

A Gestão do Conhecimento na World Wide Web: reflexões sobre a pesquisa de informações na rede

Niraldo José do Nascimento*
Jorge Tadeu de Ramos Neves **

Pesquisa na Internet com o objetivo de identificar os principais sites sobre o tema da Gestão do conhecimento, visando contribuir e servir de referência para estudantes e pesquisadores. Discute os conceitos, importância e aplicações da gestão do conhecimento nas organizações e discorre sobre pesquisas de outros autores na World Wide Web, assim como as metodologias e dificuldades encontradas. Disponibiliza, em forma de produto, uma seleção de 28 sites, considerados os mais relevantes tendo em vista a metodologia adotada, com comentários sobre seus conteúdos e os respectivos endereços (URLs).

“Online searching is like gold-washing” - Irene Wormell, Ph.D.

29

Introdução

A citação de WORMELL¹ (1998) é, na verdade, o título de um *paper* apresentado no *Online Information Scandinavia 98*², evento realizado durante o *Stockholm International Fairs*³, de 12 a 14 de maio de 1998. Tratou-se de uma conferência e encontro de negócios que procurou discutir todos os aspectos do conteúdo de publicações digitais, assim como os serviços de publicações eletrônicas e as tecnologias associadas.

O trabalho de Wormell foi um, dentre muitos outros apresentados, que tratou o problema da informação em seus vários aspectos e as tecnologias correlatas a seu tratamento, principalmente no que se refere à sua recuperação. À medida em que cresce explosivamente o volume de informações, proliferam as pesquisas, estudos e debates sobre a melhor forma de se conseguir extrair o que é realmente necessário e importante para um determinado objetivo. A citação da autora e expressões semelhantes como *data mining*⁴ dão uma idéia precisa das dificuldades encontradas por organizações e indivíduos na busca de informações e conhecimentos.

Este estudo procura discutir algumas dessas questões tomando por base uma pesquisa na *World Wide Web* – WWW cujo objetivo inicial era construir um *bookmark*⁵

* Mestrando do Programa de Pós-Graduação em Ciência da Informação da UFMG

** Professor Doutor da Escola de Biblioteconomia da UFMG

¹ Irene Wormell é chefe do Centre for Informetric Studies (<http://www.db.dk/cis/>) da Royal School of Library and Information Studies em Copenhague, Dinamarca.

² <http://www.exponova.se/online/english/index.html>

³ <http://www.stofair.se>

⁴ Data Mining é um conjunto de técnicas para descobrir padrões e regularidades em grandes conjuntos de dados. Tem sido utilizado por empresas para comporem perfis de consumidores e definirem estratégias de marketing.

⁵ Relação de endereços de páginas da Internet.

informações disponíveis, quer sejam apresentadas sob a forma de documentos e livros, ou na imensidão crescente de páginas de hipertexto da *Web*. Nesse sentido, é muito comum encontrar páginas de *bookmarks* organizadas por estudantes, pesquisadores, profissionais das mais variadas áreas e instituições diversas na WWW, remontando aos inícios da Internet, quando o compartilhamento de conhecimentos era sua premissa máxima. A construção desses *bookmarks* é uma tarefa árdua, na maioria das vezes resultado do simples ato de navegar pelas páginas sem uma metodologia específica para esse fim, ou do contato entre pessoas com interesses comuns. Um *bookmark* é inegavelmente um *mapa de conhecimento*, podendo, inclusive, apontar para pessoas, uma vez que grande parte das páginas oferece um endereço de correio eletrônico para podermos nos comunicar com o autor. Pode também apontar para fóruns de discussão (*chats*) onde é possível trocar informações e experiências sobre um assunto com diversas pessoas ao mesmo tempo, sem o limite das barreiras geográficas. Como negar que um *bookmark* que reúna, por exemplo, *links* para bases de dados *online* de patentes em diferentes países não seja uma fonte potencial de conhecimento tecnológico? Endossa-se, por outro lado, a afirmação dos autores de que:

“Um mapa do conhecimento – seja ele um mapa real, páginas amarelas do conhecimento ou um banco de dados sofisticado – indica o conhecimento, porém não o contém”.

2 Pesquisas na WWW

As pesquisas descritas a seguir foram incluídas no sentido de oferecer sustentação teórica à metodologia aplicada neste trabalho, identificar diferentes possibilidades de análises da WWW e, também, pontos comuns entre as experiências vividas pelos diferentes autores.

LARSON (1996), da *School of Information and System, University of California, Berkeley* – UCB, analisou, através de um software desenvolvido especificamente para esse fim, mais de 30 *gigabytes* (isto é, unidades de 1.000.000.000 de *bytes*) de páginas *Web* coletadas pelo programa *web crawler*, responsável pela indexação do *search engine Inktomi* da UCB. Também utilizou o motor de busca AltaVista para análise de co-citação na WWW relativo à disciplina *Ciência da Terra (Earth Science)*. Larson não descobriu uma maneira de mapear a estrutura intelectual da WWW (ou seus subdomínios com suas disciplinas particulares) além do uso mais que tradicional da categorização manual, como nos serviços disponibilizados pelo Yahoo. Contudo, ainda que esses serviços mostrem muito da WWW, parece não haver tentativas em aplicar análises para citação e co-citação⁸ no universo da rede de hipertexto. O autor considera a noção de citação e co-citação como fundamental, já que são o mecanismo primário de conexão e acesso à informação no ciberespaço. Eles podem prover um mapa da estrutura intelectual de uma disciplina, mostrando grupos significativos de autores relacionados. A tese do autor é de que, se assumirmos que a WWW é um protótipo das bibliotecas digitais do futuro, o conhecimento das técnicas desenvolvidas

⁸ Em termos da WWW, co-citação diz respeito aos links existentes entre páginas Web, isto é, quando uma ou mais páginas referenciam-se entre si. Citação implica na existência de um link para uma outra página.



para análise da estrutura intelectual nas bibliotecas convencionais pode ser muito útil na transição para um ambiente baseado em redes. A metodologia usada na pesquisa exploratória é, fundamentalmente, idêntica àquelas usadas no mapeamento de co-citação de autores e jornais científicos. Os passos utilizados foram os seguintes:

- Seleção do conjunto base de itens para o estudo;
- Recuperação das freqüência de co-citação para os itens;
- Compilação das freqüências de co-citação em uma matriz primária;
- Análises de correlação para converter a matriz primária de freqüências em coeficientes de correlação;
- Análises multivariadas da matriz de correlação, utilizando, principalmente, análises de componentes, análises de grupos ou técnicas multidimensionais (MDS – *Multidimensional Scaling*⁹).
- Interpretação e validação do mapa resultante.

