



XI Encontro Nacional de Pesquisa em Ciência da Informação
Inovação e inclusão social: questões contemporâneas da informação
Rio de Janeiro, 25 a 28 de outubro de 2010

GT 2 - Organização e Representação do Conhecimento
Modalidade de apresentação: Comunicação oral

ESTUDOS DE APLICAÇÕES DA LÓGICA PARACONSISTENTE EM UM MODELO HÍBRIDO DE INDEXAÇÃO AUTOMÁTICA

Carlos Alberto Correa
Universidade de São Paulo
Nair Yumiko Kobashi
Universidade de São Paulo

Resumo:

Pode-se afirmar que os processos de organização, recuperação e visualização de informações estão interligados por pelo menos um ponto em comum: a indexação de textos ou documentos. A indexação, um componente essencial do tratamento e organização de informações, é importante tanto para a recuperação quanto para a visualização de informações. Neste contexto, destacam-se as soluções que objetivam automatizar a indexação. As pesquisas que propõem soluções de indexação automática são desenvolvidas com base em diferentes pressupostos teóricos, entre os quais podem ser destacados: a Estatística, a Lingüística e o Controle de Vocabulários (Leiva, 1999). A grande maioria dessas soluções desenvolve modelos híbridos que combinam dois ou três desses pressupostos. Outras abordagens para o problema da indexação automática utilizam, além dos pressupostos anteriormente citados, teorias que permitem o tratamento da incerteza, da imprecisão e da vagueza. O objetivo deste trabalho é discutir o potencial teórico de utilização da lógica paraconsistente, uma lógica não clássica, com capacidade para tratar situações que envolvem incerteza, imprecisão e vagueza nos procedimentos de indexação automática.



1. Introdução

Motivada principalmente por necessidades práticas, a humanidade desenvolveu um conjunto significativo de técnicas, “tecnologias” e saberes voltados a um objetivo central: armazenar, externamente à mente humana, os incontáveis registros de conhecimentos acumulados com o passar do tempo.

Assim, podemos considerar que a fala, a escrita, a tipografia, o computador eletrônico e as bibliotecas constituem exemplos de “dispositivos” e/ou organizações que foram criados ou adaptados para serem utilizados como mecanismos externos de armazenamento dos registros anteriormente citados. Tal armazenamento era e ainda é efetuado levando-se em consideração, principalmente, critérios de valor e eventual necessidade futura de recuperação.

Com o passar do tempo, o volume dos registros externos de conhecimento não parou de crescer, decorrente de diferentes tipos de influências históricas, econômicas e sociais. O crescimento vertiginoso dos registros levou ao surgimento de diversos tipos de problemas, centrados, principalmente, nas questões associadas aos meios de armazenamento e aos processos e meios de acesso aos mesmos.

O advento dos computadores digitais e suas derivações mais recentes, como por exemplo a Internet, ao invés de resolver ou minimizar o problema de acesso à informação (um outro modo de nos referirmos aos registros do conhecimento produzido...) parece tê-lo potencializado.

Assim, torna-se imprescindível garantir a efetividade do acesso à informação por meio de mecanismos que atuem sobre grandes coleções de documentos, de modo a fornecer pistas que ajudem os usuários a encontrar os documentos de que necessitam.

2. Organização, recuperação e visualização de informações

O desenvolvimento de pesquisas para encontrar possíveis soluções para esses problemas levam ao encontro de pressupostos teóricos e metodológicos que, contemporaneamente, associam questões e estudos associados à:

- . organização e representação de informação;
- . recuperação de informação;
- . visualização de informações;
- . construção de sistemas e interfaces virtuais.



Resumidamente, pode-se afirmar que as discussões que envolvem os conhecimentos associados à organização da informação podem ser consideradas observando-se pelo menos três domínios específicos: a Classificação, a Análise Documentária e a Terminologia.

No âmbito da Análise Documentária podem ser descritos dois processos principais: a indexação e a produção de resumos. Embora Lancaster (2004) afirme que a distinção entre tais processos esteja se tornando cada vez mais difusa, muitas pesquisas ainda são efetuadas levando-se em conta a sua separação.

