



GT 8: Informação e Tecnologia

INDEXAÇÃO AUTOMÁTICA E USO SOCIAL DA INFORMAÇÃO

Modalidade de Apresentação: Pôster

Remi Correia Lapa – UFPE
Renato Fernandes Corrêa – UFPE

rmcrlp@gmail.com

RESUMO

Apresenta a indexação automática como uma ferramenta que possibilita o acesso à informação, servindo como uma das etapas que contribuem para o processo que permite o uso da informação por um determinado grupo social. Objetiva levantar e analisar os diferentes métodos de indexação automática de documentos quanto a sua aplicabilidade, referenciais teóricos, vantagens e desvantagens. Tem-se como metodologia um estudo bibliográfico aprofundado de caráter qualitativo sobre a produção literária realizada no Brasil a respeito da indexação automática. Pretende-se como resultado realizar um levantamento do estado da arte da temática indexação automática, que possa dar embasamento para trabalhos futuros de aplicação dos métodos analisados na construção de sistemas de recuperação da informação para bibliotecas digitais contendo informação científica e tecnológica.

Palavras-chave: Indexação Automática. Recuperação da Informação. Uso social da informação.

1 INTRODUÇÃO

O problema gerado pelo crescimento gradativo da produção de informação só vem a agravar uma preocupação antiga que vem atormentando muitas pessoas ao longo dos tempos. De acordo com Saracevic (1995), tal problema era (e em suas formas básicas ainda é) a tarefa de tornar mais acessível um gigantesco estoque de conhecimentos ligado à necessidade de prover disponibilidade e acessibilidade a informações relevantes, aspecto crucial em nossos dias.

Em sua época, como solução para diminuir os efeitos negativos da explosão informacional Vannevar Bush propôs o uso da tecnologia:

Consideraremos um dispositivo futuro de uso individual que é uma espécie de arquivo e biblioteca privados mecanizados. Já que é importante um nome, o chamarei de *MEMEX*. Um *MEMEX* é um dispositivo que permitirá a uma pessoa armazenar todos os seus livros, arquivos, e comunicações, e que é mecanizado de tal forma que poderá se consultado com grande velocidade e flexibilidade. (BUSH, 1945, p. 105)



O surgimento do computador possibilitou que novos mecanismos, mais eficientes e eficazes, fossem investigados e elaborados com a finalidade de organizar a enorme quantidade informacional gerada, facilitando deste modo sua recuperação e seu uso.

Instituições preocupadas com o processamento da informação têm investido em pesquisas relacionadas com indexação, armazenamento e recuperação, afirma Araújo Júnior (2007). Os três processos apontados anteriormente são as principais etapas na construção de um Sistema de Recuperação de Informação (SRI) segundo Rowley (2002).

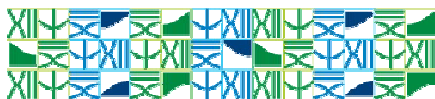
Uma boa indexação é essencial para que não ocorram ruídos nem silêncios na recuperação, problemas percebíveis quando alguém pesquisa em um banco de dados à procura de uma determinada informação e não obtém todos os documentos pertinentes às suas necessidades. Sayão apresenta esta situação na visão de um pesquisador:

Quando um pesquisador, diante de um microcomputador ligado a um banco de dados que pode estar em qualquer parte do mundo, vasculha suas estantes eletrônicas à procura de informações que definam, completem ou estabeleçam as fronteiras do seu trabalho de pesquisa, ele repete o mesmo gesto de quem mergulha na memória de seu grupo para reconstruir as lembranças comuns e dessa forma manter íntegra a sua comunidade. (SAYÃO, 1996, p. 314)

Visando melhorias na recuperação da informação são realizadas pesquisas no processo da indexação. Autores como Dias (2004), Souza (2006), Câmara Júnior (2007), Maia (2008), Lapa e Corrêa (2010) tem realizado uma série de investigações sobre a utilização da indexação automática como ferramenta para melhorar a recuperação dos documentos institucionais armazenados em bases de dados e conseqüentemente facilitar o acesso a memória institucional.

