

POLÍTICAS DE MONITORAMENTO DA INFORMAÇÃO POR COMPRESSÃO SEMÂNTICA DOS SEUS ESTOQUES

Aldo de Albuquerque Barreto
Pesquisador Titular MCT

RESUMO

Esta pesquisa se orienta para a análise morfológica do texto com a finalidade de operando de maneira automatizada, extrair informações do conteúdo escrito tanto uso para a gestão estratégica da informação localizada em estoques [3] específicos quanto para fornecer subsídios para construção de cenários de monitoração de conteúdos em linguagem natural e em língua portuguesa. Procura fornecer subsídios técnicos e teóricos para construção de *softwares* para o estudo textos e grupos de textos utilizando instrumental teórico da ciência da informação e do processamento computacional do português em linguagem natural. Representa, teoricamente, uma reflexão da percepção do conteúdo semântico das estruturas [4] de informação pelos sentidos e pela consciência. Ambiciona seu um instrumental estratégico para caracterizar através do estudo de palavras do de textos, fornecer informações para eficiência da organização de contextos informacionais. Não pretende modificar o texto em linguagem natural em nenhum momento.

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 could be planned with this type of software agent and theoretical framework. The intention of this research is 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 never to substitute natural language of information texts.

“Como sabem, eu me aventurei na escrita; mas acho que o que li é mais importante do que escrevi. Pois a pessoa lê o que gosta, porém escreve o que é capaz de escrever”.

J.L.Borges
Esse Ofício do Verso

Uma explicação teórica inicial

O objetivo do presente artigo é apresentar resultados da pesquisa sobre “A Compressão Semântica [5] nos Estoques de Informação”.

O estudo se orienta para a avaliação da estrutura do texto escrito e seu diagnóstico morfológico 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 para monitorar o conteúdo de textos em língua portuguesa através de um agente de software seja para servir a decisões estratégicas em varios níveis e objetivos do gerenciamento da informação. Como sub-produto, o estudo fornece subsídios técnicos e teóricos para construção agentes de programação indispensáveis ao

controle do cenário 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 de suas estruturas [7] pelos sentidos e pela consciência. A essência deste fenômeno, sempre raro e surpreendente, se revela pela transformação de conjuntos simbólicos em realizações de uma consciência individual ou coletiva, tanto na criação como na assimilação [8] de 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 como que em um presente contínuo, um espaço apropriado 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 delimitar nosso trabalho usaremos o conceito de informação como sendo:

“estruturas 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, pois relaciona o homem ao seu passado e ao seu futuro e a um lugar do presente, onde reside politicamente. Fica, ainda, estabelecida uma relação entre informação, com a sua produção por um emissor e o seu destino como sendo o conhecimento; destino só realizado se a informação for **percebida e aceita como tal**, colocando o indivíduo sensível em um estágio qualitativamente melhor para si mesmo e para o mundo onde se realiza a sua odisséia individual.

Como agente mediador na produção de conhecimento, damos o conceito de assimilação⁸ da informação, como sendo um processo de interação entre o indivíduo e uma determinada estrutura de informação, uma ação com apropriação que, vem gerar uma modificação em seu estado cognitivo inicial, produzindo conhecimento⁹, 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 o conhecimento é uma modificação provocada no estado cognitivo do indivíduo acrescentando, consubstanciando ou modificando seu estoque de saber. Em nossa argumentação conhecimento é um processo, um fluxo de informação que se potencializa na consciência do indivíduo.

Distinguimos, a importância de falarmos de fluxos e estoques, quando lidamos com a informação. O estoque é visto como agregado de fatos, idéias, saberes; é a informação inscrita, registrada e domiciliada, e intentada para a percepção do receptor. Entendemos que o conhecimento é um fluxo, isto é uma sucessão de eventos, que se realiza fora dos repositórios, em um espaço social e na mente de um ser pensante. É um fluxo subjetivo e diferenciado, pois é único para cada em relação ao estímulo gerado. Ainda que, o estímulo seja o mesmo, a subjetividade da apropriação o diferencia.