Para seleção do conjunto base de sites WWW foi utilizado o AltaVista, sendo o foco da pesquisa direcionado para Sistemas de Informações Geográficas, Ciências da Terra e sensoramento remoto por satélites.

Estas áreas foram escolhidas em função da familiaridade do autor com os assuntos e também porque a localidade mais referenciada, através de análises do Inktomy, era o *Xerox PARC map browser*¹⁰. Para limitar o número inicial do conjunto de itens, a pesquisa submetida ao AltaVista (utilizando o modo de pesquisa avançado) foi “link:pubweb.parc.xerox.com/map” AND “link:xtreme.gsfc.nasa.gov”, ou seja, desejava-se formar um conjunto de documentos contendo *links* tanto para o *Xerox Map browser*, como para a *home page* da NASA’s AVHRR (*Advanced Very High Resolution Radiometer*) e seus projetos de sensoramento remoto.

O resultado da pesquisa inicial apresentou um conjunto de 115 páginas da WWW contendo todos ou a maioria dos elementos. Esse conjunto foi analisado e as páginas aparentemente mais relevantes foram armazenadas para futuras análises, obtendo-se uma nova seleção, agora reduzida a 43 páginas. Destas, muitas eram páginas *bibliográficas* de serviços como Yahoo ou mantidas por indivíduos interessados nos tópicos da pesquisa. Todos os *links* dessas 43 páginas para outras páginas foram extraídos e combinados em um único arquivo, contendo 7209 URLs distintos. Esses URLs foram classificados em ordem alfabética e eliminados os que apareciam em menos do que 3 dos 43 documentos citados. Também foram eliminadas as citações que pareciam estar fora dos limites dos tópicos abordados. O resultado foi um conjunto de 332 URLs, candidatos potenciais ao estudo. Eles foram então recuperados e examinados utilizando o *browser* Netscape, sendo os sites considerados apropriados reunidos em uma *hotlist*, reduzindo o número de documentos para 125. Este conjunto foi considerado ainda muito amplo, sendo novamente reduzido, através de julgamento do pesquisador e com a ajuda de vários *best of the Web*, oferecidos por algumas páginas, a 34 sites.

⁹ Técnicas estatísticas para aplicação em bases de dados. Existem vários freewares desenvolvidos por universidades, institutos de pesquisa e organizações disponíveis para download na Internet.

¹⁰ <http://pubweb.parc.xerox.com/map/>

Tendo obtido o conjunto base, o próximo passo foi produzir a matriz primária de co-citação. Este estágio exigiu que cada link fosse combinado com outro, par a par, e enviado para pesquisa no AltaVista. Para executar este trabalho, um “web robot” foi programado para submeter automaticamente as pesquisas para o conjunto base de URLs e capturar as informações sobre frequências resultantes para posterior análise¹¹. O resultado do processo foi a matriz primária de co-citações, submetida, após sua conversão em uma matriz de correlações, a um sistema denominado SAS *Multidimensional Scaling*, que fornece como resultado uma representação espacial dos dados, consistindo em uma configuração geométrica de pontos que é, em efeito, um mapa das informações em dois ou três planos dimensionais. A conclusão do autor é de que, apesar do crescimento exponencial da WWW, sua diversidade e falta de controle natural não é, contudo, uma vasta área cibernética infértil onde alguns bits aleatórios de informação podem ser encontrados através de um *browser*. Os resultados da pesquisa, através da metodologia utilizada, mostram grupos de sites WWW razoáveis e compreensíveis, com similaridades entre os tópicos. Embora seja um estudo preliminar, o autor acredita que os métodos bibliométricos funcionam quando aplicados aos *links* das páginas de hipertexto da WWW.

Chamar a atenção para o vasto potencial de informações *online*, as múltiplas e novas possibilidades que as técnicas avançadas de pesquisa oferecem hoje e aprender como usar bases de dados online, não somente para ter acesso a documentos ou fatos, mas também traçar tendências e desenvolvimentos da sociedade e disciplinas científicas, é o objetivo do *paper* de WORMELL (1998). A autora, contudo, faz um alerta um tanto enfático: “Este tipo de informação em bases de dados é visível somente para pesquisadores inteligentes e para aqueles que aprenderam a ler nas entrelinhas das publicações eletrônicas.” Como exemplo, ela cita três análises executadas pelo *Centre for Informetric Studies – CIS*, em Copenhague.

A primeira delas, *How ‘international’ are the international journals?*, procurou mapear a internacionalidade de revistas científicas no contexto da distribuição geográfica de seus autores, referências bibliográficas e assinantes. Uma segunda denominada *Trend analysis in food science research*, teve por objetivo determinar quais áreas de pesquisa em agricultura e produção de conhecimento deveriam ser intensificadas a fim de melhorar a posição estratégica das exportações dinamarquesas no mercado mundial. Em ambas foram utilizadas, em larga escala, bases de dados de referências bibliográficas, que, segundo a autora, permitem ao pesquisador não só expandir o escopo da pesquisa a partir da recuperação de trabalhos citados em um trabalho base, como também identificar tendências e fenômenos (baseados, por exemplo, no crescimento da frequência de um termo) que podem impactar domínios da produção científica ou áreas de aplicações tecnológicas industriais. A terceira análise desenvolvida denominou-se *Informetric analyses on the World Wide Web: methodological approaches to Webometrics*.

¹¹ O autor observa que o programa web robot foi instruído para ser educado e realizar uma pausa após cada pesquisa, evitando monopolizar os serviços do AltaVista.



Os pesquisadores do CIS examinaram a idéia de utilizar métodos informétricos¹² na WWW, iniciando por estabelecer as bases da nova e emergente área, denominada *webometria*.