Neste aspecto a indexação automática ganha um enfoque social, pois auxilia um determinado grupo ao acesso e ao uso da informação produzida por uma comunidade específica. “Podemos considerar como “social” qualquer processo de produção/organização/consumo de informação, uma vez que ele acontece entre grupos [...] – ou seja, a geração e apropriação de informações só ocorrem no âmbito da sociedade, das relações sociais.” (CARDOSO, 1994, p. 107-108).

Diante do que foi exposto este trabalho tem por objetivo geral levantar e analisar os diferentes métodos aplicados no processo de indexação automática de documentos textuais produzidos no Brasil. Dentre os objetivos específicos estão: realizar estudos bibliográficos sobre indexação automática enquanto área de pesquisa; e sobre os métodos de indexação automática; analisar aplicabilidade, vantagens e



desvantagens de cada método de indexação automática levando em conta o contexto das bibliotecas digitais de informação científica e tecnológica.

2 INDEXAÇÃO AUTOMÁTICA

A indexação automática é na opinião de Silva “aquela resultante do trabalho intelectual de um profissional para checagem do valor dos termos atribuídos a um documento por um programa de computador.” (SILVA, 2004, p. 145).

Um conceito diferente apresenta a indexação automática apenas quando a análise e extração dos conceitos são realizadas por programas de computador sem a intervenção humana (neste caso o texto precisa estar em formato eletrônico).

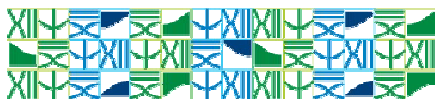
Corroborando com este conceito Araújo Júnior (2007) ao apresentar a indexação automática como um conjunto de procedimentos realizados pelo computador, permitindo identificar e selecionar os descritores que representam o assunto dos documentos sem a intervenção direta do homem.

Lancaster (2004) apresenta duas formas de se realizar a indexação automática: uma delas ocorre quando o documento é indexado por um programa que adota critérios de frequência, posição e contexto em que as palavras aparecem no decorrer do texto para então extraí-las como descritores, esta forma é conhecida como indexação automática por *extração*; a outra consiste em uma representação temática por meio de termos selecionados de um vocabulário controlado (tesauro ou lista alfabética) pelo programa.

Sobre as diferentes aplicações do computador na indexação automática Guinchat e Menou (1994) comentam que o computador utiliza diferentes métodos para localizar os termos significativos. De acordo com os autores o computador pode comparar os termos do resumo com os descritores da linguagem documental. Desta forma, cada vez que um descritor aparece no resumo, o documento é indexado com o termo correspondente.

Os autores ainda apresentam outros métodos de indexação automática como: análises estatísticas em uma amostragem dos textos; determinação da frequência com que aparecem as palavras; os métodos lingüísticos que utilizam tratamentos morfológicos para o reconhecimento das estruturas mais significativas; e os métodos sintáticos que acrescentam ao método anterior uma interpretação sintática feita pelo computador.

Na visão de Vieira (1988a), o emprego da indexação automática é uma alternativa viável, face ao grande número de documentos eletrônicos existente, tornando cada dia mais



difícil a realização a indexação manual com um mínimo de qualidade requerida para assegurar o acesso posterior à informação.

2.1 Uma ferramenta social

A indexação automática surge como uma alternativa tecnológica para amenizar os problemas do boom informacional, sua aplicação é possibilitada graças à característica interdisciplinar da ciência da informação.

“Lidar com o grande volume e a diversificação de informações registradas em variadas formas, com vistas à sua mais ampla difusão, foi o imperativo condicionante da ciência da informação. Fortemente influenciadas pelas ciências empíricas, as primeiras manifestações desse campo embrionário pretendiam estabelecer leis universais que representassem o fenômeno informacional, daí a recorrência a modelos matemáticos (teoria da informação), físicos (entropia) ou biológicos (teoria epidemiológica).” (CARDOSO, 1996, p. 73-74)

Na década de setenta, entra em cena um personagem que irá direcionar o enfoque da ciência da informação: o usuário, Cardoso (1996). Se antes os processos tecnológicos eram suficientes por si só (paradigma físico), a partir dos anos setenta são observados como um meio direcionado para beneficiar o usuário (paradigma social) como relata Araújo (2003) ao afirmar que a ciência da informação não “nasceu” como uma ciência social, mas estava ligada inicialmente à computação e à recuperação automática da informação.

