Indexing and Classification in the Light of Recent Experimental work in Psychology, Information Storage and Retrieval, vol 1, pp 3-11, 1963.
13. Farradane, J, Knowlwdge, Information and information Science, Journal of Information Science,v2,n2,1980.
14. Farradane, J, The Nature of Information, Journal of Information Science, v 1 , n 3, 1979.
15. Franck, S e Mehler, J (ed.), Cognition on Cognition, MIT Press, USA, 1994.
16. Gardner, H ,The Minds New Science : A history of the cognitive revolution, Basic Books, USA, 1987.
17. Guilford,J P , Three Faces of Intellect, Americam Psychologist, v. 14 , n. 8, 1959.
18. Habermas, J , Conhecimento e Interesse, Guanabara, Rio, 1987.

19. Habermas, J , Ciência e Técnica Como Ideologia, Edições 70, Lisboa, 1987.
20. Heidegger, M , Discurso sobre o Humanismo, Tempo Brasileiro, Rio, 1962.
21. Jakobson, R. , Lingüística e Comunicação , Cultrix, São Paulo, 1993 - Coletânea de trechos selecionados de Roman Jakobson.
22. Levy, P , A Máquina Universo, Atmed, Porto Alegre, 1998
23. Luninn L F (Ed), "Perspectives in Knowledge Utilization", Jasis (Special Issue), v44, n4, 1993
24. Mannheim, K. , Conhecimento e Sociedade
25. Ricoeur, P , Teoria da Interpretação , Edições 70, 1976, Lisboa.
26. Simon, H. "Literary Criticism: A Cognitive Approach", Stanford Humanities Review, SEHR v.4, n. 1, "Constructions of the Mind", updated in 1995.
27. Simon, H. The Sciences of the Artificial, 3rd ed., Cambridge, MA, MIT Press, 1996.
28. Wittgenstein, L. Zettel (Fichas) , Edições 70, Lisboa, 1981

Sobre o autor / About the Author:

Aldo de Albuquerque Barreto

aldoibct@alternex.com.br

Pesquisador Titular Mct