O trabalho, de autoria de ALMIND & INGWERSEN (1997), introduz a aplicação de métodos informétricos na WWW, o que foi feito através de um estudo de caso, analisando-se a proporção da *Web* dinamarquesa com as de outros países nórdicos e descrevendo um número específico de parâmetros de análises informétricas. Os resultados demonstraram que a Dinamarca corria sérios riscos de ficar atrás dos outros países nórdicos no que diz respeito à *visibilidade* na *Internet*.

Para esses autores, enquanto em outras bases de dados os descritores são encontrados em formas de autor, palavras-chave etc, na WWW os descritores são por autor ou por frequência. As páginas da Internet são assim as entidades de informação da *Web*, com *hyperlinks* entre elas atuando como citações. Como as páginas são construídas com a utilização de *tags*¹³, estas podem servir de referência para a pesquisa de assuntos ou autores. Por exemplo, através das *tags* <TITLE> e/ou <H1> é possível identificar o principal assunto de uma página e (itálico e negrito, respectivamente) também servem de referência uma vez que são utilizados pelos autores para enfatizar determinados termos ou frases de suas *Web pages*. A fonte corporativa de uma página pode ser identificada através do URL, contudo, nem sempre a instituição que a está hospedando guarda alguma relação com o seu autor. ALMIND & INGWERSEN (1997) alertam que os problemas em utilizar análises informétricas na WWW são idênticos aos encontrados nas análises das bases de dados de referências bibliográficas, tais como estruturas restritas de arquivos e falhas na validação dos dados.

Afirmam, também, que como a *Webometria* refere-se a pesquisas na WWW (usando a *informetria* ou outras medidas quantitativas) certos problemas relativos às análises são específicos a esse tipo de abordagem. Um deles é que *só é possível realizar análises assíncronas* devido à natureza dinâmica e em tempo real da *Web*. Além disso, por ser um sistema baseado em hipertexto, os dados não são estruturados, não existindo controles ou exigências para o uso dos códigos de marcação. A enorme quantidade de informações disponíveis implica em dificuldades de encontrar exatamente o que se necessita. A solução para este problema está em utilizar as bases de dados indexadas da WWW (catálogos e motores de busca) o máximo possível.

Por outro lado, criticam os autores, essas base são muito mistas e erráticas. Ainda que algumas delas venham se tornando mais profissionais, no sentido de

¹² A *informetria* abarca o estudo dos aspectos quantitativos da informação independentemente da forma com que apareça registrada e do modo como é gerada. Considera, além disso, os aspectos quantitativos da comunicação informal ou oral, do mesmo modo que as registradas e tem em conta as necessidades e usos da informação para qualquer atividade, seja ou não de caráter intelectual. Por outro lado, pode incorporar e utilizar diversos meios de medição da informação, que estão fora dos limites da *bibliometria* e da *cienciometria*.

A *bibliometria* engloba o estudo dos aspectos quantitativos da produção, disseminação e uso da informação registrada, a partir dos quais desenvolve modelos e medidas matemáticas, que servem para fazer prognósticos e tomar decisões a respeito de tais processos.

O termo *cienciometria* se define como o estudo dos aspectos quantitativos da ciência como disciplina ou atividade econômica. É parte da sociologia da ciência e tem aplicação no estabelecimento das políticas científicas e inclui, entre outras, as publicações, motivo pelo qual se imbrica, em certa medida, com a *bibliometria*.

O valor do modelo informétrico reside em sua capacidade de resumir, em termos de uns poucos parâmetros, as características de muitos grupos de dados, assim como a possibilidade que oferece de estabelecer prognósticos sobre tendências futuras e de determinar o efeito de diferentes fatores nas variáveis de interesse. Dessa forma, além das medidas que derivam dele, o modelo informétrico oferece uma base sólida para a tomada de decisões práticas. (TAGUE-SUTCLIFFE, 1994).

¹³ Mais adiante, observa-se que o help do AltaVista faz referência à utilização dessas *tags* em seus processos de pesquisa.

cobrirem amplas áreas de estudo e de oferecerem facilidades razoáveis de pesquisa, o maior problema continua sendo o tamanho e a natureza dinâmica da WWW. Denominam algumas dessas bases de índices monolíticos, por tentarem indexar praticamente a *Web* inteira, o que implica um conjunto de fontes excessivamente extenso e sujeito a freqüentes alterações. Para eles, uma estratégia mais apropriada seria a criação e distribuição de índices, como o padrão utilizado nos países nórdicos. Cada um dos países escandinavos mantém um índice compartilhado de suas bases em WWW.

O único requisito utilizado para a pesquisa era de que o URL da entidade de informação, via protocolo HTTP, deveria ser da Dinamarca. A identificação foi feita checando se o URL continha *.dk* no sufixo do servidor. Para assegurar o nível mais alto possível de retorno na identificação das páginas que preenchessem esse parâmetro, diferentes *search engines* foram utilizados, o que é uma forma de polirepresentação, na qual diferentes fontes oferecem diferentes resultados cognitivos e funcionais, bem como as representações dos objetos para os quais a pesquisa está sendo realizada. A desvantagem desse método é que os resultados das representações dos documentos são muito individuais e não contêm os mesmos elementos de dados. Os índices utilizados foram o Lycos e o OpenText, julgados por várias fontes como os mais indicados, na época em que a pesquisa foi realizada. Examinar páginas conhecidas da *Web* para procurar por outras é uma tarefa laboriosa e que consome muito tempo, advertem os autores, mas recomendada no sentido de completar outros resultados de pesquisa.

3 Metodologia

A metodologia inicial adotada para o mapeamento consistiu em lançar no AltaVista, o termo *knowledge+management* na caixa padrão de pesquisa (*standard search box*) que, segundo o próprio AltaVista, “quase tudo que você precisa pesquisar pode ser encontrado rapidamente e com melhores resultados usando a caixa padrão de pesquisa, na qual o serviço de pesquisa do AltaVista classifica os resultados apresentando os mais relevantes primeiro.” (AltaVista Help).