O mesmo autor cita Fairthorne (apud LANCASTER, 2004), que enfatizou o fato de ser a indexação o problema fundamental bem como o mais dispendioso obstáculo para uma efetiva recuperação de informação. A expressão recuperação de informação foi utilizada pela primeira vez em 1951, por Calvin Mooers, que a definiu como a disciplina que *“trata dos aspectos intelectuais da descrição da informação e sua especificação para busca e também de qualquer sistema, técnicas ou máquinas que são empregadas para realizar a operação”* (MOOERS APUD FERNEDA 2003, p.11).

Contemporaneamente, a Análise documentária, a Recuperação de informações e a Visualização de informações podem ser associadas tanto à Ciência da Informação quanto à Ciência da Computação (BAEZA-YATES E RIBEIRO-NETO, 1999).

O funcionamento da recuperação de informação se dá, via de regra, através de sistemas informatizados. De acordo com Belkin (1993), tal funcionamento depende do desenvolvimento de representações e substitutos para os textos armazenados. Tais representações (resumos e índices) são obtidas por meio de sucessivas operações de análise e síntese, processos estudados no âmbito da Análise Documentária.

As dificuldades com os processos de recuperação de informação podem ser minimizadas à medida que os usuários conheçam melhor o espaço de busca em que estão atuando, aumentando assim, a efetividade dos resultados das pesquisas. Este é um dos objetivos principais das pesquisas relacionadas à visualização de informações e de domínios de conhecimento.

Muitos sistemas de visualização de informações retratam a estrutura semântica global de uma coleção de documentos. Assim, os usuários podem utilizar essa visualização estrutural como base para suas buscas subsequentes (BORNER et al. 2003).

A visualização de informações estuda o desenvolvimento de técnicas de representação visual de conjuntos de dados que não apresentam uma geometria intrínseca ou uma natureza gráfica. Refere-se à representação gráfica de dados complexos, multidimensionais e volumosos, tais como:



hierarquias muito grandes, bases de dados textuais ou dados estatísticos. A expressão também é associada à representação dos conjuntos de dados não estruturados ou abstratos (documentos, coleções, sites da WWW, entre outros).

A visualização de informações tem potencial para ajudar as pessoas a encontrarem a informação de que precisam de forma mais efetiva e intuitiva (CHEN, 2006). Dessa forma, a visualização pode ser vista como:

- uma maneira de diminuir o esforço necessário para explorar coleções de informações;
- uma ferramenta que auxilia a identificar padrões e a obtenção de *insights* em situações que envolvem a manipulação e acesso a grandes quantidades de informação;
- um modo de enxergar o “invisível” ou de “descobrir” o “desconhecido”.

O desenvolvimento de sistemas de visualização de informações não estruturadas (documentos textuais ou de outro formato) é subsidiado, de acordo com Zhu & Chen (2005), por estudos realizados nas seguintes dimensões de pesquisa: representação de informação, interação usuário-interface e análise da informação.

A análise da informação é utilizada com o objetivo de reduzir a complexidade e extrair estruturas proeminentes de grandes quantidades de informação não estruturada. Para atingir tais objetivos, a análise é desenvolvida em dois estágios: indexação e análise propriamente dita.

Neste contexto, a indexação é vista como um modo de extrair a semântica da informação para representar seu conteúdo. Assim, diferentes algoritmos são utilizados para diferentes tipos de mídias, tais como: textos, imagens, áudio e vídeo.

No segundo estágio, de análise propriamente dita, são utilizadas técnicas de classificação e agrupamento (clustering).

Essencialmente, a visualização de domínios do conhecimento configura-se como um conjunto de técnicas que busca propiciar uma visão geral de um domínio específico do conhecimento. Neste caso, o foco muda da busca por documentos potencialmente relevantes para a visualização de contextos disciplinares ou temáticos, em geral representados por agrupamentos de documentos. Tais técnicas podem ser muito úteis para o desenvolvimento de revisões sistemáticas, auxiliar na recuperação de documentos, instruir decisões políticas, avaliar a performance da pesquisa, gerenciar programas de ciência e tecnologia ou inteligência competitiva.