Quando falamos da inteligência pensamos na introdução dinâmica da informação dentro da 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 início, do que nunca iniciou antes e só se realiza na pluralidade política indo resultar sempre em uma modificação como resultado da ação inicial. Ainda que, ocorra uma volta para uma permanência ao estado inicial da partida a ação em si não deixou de percorrer seu processo e gerou efeitos e nada voltará a ser como antes. Toda inovação é uma ação de inteligência social.

Penso no significado do conceito de saber, como o conhecimento que aceitei, apropriei e acumulei nas partições de minha mente. É um estoque que posso evocar para re-elaborar novamente. Pode ser acrescido, sedimentado ou modificado por novo conhecimento. Mas é meu e de mais ninguém; agrega-se socialmente 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 em suas idéias. Esta composição pode ser linear ou não.

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

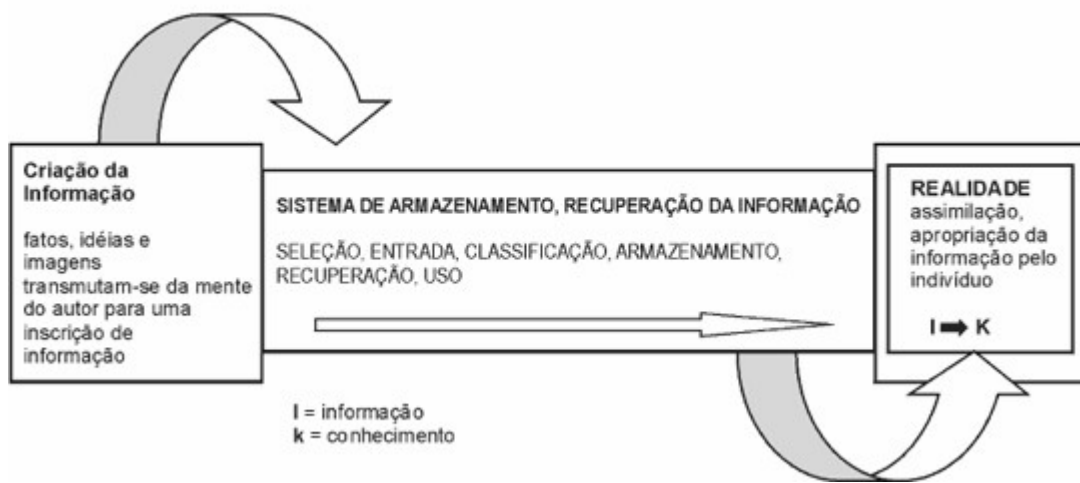
Os fluxos de informação de primeiro nível

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 reunião, armazenamento e recuperação da informação, fluxos que se guiam por uma racionalidade pragmática de organização e controle de repositórios.

Estes fluxos internos possuem uma premissa de razão prática e produtivista, com ações pautadas por decisões e um agir baseado em princípios. Este é o mundo do gerenciamento 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 mostram a Essência[10] do fenômeno de informação e da transformação, entre a linguagem do pensamento do um emissor e a linguagem de inscrição da informação.



desapropriação

apropriação cognitiva

Assim, no fluxo extremo a direita, a premissa se transforma na promessa, a promessa ou a esperança em que a informação gerada pelo autor possa ser assimilada como um conhecimento pelo receptor. No extremo esquerdo do fluxo acontece um fenômeno de transferência da informação cuja Essência está na passagem das ideias da mente do autor em uma linguagem de pensamento para um texto de informação editada na inscrição do texto; assim a narrativa mental do autor se transforma em um texto expresso em uma linguagem de edição. Um único fluxo, duas linguagens.

Aqui, 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 que se fez pública para um subjetivismo do receptor que o quer privado. Um desfalecer para um renascimento.

Lá no fluxo extremo a esquerda acontece uma desapropriação cognitiva, quando o pensamento se arranja em informação com uma linguagem de inscrição própria, intencionalmente dirigida para tornar-se pública. A direção da passagem ocorre 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 à informação cumprir uma meta de conhecimento. Eros e Hipnos quando a mitologia se fizer palavra.

Particularmente nos preocupa, no momento o fluxo esquerdo, onde o estudo do texto, enquanto uma estrutura de informação;

^[12]

o texto e sua estrutura morfológica, com um emoque quantitativo e qualitativo do seu conteúdo.