A opção de colocar o operador (+) foi no intuito de estender a pesquisa, uma vez que este formato significa solicitar ao motor de busca que procure pelo termo *knowledge* e que os documentos encontrados contenham também o termo *management*. Desse modo, termos relacionados à gestão do conhecimento, como *knowledge garden* e *knowledge ecology* também poderiam ser resgatados.

Optou-se também por pesquisar todos os primeiros duzentos documentos trazidos como resposta, que é o número máximo de documentos disponibilizados na forma básica de pesquisa do AltaVista. Acreditava-se ser uma amostra suficientemente significativa, uma vez que os resultados são organizados em ordem de importância pelo mecanismo de busca, ainda que essa *importância* seja relativa¹⁴.

¹⁴ O help do AltaVista não fornece muitos detalhes a respeito da classificação dada aos documentos. No entanto, uma das FAQ's (Frequently Asked Questions), para desenvolvedores na Web, pergunta como fazer para que determinado Web Site fique bem classificado nos resultados. A resposta indica que este mecanismo de pesquisa utiliza-se das tags do documento em html, como <TITLE> (título da página), <H1> (título de nível 1), (texto em itálico), (texto em negrito) etc, para fazer suas classificações. São desaconselhados artifícios como a repetição de um mesmo termo muitas vezes na página ou o uso excessivo de palavras chave.



Cada um dos documentos listado foi visitado, fazendo-se uma avaliação de seu conteúdo. Para cada documento, foi feito ainda, o *encadeamento para trás (chaining)*, ou seja, buscou-se as origens de cada documento procurando identificar o responsável por sua elaboração ou disponibilização na WWW, tornando-se esta a primeira categoria da pesquisa.

As categorias analisadas foram:

- *Identificação do site hospedeiro*¹⁵:
 - ✓ *Comercial (COM) – Empresas e organizações comerciais;*
 - ✓ *Acadêmico (EDU) – Universidades, faculdades e entidades ligadas diretamente ao meio acadêmico;*
 - ✓ *Organizações em geral (ORG) – Organizações sem fins lucrativos, associações etc.*
- *Pertinência das informações contidas nos sites como tema knowledge management*:
 - ✓ *Pertinente (P) – o documento é totalmente pertinente ao tema;*
 - ✓ *Pertinência relativa (PR) – o documento é mais ou menos pertinente ao tema (Ex.: informações sobre um software para gestão de conhecimentos);*
 - ✓ *Não pertinente (NP) – as informações não guardam nenhuma relação ao tema.*
- *Quantidade e qualidade das informações disponibilizadas (existência de papers ou white papers, links para outros sites sobre knowledge management – KM, serviços disponíveis etc.):*
 - ✓ *Bom (B) - há muita informação e são de boa qualidade;*
 - ✓ *Regular (R) – há pouca informação, mas têm boa qualidade;*
 - ✓ *Insuficiente (I) – há pouca ou nenhuma informação.*

Tem-se conhecimento de uma vasta literatura a respeito de avaliação de páginas Web, gerada principalmente por educadores e bibliotecários. Segundo MOSTAFA & TERRA (1998), “essa enorme massa de literatura de avaliação de fontes eletrônicas abrangem em maior ou menor grau os cinco critérios de avaliação de fontes impressas tão conhecidos dos bibliotecários: acuidade, autoridade, objetividade, atualização e cobertura”. Para essas autoras, somam-se ainda outros elementos como o nível de interatividade da página, já que a Internet é um ambiente interativo. Considerando-se os objetivos deste trabalho e o volume de consultas a ser realizado, decidiu-se simplificar ao máximo as categorias de análise mas, procurando, por outro lado, manter um referencial básico que permitisse uma avaliação objetiva em um segundo momento.

A relação dos sites obtida foi previamente transferida para o processador de textos Word, em forma de tabela, a fim de facilitar a documentação das observações sobre cada documento. O próximo passo foi eliminar os documentos considerados inadequados tomando por base as categorias de análise. Excluiu-se, dessa forma, os documentos considerados *não pertinentes* e aqueles em que as informações disponíveis eram insuficientes (I). O resultado, contudo, foi considerado insatisfatório.

¹⁵ Apesar do critério de abreviaturas utilizado ser idêntico ao utilizado na identificação dos sites da Internet, alguns dos endereços visitados não continham no URL essas siglas, sendo feita, então, uma classificação pessoal.

usados: a pesquisa por palavras-chave e a pesquisa baseada em conceitos, sendo o primeiro mais comum. Na pesquisa por palavras-chave, se o autor do documento não as define (utilizando *metatags* da linguagem HTML) o mecanismo de busca irá fazê-lo, procurando por palavras no início do documento e palavras que se repetem várias vezes, eliminando, é claro, artigos, conectivos e outros termos. Os maiores problemas apresentados são a incapacidade de analisar o contexto em que as palavras estão sendo usadas e variáveis como tempos verbais, singular e plural etc.

Existem vários métodos para a pesquisa baseada em conceitos (muito utilizada pelo Excite). A idéia consiste em analisar termos próximos a uma determinada palavra. Por exemplo, se a palavra *coração* está próxima de outras como coronárias, sangue, artérias, colesterol etc., o mecanismo irá enquadrar a página na categoria de *medicina*. Se, por outro lado, os termos próximos a coração são flores, amor, paixão, namoro etc. o documento poderá ser enquadrado na categoria *romance*. De qualquer modo, é muito comum encontrar documentos que não têm nenhuma relação com o termo pesquisado. No caso desta pesquisa, o termo *knowledge* parece ter contribuído para acentuar esse problema, uma vez que é utilizado em muitos e diferentes contextos.

Com relação ao caso citado no tópico Metodologia, de se optar em lançar no AltaVista os termos *knowledge+management* ou *knowledge management*, uma pesquisa posterior, praticamente simultânea, revelou não haver a menor diferença entre as duas opções, uma vez que a comparação dos dois resultados mostrou exatamente os mesmos documentos na mesma ordem de classificação. Isso se deve, provavelmente, ao extenso número de documentos relacionados ao tema, não sendo possível afirmar que com a utilização de outro termo, ou em outro momento, se alcançasse o mesmo resultado.

Comparação entre os resultados apresentados pelos diferentes mecanismos de busca

As listas de documentos apresentadas por cada um dos mecanismos foram transferidas para o Word, classificadas e comparadas entre si, para verificar ocorrência de referências a um mesmo documento.