Uma série de eventos podem ser utilizados para descrever o modo como uma visualização de domínios é criada. Dessa forma, segundo Borner et al. (2003), podem ser enumerados os seguintes passos que compõem o processo de visualização: (1) extração de dados; (2) definição das



unidades de análise; (3) seleção de medidas; (4) cálculo de similaridades entre as unidades; (5) ordenação ou atribuição de coordenadas para cada unidade e (6) exibição e interação ou uso da visualização para análise e interpretação.

A etapa de extração de dados consiste na aquisição dos objetos de informação, em geral documentos, que serão utilizados na construção da visualização. A seleção das unidades de análise deve ser direcionada para o tipo de questão que se espera seja respondida ou analisada com a visualização. As unidades mais comuns a serem utilizadas são: revistas, documentos textuais, autores e descritores. A seleção de medidas envolve o desenvolvimento de cálculos de frequência e contagem que servirão de base para o cálculo de similaridades entre os objetos de informação. Em geral, são utilizados como medidas os seguintes indicadores: ocorrência e co-ocorrência de termos, co-citação e co-autoria.

O cálculo de similaridades é efetuado com o objetivo de estabelecer “relações de distância” entre os objetos de informação (ou seus agrupamentos) a serem exibidos. Nesta fase são utilizadas as medidas de similaridade escolhidas na fase anterior. Também é muito utilizado, nesta fase, o modelo de espaço vetorial, originalmente desenvolvido para recuperação de informação (SALTON, 1989). Este modelo é composto de três estágios: indexação do documento, ponderação de termo e cálculo dos coeficientes de similaridade.

Na etapa de ordenação, as medidas de similaridades obtidas são utilizadas na construção do mapa ou visualização que será apresentada, pelo sistema, na tela de um computador. Finalmente, a visualização é exibida na tela, sendo oferecidas facilidades de interação para que o usuário possa efetuar as análises que deseja.

2.1 O papel da indexação

Conforme visto nos parágrafos anteriores, o processo de indexação assume grande importância nos procedimentos de organização, recuperação e visualização de informações. A figura a seguir resume tal convicção.