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 a própria natureza do significante” (Foucault,

^[13]

1992)

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, [14] 1967) .

Existindo uma linguagem do texto, com configurações estruturais delineadas, seria possível utilizar mecanismos de análise, para obter dados de seu conteúdo, dentro de uma tática de compressão semântica de sua substância:

“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) ” .

Na Universidade de Toronto, Canadá, existem evidências neurocognitivas de que duas diferentes linguagens existem no processo de criar a informação; uma sendo a do pensar, que antecede uma segunda linguagem a do editar, do formatar o texto. Estas linguagens teriam características diferenciadas. Nos estudos do grupo canadense que edita a “Revista Texte e Informatique” podem-se encontrar indicações de que a linguagem do pensamento se processa em sentenças pequenas; usando, freqüentemente, poucas palavras, e as palavras com um pequeno número de letras. Assim existirá em uma segunda fase de edição, uma *nova linguagem*, com características mais formais na estrutura e no estilo, Esta segunda linguagem aparece como que encobrindo a

[15] linguagem do pensamento. (Lancashire, 1993) . Nesta mesma linha Walter Ong discute a maneira como a escrita distancia o autor e do seu pensar. Ong indica as características da linguagem do pensamento de um ator. *Elaboramos* sobre o texto de

[16] Walter Ong (Ong, 1988) para indicar o que entendemos ser **as características básicas da linguagem do pensamento**, onde o autor organiza sua narrativa, antes de sua transposição para a linguagem 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 o sentido com frases pequenas e com palavras curtas, formadas por 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 expressá-las com uma relativa proximidade das vivências do cotidiano.

Por outro lado à linguagem **de edição ou da escrita** tem também características delineáveis:

- é explícita, formal, de padrões normativos e de procedimentos formalizados; usa uma prática de intentos pré-elaborados;
- procura eliminar as repetições das expressões e conceitos;
- procura eliminar a redundância e palavras indeterminadas;
- utiliza figuras de linguagem para facilitar a agilidade e a fixação das expressões;

- é rica nas relações entre palavras, e não elabora só com a realidade sensível; é portanto rica em metáforas; possui uma grande fluência de termos, usa expressões peculiares de um contexto informacional e pode possuir 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, que podem ser simples ou complexas e com uma grande liberdade de elaborar significados.
- é uma linguagem morfológicamente coerente e passível de ter alguma definição de padrões e procedimentos.

Assim 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 analisar e monitorar a informação. O *quantum* de informação disponível é abundante; cabe pois, uma articulação estratégica para conhecer, avaliar e filtrar a informação relevante [18] e prioritária [19] passível de ser mais facilmente transformada em conhecimento, de ser controlada e avaliada para atingir um fim específico. Uma política de monitoramento da informação, por compressão semântica, tem por base a seguinte articulação teórico-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 freqüentes*, 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 modifica-se com maior rapidez em função do tempo;

- As palavras que indicam a relevância de um texto, são as de maior de maior frequência e situam-se dentro de um conjunto (cluster) de palavras com uma frequência limite. A partir deste limite a relação frequência de ocorrência da palavra e relevância apresenta rendimentos decrescentes de escala, em relação à relevância. Existe uma condição de rendimentos decrescentes em escala indicando que, quanto menor a frequência da palavra no texto menor, também, será a importância daquelas palavras para o conteúdo do texto considerado;

Assim a indicação da frequência limite, para explicar a relevância do texto, pode ser determinado pelo coeficiente de relevância [22] abaixo indicado, que é calculado para cada conjunto de palavras de uma mesma frequência. Por exemplo: A formula permite indicar que na frequência 10, 80 % das palavras são relevantes, e a partir daí haverá um continua perda na taxa de relevância, dos conjuntos subsequentes, sendo assim, a frequência de palavras aparecendo 10 vezes no texto será a frequência limite para estudo daquele conjunto específico de documentos;