O resultado pode ser observado na FIG. 3 ao lado:

A primeira comparação, realizada entre as duas pesquisas feitas no AltaVista, reflete a própria dinâmica da Web sobre o assunto pesquisado, demonstrando que em um período aproximado de 10 meses, 71% dos sites relacionados foram substituídos por outros, a princípio, mais relevantes. Outro fator que chama a atenção é o bom desempenho na comparação entre os mecanismos de busca Copernic 98 e Metaminer. O percentual de documentos que coincidiram (13%) é relativamente significativo se comparado com os demais, tendo contribuído para melhorar o produto final apresentado.

Conclusões

Como já foi relatado, na primeira fase foram analisados individualmente cada um dos 200 documentos relacionados na pesquisa do AltaVista. É impossível não recorrer mais uma vez à citação de WORMELL(1998), *Online searching is like gold-*

Mecanismos de Busca Comparados	Sites que coincidiram	Percentual	Documentos que coincidiram	Percentual
AV1 vs AV2	58	29%	15	8%
AV2 vs Metaminer	12	14%	3	4%
AV2 vs Copernic	21	15%	5	4%
Copernic vs Metaminer	15	21%	9	13%
AV2 vs Metaminer vs Copernic	9	10%	0	0%

FIGURA 3 – Comparação de resultados dos mecanismos de busca utilizados

Observações:

- AV1 – Primeira pesquisa realizada no Altavista
- AV2 – Segunda pesquisa realizada no Altavista
- Os percentuais foram calculados tomando-se por base as médias aritméticas dos percentuais individuais de cada mecanismo de busca.

washing, quando, no resultado final, obtém-se uma relação com 36 *pepitas*, sendo que muitas delas vêm acompanhadas de algumas *impurezas*. Na segunda fase, foi possível depurar mais o processo, sendo a relação de sites reduzida para 28, apresentando, porém, maior qualidade. O esforço despendido nessas análises é muito grande e apresentam-se outras dificuldades que colaboram para que este tipo de pesquisa se torne ainda mais árduo. Falta de objetividade nos documentos, linguagem informal, páginas órfãs (dificuldade em localizar o link da página principal para a página referenciada), ausência de uma definição clara dos propósitos da organização, dificuldade em localizar as páginas do site que disponibilizam artigos, *papers* e *links*, excesso de informações e/ou falta de organização no *site*, entre outras.

A conclusão óbvia que pode-se obter deste trabalho é que é perfeitamente possível construir um *bookmark* da *Web* sobre um determinado tema, utilizando-se ferramentas de busca amplamente conhecidas. Mas, existem também uma série de possibilidades e metodologias de pesquisa que podem ser reproduzidas ou criadas para esses objetivos. Durante a preparação para a pesquisa, tinha-se conhecimento de que muitos *bookmarks* são construídos a partir de navegações pela Internet, de indicações de outros pesquisadores, da apropriação de outros *bookmarks* já existentes, enfim, através da coleta não estruturada de informações. Outros mecanismos de busca igualmente famosos como o Yahoo, o Excite, o Infoseek etc, não eram ignorados, contudo, sabia-se que o AltaVista detinha uma das maiores bases de dados sobre os mais diferentes assuntos e que, provavelmente, poderia oferecer os melhores resultados.

Foi a partir dessa premissa que o projeto de pesquisa foi elaborado, ou seja, a coleta estruturada de informações sobre gestão do conhecimento. Porém, durante o desenrolar dos trabalhos, tomou-se conhecimento da existência do Copernic 98, um agente de busca que não tem bases próprias, mas pesquisa em vários canais trazendo, supostamente, os melhores resultados. Metodologia semelhante é utilizada pelo MetaMiner, parte de um conjunto de agentes Web chamado *Familia Miner*, desenvolvido no LATIN - Laboratório para Tratamento da Informação, do Departamento



de Ciência da Computação – DCC, da Universidade Federal de Minas Gerais, sendo igualmente agregado ao trabalho. Não estava, entre os objetivos deste trabalho, analisar em profundidade esses diferentes instrumentos e outros citados.

Existem aspectos altamente positivos associados a este tipo de pesquisa. O primeiro deles é a consolidação de um produto final, que pode servir de base para novos estudos. Na verdade, o *bookmark* sobre gestão do conhecimento é apenas uma pequena parte deste produto. Existe também a relação completa e comentada de todos os documentos visitados¹⁶ e, principalmente, a própria descrição da pesquisa, relatada neste artigo. Acredita-se que a partir de sua leitura é possível estabelecer uma série de *insights* que irão dar origem a novas pesquisas e à utilização de novos métodos, existentes ou a serem desenvolvidos.

Outro aspecto, é a incorporação de conhecimentos laterais à pesquisa. Foi dessa forma que, através de uma pesquisa acidental em um grupo de discussões de alunos do curso de MBA do Departamento de Administração da Universidade de Oklahoma, soube-se que a IBM está desenvolvendo um novo agente de pesquisas na Web denominado *Clever's hits technology*, que ordena os resultados a partir de uma medição dos *links* existentes entre eles, que em seguida são filtrados por outros agentes. De acordo com a IBM, este agente tem conseguido melhores resultados do que, por exemplo, o Yahoo, cujas bases são compiladas manualmente. Os exemplos apresentados no site da IBM¹⁷ parecem ser muito promissores.

Knowledge Management in the World Wide Web: some ideas about searching information on the net.

Presents a research project which will identify the most significant sites on the Internet about knowledge management. The concept of knowledge management, as well as its importance to and applications in organizations is discussed. Related research is reviewed, compared and analysed in regards to the methodology used and problems identified by different authors. The product of the research project will be an evaluated and commented list of the 28 most relevant sites, selected according to pre-defined criteria.