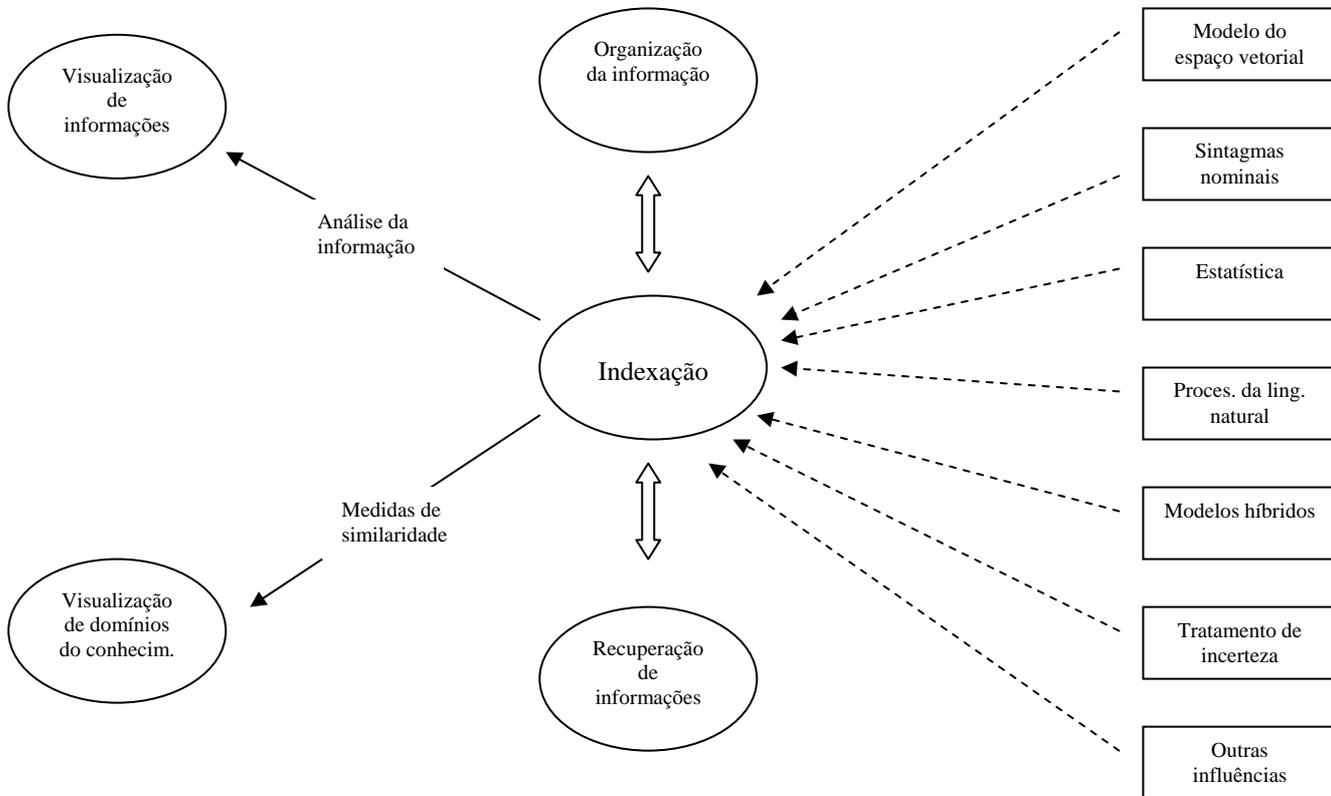


Figura 1. Importância da indexação (Esquema dos autores)

As setas tracejadas mostram algumas teorias, modelos e campos de estudos que influenciam o desenvolvimento de técnicas de indexação, principalmente aquelas ligadas à indexação automática. As setas contínuas mostram a influência da indexação através de processos específicos, nos estudos sobre visualização; as setas cheias indicam sua interação com campos de conhecimentos associados tanto à Ciência da Informação quanto à Ciência da Computação.

2.2 Um sistema de visualização de informações – Projection Explorer

Ao longo das últimas décadas, muitos sistemas para visualização de informações e de domínios de conhecimento foram desenvolvidos. Os sistemas possuem propósitos diferentes e se utilizam também de metáforas diferentes para representar, nas telas dos computadores, coleções de documentos ou alguns de seus atributos. Vários exemplos destes sistemas podem ser vistos em Borner et al (2003), Chen (2006) e Zhu & Chen (2005).

No Brasil, foi desenvolvido no Instituto de Computação e Matemática Computacional da USP/São Carlos um sistema de visualização que vem sendo utilizado por diferentes grupos de

pesquisadores para testes, simulações e pesquisas. Tal sistema é conhecido por PEx – Projection Explorer (PAULOVICH et al, 2007, PAULOVICH et al, 2008)¹.

O sistema PEx utiliza pequenos círculos coloridos como metáfora visual para representar documentos individuais de uma coleção. O aplicativo efetua uma série de cálculos para determinar o grau de similaridade entre os documentos da coleção e, dessa forma, estipula a “distância” entre os objetos informacionais, conforme descrito no final da seção 2. Tal procedimento é necessário porque quando tais objetos forem representados, através da metáfora visual do aplicativo, eles serão mostrados tão próximos quanto o cálculo determinar. Além disso, são efetuados procedimentos de agrupamento e classificação que, por fim, serão utilizados na construção da visualização final propriamente dita.

A figura a seguir descreve, resumidamente, as fases que compõem o processo de criação da visualização de uma coleção de documentos com o Projection Explorer.