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

- As palavras de frequência igual a um se agregam em *clusters* de significância . Estes clusters são delimitados pelo tamanho das palavras, (em número de letras) contidas naquele cluster. Por exemplo: cluster de palavras de 20 letras, de 18 letras, de 12 letras, e assim em diante. Existe um rendimento decrescente em escala [23] da importância, das palavras, destes clusters. Isto é quanto menor o tamanho das palavras menos importante será o cluster. A formula do coeficiente de relevância acima indicado é uma opção de trabalho, para o analista determinar o do cluster limite para o trabalho;
- Existe uma relação positiva entre o tamanho de uma palavra de um texto e a sua qualidade de relevância e de prioridade. Isto é palavras maiores são mais significantes;
- As configurações de zonas de qualidade intensa, com relevância e prioridade podem ser determinadas automaticamente por agentes de software, sem interferência humana. desde que definidos em algoritmos os parâmetros indicados;

Os preceitos operacionais básicos , validados em pesquisa realizada, agregado ao agente de software formam um instrumental de monitoramento, gestão e controle da informação de enorme poder de análise de grandes quantidades de textos em um assunto específico.

DESENVOLVIMENTO DE UM SOTWARE 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 tipo *freeware*: são elaborados para um estudo específico de um determinado pesquisador, e não atendem a necessidade de um outro estudo específico; b) seu desenvolvimento se direciona para a língua em que foi programado, não reconhecendo particularidades do português como acentuação e cedilha, por exemplo; c) grande parte destes instrumentos foi projetada para análise literária, como poesia e ficção e não suportam corretamente a literatura de ciência e tecnologia.

Foi indispensável construir um software específico para o desenvolvimento da pesquisa.

PROCEDIMENTOS OPERACIONAIS:

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

A análise mostrou a validade de todo o raciocínio operacional anteriormente citado. Cuidado especial foi dado às palavras de frequência 1 que, indicam a individualidade de um texto, seus métodos de transformação, 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 para a prioridade do texto considerado; 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	α	
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.

α = 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	media	media	
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 uma progressão decrescente ao tamanho das palavras, medido pelo seu número de letras. A importância das palavras nos clusters tem uma evolução como em uma progressão decrescente.

Os Clusters evoluem 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 número de 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 DAS 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 -revocação de palavras da memória com grande liberdade conceitual - pensamento divergente na recongnição dos conceitos

<p>INTERMEDIÁRIAS</p> <p>Tipo: sociais aplicadas (com um aporte do discurso tecnológico)</p>	<p>-código mais formalizado e visível</p> <p>-condições semânticas controladas</p> <p>- discurso semitécnico</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>-revocação de conceitos da memória com orientação ao assunto</p> <p>- predomina o pensamento orientado a reconhecimento convergente dos conceitos na memória</p>
<p>DURAS</p> <p>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 às figuras de linguagem</p>	<p>-pouca fluência de idéias e conceitos</p> <p>-precisão na revocação dos conceitos da memória</p> <p>-pensamento altamente convergente na revocação das palavras</p>

Conceito [\[24\]](#); Pensamento Convergente [\[25\]](#); Pensamento Divergente : [\[26\]](#)

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 um simples algoritmo de contagem de palavras em um texto ou em 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 um diagnóstico específico.

Trabalha-se com este modelo para verificar desde cenários de condições estratégicas da informação aos mais diversos tipos de monitoramento; a gestão da informação para estabelecer zonas de qualidade intensa [\[27\]](#) em um determinado contexto informacional é uma das qualidades direcionadas a um grupo de usuários de um mesmo contexto informacional.

Do ponto de vista da gestão e monitoramento de seus estoques a informação relevante e prioritária pode ser agregada em um mesmo espaço. Na elaboração de uma política de compressão cabe ao administrador do estoque a responsabilidade de selecionar indicadores representativos para esta compactação, que pode elaborada em diferentes níveis de automação.

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

ZONAS DE QUALIDADE DA INFORMAÇÃO

Com a mesma intenção tem-se a tabela 2X2 ponderando relevância e prioridade

TABELA 2X2

A zona de excelência seria a zona marcada, que poderia ainda ser fragmentada em diferentes níveis de qualidade. A zona dos não relevantes e não prioritários seria descartada para àquele receptor ou grupo de receptores semelhantes. As demais zonas necessitam de um trabalho de identificação estratégica junto ao grupo ou de estratégia de monitoramento.

A zona relevante/não prioritários 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 dos prioritários /não relevantes, ou pode indicar que existem informações sem utilidade 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, e uma melhor apropriação visando o conhecimento.