Referências bibliográficas

- ALMIND, Tomas C., INGWERSEN, Peter. Informetric Analyses on the World Wide Web: Methodological Approaches to "Webometrics". *Journal of Documentation*, v. 53, n. 4, p. 404-426, Sept. 1997.
- DAVENPORT, Thomas H., PRUSAK, Laurence. *Conhecimento Empresarial: como as organizações gerenciam seu capital intelectual*. Rio de Janeiro: Campus, 1998.
- LARSON, R. *Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace*. 1996. [online] Disponível na Internet via WWW. URL: <http://sherlock.berkeley.edu/asis96/asis96.html>. Documento capturado em 30/08/98.
- MALHOTRA, Yogesh. *Knowledge Management for the New World of Business*. [online] Disponível na Internet via WWW. URL: <http://www.brint.com/km/whatis.htm>. Documento capturado em 02/10/98.

¹⁶ Não a reproduzimos neste trabalho por limitação de espaço.

¹⁷ <http://www.almaden.ibm.com/cs/k53/clever.html>

MOSTAFA, Solange P., TERRA, Marisa. As Fontes Eletrônicas de Informação: Novas Formas de Comunicação e de Produção do Conhecimento. São Paulo em Perspectiva, Fundação SEADE, São Paulo, v. 12, n. 4, out./dez. 1998.

NONAKA, Ikujiro, TAKEUCHI, Hirotaka. Criação de Conhecimento na Empresa. Rio de Janeiro: Editora Campus, 1997.

STEWART, Thomas A. Capital Intelectual. Rio de Janeiro: Campus, 1998.

SVEIBY, Karl Erik. A Nova Riqueza das Organizações. Rio de Janeiro: Campus, 1998.

TAGUE-SUTCLIFFE, Jean. Introducion a la Infometria. Capturado em 11/11/98. Online. Disponível na Internet via WWW. URL: <http://infonew.sld.cu/revistas/aci/aci05394.htm>

WORMELL, I. Online searching is like gold-washing. Capturado em 30/08/98. Online. Disponível na Internet via WWW: <http://www.db.dk/cis/texts/artikel3.htm>.

A História da Família Miner. Capturado em 26/04/99. Online. Disponível na Internet via WWW. URL: <http://www.miner.uol.com.br/história.html>.

Ajuda e Dicas para a Família Miner. Capturado em 26/04/99. Online. Disponível na Internet via WWW. URL: <http://www.miner.uol.com.br/ajuda.html>.

AltaVista Help – Basics. online Disponível na Internet via WWW. URL: <http://www.altavista.com/av/content/help.htm>.

AltaVista Help – Advanced. .online. Disponível na Internet via WWW. URL: http://www.altavista.com/av/content/help_advanced.htm.

AltaVista Help – Refine. Online. Disponível na Internet via WWW. URL: http://www.altavista.com/av/content/help_refine.htm.

AltaVista Feedback. online. Disponível na Internet via WWW. URL: <http://www.altavista.com/av/content/questions.htm>.

Copernic – Product Info. Capturado em 08/02/99.

Copernic Product Info. Capturado em 08/02/99. online Disponível na Internet via WWW. URL: <http://www.copernic.com/plus/index.html>.



ANEXO

RELAÇÃO COMENTADA DE SITES SOBRE GESTÃO DO CONHECIMENTO NA WWW

SITE SELECIONADO	URL	Categorias
<p>@BRINT – Business, Management and Technology Research Excelente <i>site</i> sobre KM, fundado por Yogesh Malhotra e Mini Malhotra, com muitos <i>papers</i>, fóruns de discussão, <i>links</i> e referências bibliográficas. O objetivo da organização é se tornar uma referência mundial na área de conhecimento e práticas de E-Business, tecnologia e Knowledge Management, de modo a aumentar a performance organizacional das empresas. O <i>site</i> é considerado um “portal” na área de gestão do conhecimento, opinião compartilhada por várias publicações de renome.</p>	<p>http://www.brint.com/</p>	<p>Tipo: COM Pertinência: P Qlde/Qtde: B</p>
<p>Sveiby Knowledge Management Excelente <i>site</i> com muita informação disponível e interessante sobre gestão do conhecimento. Mantido por Karl-Erik Sveiby, pesquisador da Queensland University of Technology na Austrália. Conduz trabalhos relacionados ao desenvolvimento de estratégias focadas no conhecimento e implementação de sistemas para avaliação de ativos intangíveis e capital intelectual.</p>	<p>http://www.sveiby.com.au</p>	<p>Tipo: COM Pertinência: P Qlde/Qtde: B</p>
<p>KM Metazine Revista sobre KM, com artigos completos, gratuitos e interessantes. A revista é mantida pela KTI (Knowledge Transfer International) que nas palavras de seus proprietários “é uma organização unida pela crença de que no mundo dos negócios, as respostas para as questões colocadas pelas informações e o aumento da performance só podem ser obtidas pelo atendimento das necessidades humanas de conhecimento”. Vale a pena visitar.</p>	<p>http://www.ktic.com/</p>	<p>Tipo: COM Pertinência: P Qlde/Qtde: B</p>
<p>Knowledge Management Server at the University of Texas at Austin. Página do servidor sobre Gestão do Conhecimento da Universidade do Texas em Austin. Há uma série de indicações bibliográficas, <i>papers</i>, <i>links</i> para revistas e para outras organizações que tratam o tema e <i>softwares</i> para KM. É também possível, submeter <i>papers</i> para publicação no <i>site</i>.</p>	<p>http://www.bus.utexas.edu/kman/index1.htm</p>	<p>Tipo: EDU Pertinência: P Qlde/Qtde: B</p>
<p>Community Intelligence Labs A CoIL se define como uma incubadora de projetos no centro de um ecossistema de negócios composto pelos clientes, parceiros e aliados. Trata de temas como capital intelectual, comunidades virtuais e inteligência organizacional, com <i>links</i>, <i>papers</i>, entrevistas e indicações bibliográficas. Introduz o termo Knowledge Ecology e Knowledge Garden e faz uma comparação com o termo Knowledge Management. Vale a pena visitar.</p>	<p>http://www.co-i-l.com</p>	<p>Tipo: COM Pertinência: P Qlde/Qtde: B</p>