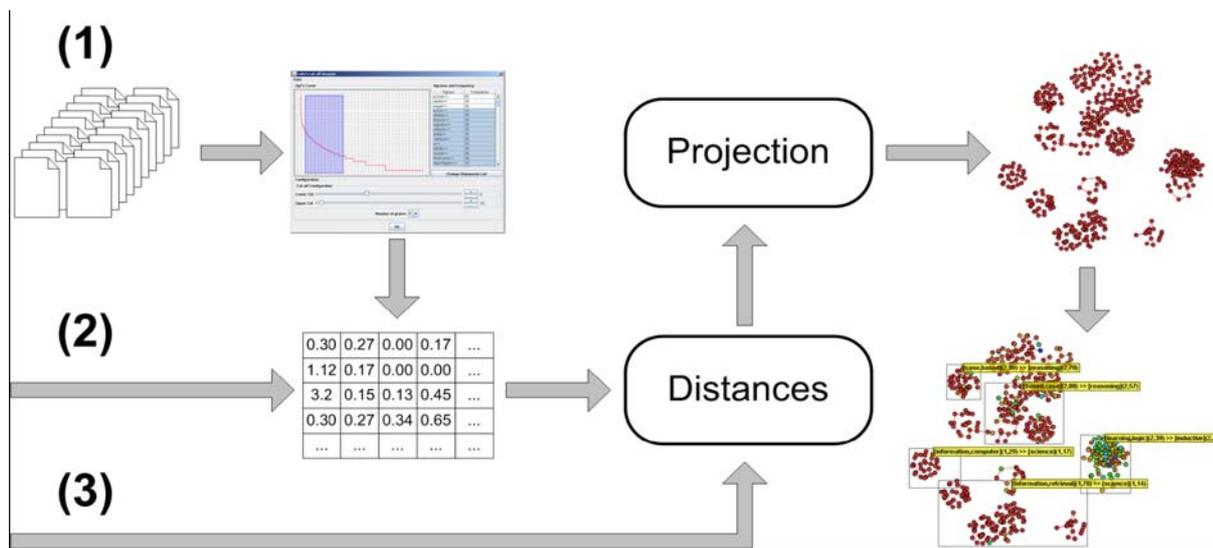


Figura 2 – Processo de visualização de objetos informacionais

Fonte: Paulovich et al, 2007

Na Fig. 2, podem ser observadas as diferentes maneiras de alimentar o sistema com os dados a serem visualizados. Tais dados podem ser: uma coleção de documentos (1); uma tabela estruturada de dados relativos a um conjunto de documentos (2) ou uma matriz de distâncias (3) (PAULOVICH et al, 2007). A partir do estabelecimento das distâncias entre os elementos

¹ O sistema Projection Explorer pode ser obtido, gratuitamente, em <http://infoserver.lcad.icmc.usp.br/infovis2/PEX>.

individuais é realizada a redução de dimensionalidade, também chamada de projeção e, finalmente, os objetos são visualizados.

Pode-se observar, na Figura 2, que ao alimentar o sistema com uma coleção de documentos, o sistema estabelecerá uma tabela de dados relativos aos pesos que podem ser atribuídos aos termos de indexação de cada documento, de acordo com os cálculos estabelecidos no modelo de espaço vetorial (SALTON, 1989).

As Fig. 3 e 4, a seguir, mostram exemplos de visualizações de coleções de documentos criadas com o Projection Explorer.