Esta seria 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; melhoramentos para o receptor e sua qualidade de interatuação com o conjunto de documentos. Interatuação, apropriação, conhecimento.

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

BIBLIOGRAFIA

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 à geração do conhecimento, Relatório apresentado ao CNPq em fevereiro de 2000. Publicado, ainda, 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 1993 (Relatório Apresentado ao CNPq)
4. Barreto, A de A, A Informação e o Cotidiano Urbano, IBICT, 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 Pres, 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, Knowledge, 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, American 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 Jacobson.
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. Mannhein, K , Conhecimento e Sociedade, em Sociologia,
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

NOTAS

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

[2] o presente artigo é fruto de pesquisa realizado com auxilio 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 –um método de análise nunca de redução ou de substituição de conteúdos; a ação ou efeito de um 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 explicar o 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 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 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 – uma intenção com vigor de propósitos de acabamento; 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, julgamento do valor, percepção, pensamento e memória. 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 para um determinado sujeito, isto 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, série, classe, importância

relativa quando da prática de alguma coisa. Em nossa análise a prioridade se associa à 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 , ∞) anos .

[22]

No Curso de Ciência da Informação no Rio de Janeiro. m 2001, foi oferecida 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 e referenciamos.

[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 esta produzindo.

[24]

: a menor unidade de uma estrutura significante, com condições representacionais; para nos o mesmo que **uma palavra**

[25]

: aquele em que a seleção das palavras do texto se direciona a uma cadeia de ligações precisa, determinada, convencional, pontual

[26]

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 propriedade de por sua singularidade, seu viés.

[28]

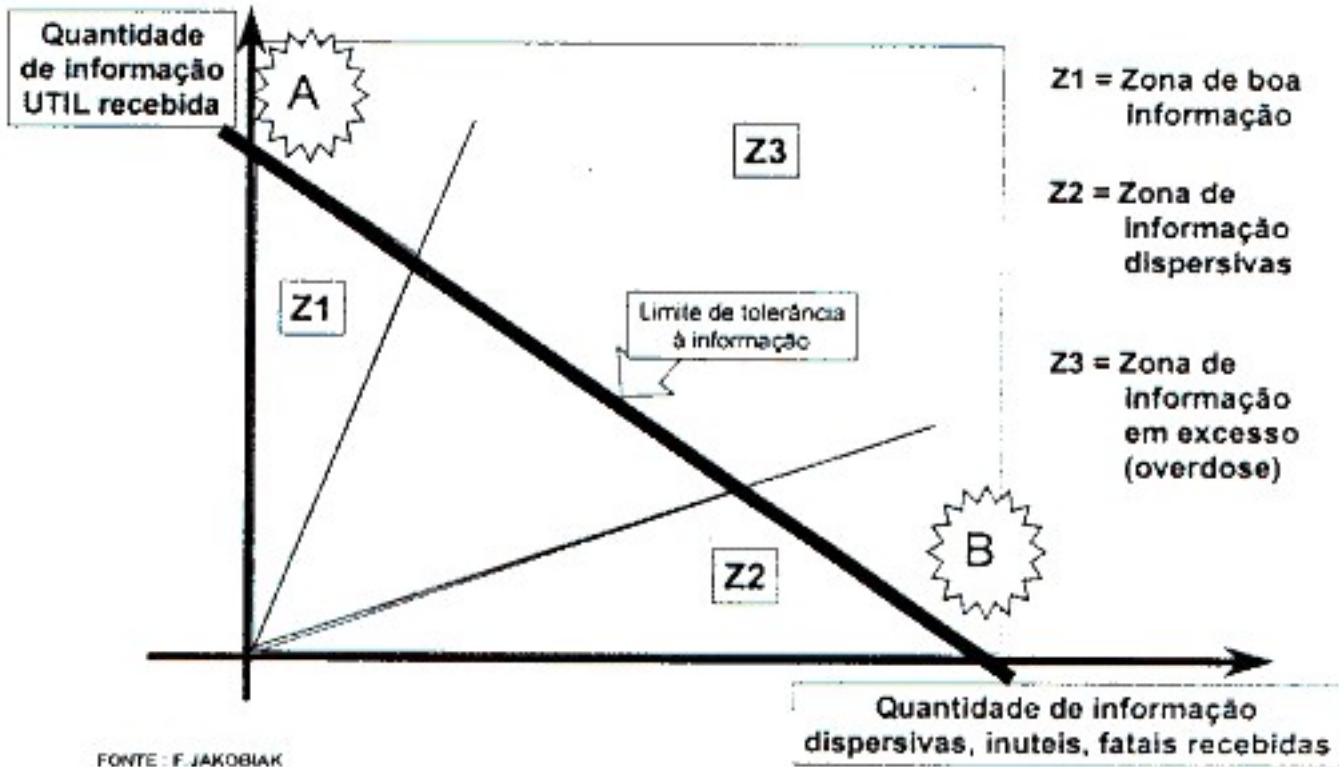
JAKOBIAK F Maîtriser l'Information Critique, Les Editions Organization, Paris, 1988