Neste aspecto, Capurro e Hjørland (2007) apresentam a informação não mais como um ‘objeto’ e sim como uma ‘mudança’ no conhecimento do usuário. Esta mudança na percepção da função da ‘informação’ recuperada em um SRI permite que a indexação automática seja observada como uma ferramenta social, no momento em que possibilita ao usuário encontrar e usar a informação desejada.

Ora, se a informação não é apenas uma ‘coisa’ a ser fisicamente observada, e sim historicamente construída, pois é ela que ‘dá forma a alguma coisa’, pode-se concluir que os sujeitos criam mecanismos informacionais para reconhecer, interpretar e transmitir significados – onde ela está inserida dentro de um contexto cultural e social. (Nascimento e Marteleto, 2004, p.9)

Ainda de acordo com Nascimento e Marteleto (2004), não há dúvidas que os processos de produção, transferência e uso das informações são sociais, já que eles ocorrem entre a sociedade e suas relações sociais.

4 PROCEDIMENTOS METODOLÓGICOS



4.1 Caracterização da Pesquisa

Trata-se de uma pesquisa com características tanto exploratórias quanto descritiva, pois tem como finalidade realizar uma análise conceitual sobre a indexação automática e respectivos métodos. Além disso, esta investigação pretende analisar quais dos diferentes procedimentos de indexação automática são mais eficientes e eficazes na recuperação da informação e estudar como cada método é interpretado e utilizado pelos pesquisadores na literatura científica.

Como primeiro procedimento metodológico, esta pesquisa realizará uma revisão de literatura através da coleta, seleção e leitura de livros, artigos de periódicos científicos e literaturas cinzentas (monografias, teses e dissertações) encontradas em bibliotecas ou disponíveis na internet que tratem sobre indexação automática, levantamento e análise das ferramentas que auxiliam na indexação de documentos textuais.

Este trabalho possui como material de estudo as fontes de informação que abordam a aplicação da indexação automática como etapa na construção de um Sistema de Recuperação de Informação, visando tornar o processo de recuperação dos documentos mais eficiente e eficaz.

Faz parte da próxima etapa metodológica, localizar os artigos acadêmicos, teses e dissertações que apresentam: no título; e/ou nas palavras-chave; e/ou no resumo, os termos: indexação automática, recuperação de informação, sistema de recuperação de informação. Neste processo, a pesquisa será realizada em bases de dados virtuais, tais como: BRAPCI¹, BDTD Nacional², Google Acadêmico³, entre outras.

O procedimento seguinte é a apreciação crítica dos documentos selecionados separando-os em grupos de acordo com cada tipo de procedimento de indexação automática.

Além disso, inclui uma fase onde será realizado um levantamento de indicadores para avaliação dos métodos de indexação automática, bem como uma análise comparativa destes métodos.

4.2 Resultados Esperados

Espera-se obter através desta revisão bibliográfica um levantamento e uma análise dos diferentes métodos de indexação automática de documentos textuais em português que serão apresentados através de tabelas comparativas.

¹ <http://www.brapci.ufpr.br>

² <http://bdttd.ibict.br/>

³ <http://scholar.google.com.br/sczhhp?hl=pt-BR&tab=ws>



Outro resultado aguardado na conclusão deste trabalho é o de analisar a aplicabilidade de cada método de indexação automática, apresentando suas vantagens e desvantagens, observando o contexto das bibliotecas digitais de informação científica e tecnológica.

A partir deste estudo, pretendemos contribuir ao fornecer embasamento a futuros trabalhos que se proponham a aplicar os métodos analisados na construção de Sistemas de Recuperação da Informação para bibliotecas digitais contendo informações científicas e tecnológicas.