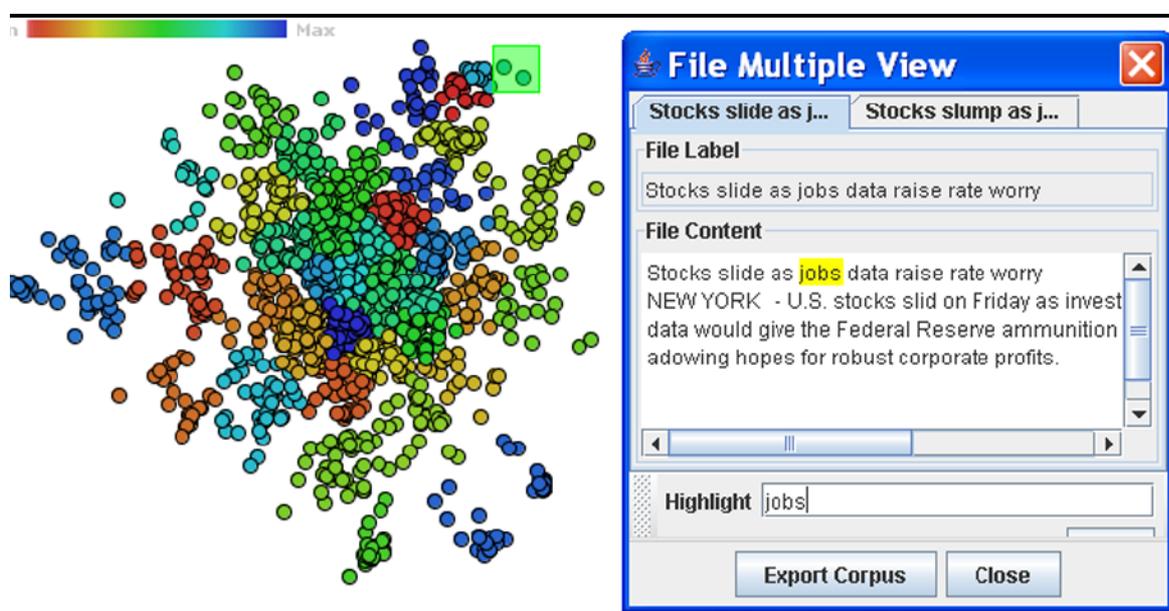


Figura 3 – Exemplo de visualização – PEx
Fonte: Paulovich et al, 2007

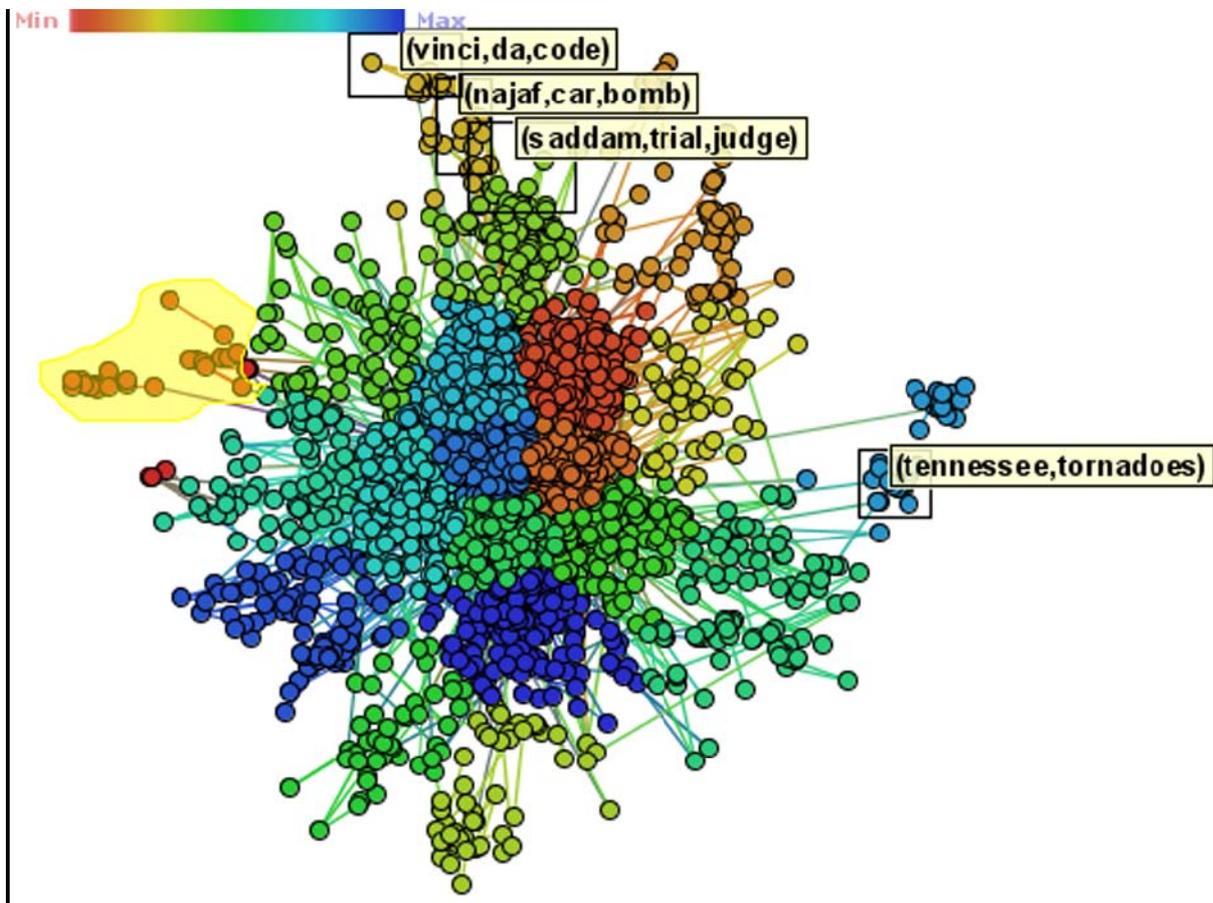


Figura 4 – Exemplo de visualização no PEx com identificação de agrupamentos de textos
Fonte: Paulovich et al, 2007