[29]

Tensão cognitiva : relativa a uma situação de comunicação em que ocorre uma freqüente situação de exposição a grandes volumes de informação que necessitam ser avaliadas para a escolha de uma opção que, seria a mais útil para o receptor

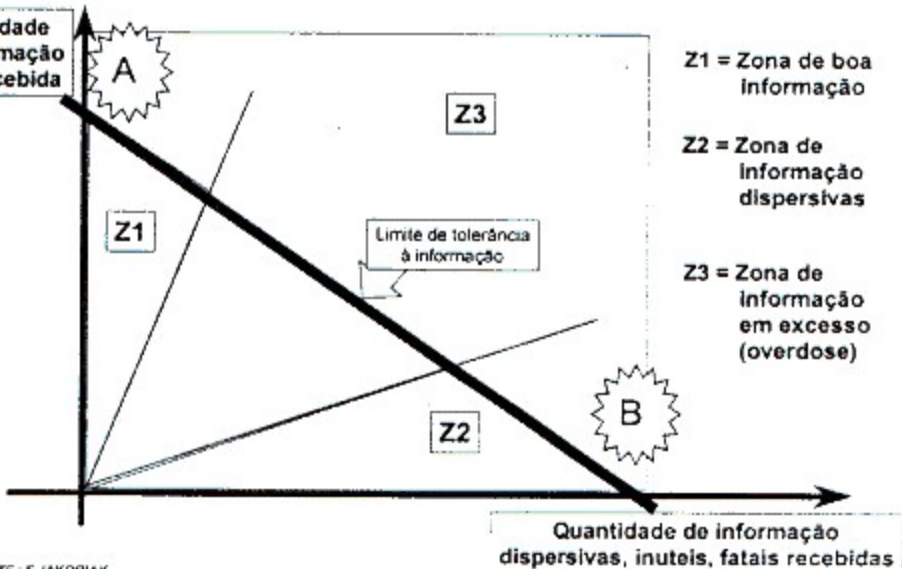


$$\textit{Coeficiente} = \frac{\sum \textit{Palavras.relevantes}}{\sum \textit{Palavras.cluster}} \times 100$$



FONTE : F. JAKOBIAK

Quantidade de informação UTIL recebida



Z1 = Zona de boa informação

Z2 = Zona de informação dispersivas

Z3 = Zona de informação em excesso (overdose)

Quantidade de informação dispersivas, inúteis, fatais recebidas

RELEVÂNCIA X PRIORIDADE

RELEVÂNCIA

P
R
I
O
R
I
D
A
D
E

Documentos
Relevantes

Documentos
Prioritários

Documentos
Relevantes

Documentos NÃO
Prioritários

Documentos NÃO
Relevantes

Documentos
Prioritários

Documentos NÃO
Relevantes

Documentos NÃO
Prioritários

RELEVÂNCIA X PRIORIDADE

RELEVÂNCIA

**P
R
I
O
R
I
D
A
D
E**

**Documentos
Relevantes**

**Documentos
Prioritários**

**Documentos
Relevantes**

**Documentos NÃO
Prioritários**

**Documentos NÃO
Relevantes**

**Documentos
Prioritários**

**Documentos NÃO
Relevantes**

**Documentos NÃO
Prioritários**