<p>ENTOvation International Ltd A missão do Entovation é promover a prosperidade de empresas em todo o mundo através da otimização dos fluxos de conhecimento. Fundado por Debra M. Amidon, é composto por uma equipe de especialistas em conhecimento. Disponibiliza artigos e <i>papers</i>, além de duas imagens muito interessantes sobre o tema. Na página http://www.entovation.com/kleadmap/index.htm há um criativo mapa apontando para “líderes do conhecimento” em diferentes países do mundo. Clicando sobre os nomes, obtém-se dados para contato com cada um deles. Faltou, contudo, explicar os critérios de seleção dos mesmos.</p>	Http://www.entovation.com	Tipo: COM Pertinência: P Qlde/Qtde: B
<p>International Knowledge Management Network Site holandês para debates e troca de idéias sobre Gestão do Conhecimento. Composto por aproximadamente 600 membros, de diferentes áreas, disponibiliza vários <i>papers</i> sobre KM e muitos <i>links</i> interessantes.</p>	http://kmn.cibit.nl	Tipo: ORG Pertinência: P Qlde/Qtde: B
<p>Business Processes Resource Centre O BPRC é em centro de pesquisas da Universidade de Warwick, no Reino Unido, que elegeu 3 maiores temas de trabalho: Complexidade, Gestão do Conhecimento e Desenvolvimento Profissional e Reestruturação de Negócios. Na página sobre Gestão do Conhecimento é possível encontrar <i>papers</i> interessantes, <i>links</i> para vários <i>sites</i> e eventos sobre o tema.</p>	http://bprc.warwick.ac.uk	Tipo: EDU Pertinência: P Qlde/Qtde: B
<p>FEND - FEDERATION FOR ENTERPRISE KNOWLEDGE DEVELOPMENT FEND é uma “federação”, sem fins lucrativos de empresas e universidades, para a colaboração em pesquisas, desenvolvimento, testes e disseminação do estado-da-arte em Gestão do Conhecimento e Inovação. Disponibiliza artigos e <i>links</i>. O <i>site</i> oferece um fórum de livre acesso para discussão sobre o tema denominado, sugestivamente, de Ágora.</p>	http://www.fend.es	Tipo: ORG Pertinência: P Qlde/Qtde: B
<p>Ernst & Young Site da empresa de consultoria Ernst & Young que se define como uma das empresas líderes mundiais em contabilidade, tributação e consultoria, distribuindo idéias inovativas para clientes locais e globais em um amplo segmento de indústrias. É possível encontrar textos interessantes sobre KM, desta organização, nos endereços: http://www.ey.com/consulting/kbb/default.asp e http://www.ey.com/knowledge/centers.asp</p>	http://www.ey.com	Tipo: COM Pertinência: P Qlde/Qtde: R
<p>MG Taylor Corporation Empresa que armazena, projeta, constrói, manufatura e opera ambientes colaborativos e de alta performance para trabalho, escritórios, casa, viagens e locais virtuais. O <i>site</i> apresenta um dos vários modelos disponíveis sobre diversos assuntos ligados à gestão de empresas e negócios na atualidade. Os modelos são muito bem feitos, através de Glyphs (hieróglifos) e totalmente explicados. Muito interessante e inovador.</p>	http://www.mgtaylor.com	Tipo: COM Pertinência: P Qlde/Qtde: B



<p>Knowledge Media Institute æ KMI O KMI Knowledge Media é patrocinado por empresas como a Apple Computers, a Sun Microsystems e outras. Abrange um amplo leque de pesquisas centradas em tecnologias de aprendizagem, incluindo colaboração através da Internet, multimídia, memória organizacional, documentos digitais, simulação, representações formais e informais de conhecimento, em suma, abordagens inovativas para compartilhar, acessar e compreender o conhecimento. Através de pesquisa interna, pelo <i>search engine</i> Excite, disponível no servidor, localizou-se uma série de <i>papers</i> interessantes.</p>	<p>http://kmi.open.ac.uk</p>	<p>Tipo: ORG Pertinência: P Qlde/Qtde: B</p>
<p>Object Knowledge Fundada em 1992, a Object Knowledge é uma empresa de consultoria na área de conhecimento. Seu principal foco são os processos de transferência do conhecimento. Não possui <i>papers</i>, mas contém uma boa coleção de <i>links</i> para <i>sites</i> sobre KM.</p>	<p>http://www.objectknowledge.com</p>	<p>Tipo: ORG Pertinência: P Qlde/Qtde: B</p>
<p>Knowledge On-Line Este <i>site</i> se define como um fórum central de informação e comunicação para Chief Knowledge Officers e trabalhadores do conhecimento. Após registro prévio, têm-se acesso gratuito a alguns pequenos artigos de renomados autores e a uma boa página de <i>links</i></p>	<p>http://www.knowledge.org.uk</p>	<p>Tipo: ORG Pertinência: P Qlde/Qtde: B</p>
<p>Collaborative Strategies Empresa de consultoria, de São Francisco (EUA), que tem como foco as tecnologias colaborativas (groupware) e a Gestão do Conhecimento. O <i>site</i> traz alguns pequenos artigos, <i>links</i> e um <i>white-paper</i> muito interessante, chamado "Electronic Collaboration on the Internet and Intranets", resultado de uma pesquisa feita em 400 empresas, sobre a utilização dessas ferramentas.</p>	<p>www.collaborate.com</p>	<p>Tipo: COM Pertinência: P Qlde/Qtde: R</p>
<p>NTG International Inc. Empresa de consultoria em tecnologia de Informação e design. Oferece uma série de <i>links</i> para <i>sites</i> e artigos na área de Gestão do Conhecimento. Esta página pode ser diretamente acessada através do endereço: http://www.ntgi.net/ntg/y2k/info/kmfr.htm.</p>	<p>http://www.ntgi.net/</p>	<p>Tipo: COM Pertinência: P Qlde/Qtde: R</p>
<p>Qlde/Qtde: R Knowledge Management Consortium International (KMCI) Sociedade de profissionais em KM, formada por executivos, gerentes do conhecimento, cientistas e outros, com o objetivo de desenvolver uma visão compartilhada e alinhada em ações sobre Gestão do Conhecimento para negócios e outros tipos de organizações. O <i>site</i> é organizado por tópicos como pesquisa, discussão, ciência, eventos, novidades, etc. Apesar dos papers estarem restritos a membros, há um fórum de discussões e muitas informações sobre trabalhos e projetos em KM</p>	<p>http://www.km.org/</p>	<p>Tipo: ORG Pertinência: P Qlde/Qtde: R</p>