O Projection Explorer oferece várias facilidades entre as quais podem ser citadas:

- possibilidade de estabelecer rótulos para os agrupamentos de objetos selecionados;
- estabelecimento de cores diferentes para diferentes grupos de objetos;
- comparação de mapas de documentos criados a partir de diferentes técnicas de projeção;

3. Indexação automática

A utilização maciça dos computadores como meio para o armazenamento dos estoques de informação e como ferramenta para recuperação de informação acabou por influenciar os procedimentos relativos a esses dois processos executados pelos seres humanos. Assim, parece natural que os computadores começassem a ser utilizados para reproduzir ou executar essas atividades humanas.

Nesse sentido, a atividade de indexação passou a ser alvo dos mais diversos tipos de experimentos objetivando realizá-la de forma automatizada. Deve-se observar, ainda, que é possível



enumerar diversos motivos tanto a favor quanto contra a automatização da indexação (LEIVA, 1999; LANCASTER, 2004).

A quantidade de experimentos computacionais envolvendo a indexação automática não parou de crescer ao longo do tempo e a diversificar-se quanto aos métodos e técnicas empregadas. Tais eventos levaram ao surgimento de diferentes abordagens de pesquisa e conduziram a indexação automática a interagir com diferentes disciplinas direcionadas ao desenvolvimento de aplicações computacionais.

Os principais métodos empregados na indexação automática envolvem a utilização de estatística, métodos linguísticos, vocabulários controlados e sistemas híbridos (que se utilizam da combinação dos três primeiros métodos) (LEIVA, 1999; LANCASTER, 2004; KOWALSKI e MAYBURY, 2000). Tais métodos deram origem a uma grande quantidade de pesquisas e sistemas de recuperação de informação, com seus respectivos mecanismos de indexação automática.

Entretanto, as limitações de desempenho das soluções de indexação automática vêm estimulando os pesquisadores tanto da Ciência da Computação quanto da Ciência da Informação a utilizarem outras teorias ou abordagens não tradicionais.

3.1 Indexação automática e incerteza

Alguns autores consideram que os processos associados à recuperação de informações envolvem incerteza, vagueza e imprecisão (WILSON, 1997; CRESTANI E PASI, 2003). Nesse contexto, torna-se necessário a utilização de abordagens teóricas específicas para lidar com essas características, tais como as lógicas não clássicas, a teoria das probabilidades ou a teoria da evidência de Dempster-Shafer.

O uso de lógicas não clássicas no contexto da recuperação de informações pode ser observado em trabalhos que utilizam a lógica difusa, tanto nos processos de indexação automática quanto para a modelagem formal dos sistemas de recuperação de informações (OGAWA AT AL., 1991; BORDOGNA E PASI, 1995; MOLINARI E PASI, 1996).

Outra solução envolve a utilização da teoria da evidência de Dempster-Shafer. Essa teoria, de cunho essencialmente matemático, permite o tratamento formal de situações em que o fenômeno da incerteza se mostra presente. Dessa forma, pesquisas foram desenvolvidas com o intuito de utilizar esta teoria no contexto da indexação automática e da recuperação de informação (LALMAS E RUTHVEN, 1998).

Contudo, outras teorias também fornecem subsídios para o tratamento da incerteza, vagueza e imprecisão, entre elas a lógica paraconsistente, que será detalhada a seguir.



4. Lógica paraconsistente

A Lógica clássica, de origem aristotélica, está fundamentada em três princípios: (i) princípio da identidade – que estabelece que todo objeto é idêntico a si mesmo; (ii) princípio da não contradição – que afirma que uma proposição não pode ser verdadeira e falsa ao mesmo tempo; e (iii) princípio do terceiro excluído – que estabelece que uma proposição é verdadeira ou falsa, não havendo uma terceira possibilidade.