ABSTRACT

Presents automatic indexing as a tool that provides access to information, serving as one of the steps of the process that allows the use of information by a particular social group. This work aims to select and analyze the applicability, theoretical advantages and disadvantages different methods of automatic indexing of documents. Its methodology has been a bibliographical study about the literary production carried out in Brazil about the automatic indexing. It is intended as a result to survey the state of the art automatic indexing issue, which may give foundation for future work the application of methods discussed in the construction of information retrieval systems for digital libraries containing scientific and technological information.

Keywords: Automatic indexing. Information retrieval. Social use of information.

REFERÊNCIAS

ARAÚJO, C. A. A. A ciência da informação como ciência social. **Ciência da Informação**, Brasília, v. 32, n. 3, p. 21-27, set./dez. 2003.

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

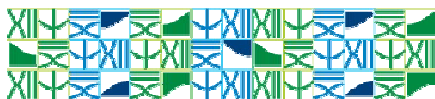
BUSH, V. As we may thing. **Atlantic Monthly**, v.176, 1, p.101-108, 1945.

CÂMARA JÚNIOR, A. T. Indexação automática de acórdãos por meio de processamento de linguagem natural. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. **Anais...**, Salvador: ENANCIB, 2007.

CARDOSO, A. M. P. Retomando possibilidades conceituais: uma contribuição à sistematização do campo da informação social. **Revista da Escola de Ciência da Informação da UFMG**, v. 23, n. 2, p. 107-114, jul./dez. 1994.

CARDOSO, A. M. P. Pós-modernidade e informação: conceitos complementares? **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 63-79, jan./jul. 1996.

CAPURRO, R.; HJORLAND, B. O conceito de informação. **Perspectiva em Ciência da Informação**, v.12, n. 1, p. 148-207, jan./abr. 2007.



DIAS, M. A. L. **Extração automática de palavras-chave na língua portuguesa aplicada a dissertações e teses da área das engenharias** [dissertação]. Campinas, SP; 2004.

GUINCHAT, C.; MENOU, M. **Introdução geral às ciências e técnicas da informação e documentação**. 2. ed. corr. aum. Brasília: IBICT, 1994. 540 p.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2. ed. ver. atual. Brasília: Briquet de. Lemos, 2004.

LAPA, R. C.; CORRÊA, R. F. Seleção de descritores para a indexação automática de teses e dissertações da UFPE. In: Encontro Nacional de Estudantes de Biblioteconomia, Documentação, Gestão e Ciência da Informação, 33., 2010, João Pessoa. **Anais eletrônicos...** João Pessoa, 2010. Disponível em <<http://dci.ccsa.ufpb.br/enebd/index.php/enebd/article/view/20>>, Acesso em 27 jul. 2011.

MAIA, L. C. G. **Uso de sintagmas nominais na classificação automática de documentos eletrônicos**. 2008. Tese (Doutorado em Ciência da Informação) – Universidade Federal de Minas Gerais – UFMG. Belo Horizonte, 2008.

NASCIMENTO, D. M.; MARTELETO, R. M. A informação construída nos meandros dos conceitos da teoria social de Pierre Bordieu. **DataGramaZero: Revista de Ciência da Informação**, v. 5, n. 5, out. 2004.

ROWLEY, J. **A biblioteca eletrônica**. 2. ed. Brasília: Brinquet de Lemos, 2002.

SARACEVIC, T. Interdisciplinarity nature of information science. **Ciência da Informação**, Brasília, v. 24, n. 1, p. 36-41, 1995.

SILVA, M. R.; FUJITA, M. S. L. A prática de indexação: análise da evolução de tendências teóricas e metodológicas. **Transinformação**, Campinas, v. 16, n. 2, p. 133-161, maio/ago. 2004.

SAYÃO, F. Bases de dados: a metáfora da memória científica. **Ci. Inf.**, Brasília, v. 25, n. 3, p. 314-318, set./dez. 1996

SOUZA, R. R. Uma proposta de metodologia para indexação automática utilizando sintagmas nominais. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, n. esp., p.42-59. 2006.

VIEIRA, S. B. Indexação automática e manual: revisão de literatura. **Ci. Inf.**, Brasília, v. 17, n. 1, p. 43-57, jan./jun. 1988.