<p>Knowledge Science Institute (KSI) O KSI foi fundado pela Universidade de Calgary em 1985 com o propósito de estudar todos os aspectos da economia do conhecimento. O instituto disponibiliza vários <i>papers</i> integrais, <i>links</i> constantemente atualizados, um jornal mensal intitulado "International Journal of Human-Computer studies / Knowledge Acquisition" (IJHCS), <i>workshops</i> e projetos interessantes como o um sistema de psicologia para modelagem dos processos de conhecimento humano e um programa de pesquisa para promover a colaboração internacional em sistemas de manufatura.</p>	<p>http://ksi.cpsc.ucalgary.ca/KSI/ Tipo: EDU Pertinência: P Qlde/Qtde: B</p>
<p>SLA - Special Libraries Association. A SLA foi criada com o objetivo de envolver os membros e patrocinadores corporativos para a realização de pesquisas, fornecimento de bolsas de estudo, manutenção de relações públicas e outros programas importantes para profissionais da informação. Vende <i>papers</i> de suas conferências e outras publicações. Traz, também, uma série de referências a artigos (alguns com links) sobre KM, livros, <i>sites</i> e <i>softwares</i>.</p>	<p>http://www.sla.org Tipo: ORG Pertinência: P Qlde/Qtde: R</p>
<p>APQC - American Productivity & Quality Center Organização sem fins lucrativos fundada por 550 empresas mundiais. Seu objetivo é a gestão de mudanças, o desenvolvimento de processos, o incremento do conhecimento e a melhoria da performance, tornando as empresas mais ágeis, criativas e competitivas. Possui <i>papers</i>, livros, resumos e uma série de estudos de caso sobre KM em empresas que podem ser acessados através da página http://apqc.org/knowledge/pubs.htm</p>	<p>http://apqc.org Tipo: ORG Pertinência: P Qlde/Qtde: B</p>
<p>David Skyrme Associates Empresa de consultoria, cujo <i>site</i>, tem por objetivo fornecer <i>insights</i> para a economia global emergente, onde ativos intangíveis, como informação e conhecimento, irão determinar a futura competitividade. Comercializa um relatório sobre KM chamado Creating the Knowledge-based Business. Contém também pequenos textos interessantes chamados Management Insights, disponibilizando <i>links</i> e boas indicações bibliográficas.</p>	<p>http://www.skyrme.com Tipo: COM Pertinência: P Qlde/Qtde: R</p>
<p>XEROX Palo Alto Research Center Centro de Pesquisas da Xerox em Palo Alto, EUA. Criado em 1970, foi constituído por uma equipe de pesquisadores com a missão de criar "a arquitetura da informação". Principais temas de pesquisa: <i>smart matter</i>, emerging document types, knowledge ecologies, network devices & document services e document information fabric. Disponibiliza alguns <i>papers</i>, mas, em sua maioria, excessivamente técnicos.</p>	<p>http://www.parc.xerox.com Tipo: COM Pertinência: PR Qlde/Qtde: R</p>



<p>InfoQuest! Information Services Empresa de consultoria para bases de dados <i>online</i>, pesquisa, inteligência competitiva e marketing. Simples e objetiva. Contém poucos <i>papers</i>, mas uma <i>Listserv</i> para bibliotecários e especialistas em informação designada para ser um fórum para discutir KM nas organizações é interessante. Instruções para acesso à lista podem ser encontradas na página http://www.tbchad.com/libknow/libknow.html</p>	<p>http://www.tbchad.com</p>	<p>Tipo: COM Pertinência: P Qlde/Qtde: R</p>
<p>CIO Magazine Archive Revista americana dedicada a CIO's, com artigos pequenos e interessantes sobre KM, Capital Intelectual e vários outros assuntos, de autores como Laurence Pruzak, Tom Davenport, Leonard M. Fuld, entre outros. Artigos específicos sobre KM podem ser encontrados na página http://www.cio.com/archive/index_knowledge_managem ent.html</p>	<p>http://www.cio.com/</p>	<p>Tipo: COM Pertinência: PR Qlde/Qtde: B</p>
<p>AIIM - Association for Information and Image Management Fórum de discussões sobre utilização de tecnologias na gestão de documentos e informações. Reúne páginas de serviços de mais de 1000 empresas ligadas à gestão de documentos. O <i>site</i> possui uma livraria, a revista <i>Inform Magazine</i> com alguns artigos interessantes disponíveis e consultoria <i>online</i>. Também possui uma base de dados com mais de 30.000 artigos disponíveis apenas para membros. Site interessante, merece ser visitado.</p>	<p>http://www.aiim.org/</p>	<p>Tipo: ORG Pertinência: PR Qlde/Qtde: R</p>
<p>Fulcrum - The Knowledge Network Company A empresa mudou do endereço http://www.fulcrum.com/. Comercializa <i>softwares</i> para KM e documentação. Na época da visita, disponibilizava dois <i>papers</i> interessantes sobre KM.</p>	<p>http://www.pcdocs.com/.</p>	<p>Tipo: COM Pertinência: PR Qlde/Qtde: R</p>
<p>EAGLES - Expert Advisory Group on Language Engineering Standards Iniciativa da Comissão Européia dentro do programa de pesquisa de linguagens e engenharia que fica hospedado na Universidade de Genebra. Contém poucos <i>papers</i> disponíveis online e outros sob encomenda (pagos). Contudo, na página http://www.issco.unige.ch/ewg95/node212.html tem início uma série de documentos sobre avaliação de Sistemas de Gestão do Conhecimento.</p>	<p>http://www.ilc.pi.cnr.it/EAGLES/home.htm</p>	<p>Tipo: ORG Pertinência: PR Qlde/Qtde: R</p>
<p>Learning-Org Dialog on Learning Organizations Site dedicado à troca e disponibilização de mensagens sobre Learning Organizations patrocinado por Innovation Associates, Inc e mantido pelo consultor Richard Karash. Muitas mensagens dizem respeito à KM.</p>	<p>http://www.learning-org.com</p>	<p>Tipo: COM Pertinência: PR Qlde/Qtde: R</p>