A modificação de um destes princípios dará surgimento a um novo tipo de lógica, com reflexos nos métodos que a compõem, entre os quais podem ser citados: o cálculo de predicados de primeira ordem, as tabelas verdade e os procedimentos de prova.

A lógica paraconsistente, cujos fundamentos podem ser associados aos trabalhos iniciais do polonês Stanislaw Jaskowski e do brasileiro Newton da Costa, derroga o princípio da não contradição. Ou seja, ela tolera a existência de proposições contraditórias e permite o desenvolvimento de métodos semelhantes ao da lógica clássica, servindo de fundamento para o surgimento de cálculos e teorias em que o fenômeno da contradição ocorra. Além disso, a tolerância à existência das contradições possibilitou o desenvolvimento de métodos capazes de fornecer um arcabouço teórico para o tratamento da incerteza, imprecisão e vagueza (DA COSTA et al., 1999).

Uma das variações da Lógica paraconsistente é a Lógica paraconsistente anotada. Nesta lógica, atribui-se, a uma proposição, duas variáveis: os graus de crença e de descrença, com valores que variam no intervalo $[0,1]$. Esta abordagem torna possível o estabelecimento de quatro estados lógicos (inconsistente, verdadeiro, falso, indeterminado), em vez dos dois estados previstos na lógica clássica (verdadeiro e falso). Os estados lógicos são determinados com base nos dois valores da anotação, descritos por um par (μ_1, μ_2) , que representam, respectivamente, o grau de crença e o grau de descrença atribuído a uma proposição. Como exemplo, consideremos p uma proposição com graus de crença e descrença (μ_1, μ_2) . O par $(1.0, 0.0)$ significa crença total e ausência de descrença na proposição ou, dito de outra forma, a mesma é verdadeira. Analogamente, podem ser descritos os seguintes estados lógicos: $(1.0,1.0)$ – inconsistente (crença total e descrença total); $(0.0,1.0)$ – falso (ausência de crença e descrença total); e $(0.0,.0.0)$ – indeterminado (ausência de crença e ausência de descrença). Pode-se observar que estes são os valores extremos passíveis de serem atribuídos aos graus de crença e descrença. Outros valores, pertencentes ao intervalo $[0,1]$ podem ser atribuídos, permitindo amplas possibilidades de estabelecer estados intermediários aos quatro estados lógicos anteriormente definidos. Uma análise pode ser efetuada de modo que, para



uma dada proposição, pode-se avaliar sua tendência. Por exemplo, uma proposição com valores anotados de (0.8,0.2) pode ser entendida como sendo muito próxima de ser verdadeira.

5. Indexação automática e lógica paraconsistente

A utilização da lógica difusa e da teoria de Dempster-Shafer na indexação automática pressupõe o estabelecimento de algum tipo de função que estabeleça um grau de agregação entre um ou mais termos e um documento no qual os mesmos ocorram. Este mecanismo é nomeado de função de pertinência na lógica difusa (BORDOGNA E PASI, 1995) e função de crença na Teoria de Dempster-Shafer (LALMAS E RUTHVEN, 1998). A solução que envolve a teoria de Dempster-Shafer avalia também o grau de plausibilidade para uma dada proposição. Contudo, ambas as abordagens desconsideram avaliações que indiquem a “desagregação” de um termo a um documento.

Conforme vimos anteriormente, a Lógica paraconsistente anotada utiliza, formalmente, uma variável para o grau de crença (ou de agregação) e uma variável que indica o grau de descrença (ou “desagregação”) quanto a uma proposição. Consideremos, por exemplo, a proposição p_t – o termo t é um bom termo para indexar o documento D . Se atribuirmos valores μ_1 e μ_2 a p_t , com ambos os graus estabelecidos por meio de algum cálculo ou heurística, pode-se estabelecer, com precisão, o quanto a proposição se aproxima de um dos quatro estados lógicos conhecidos. Um termo t , para o qual os valores de crença e descrença fossem, respectivamente, 1 e 0, poderia ser considerado como um bom termo de indexação, pois a proposição p_t seria considerada verdadeira. Já um termo com valores 0 e 1 não seria um bom termo de indexação, uma vez que a proposição seria considerada falsa. Pode-se destacar que esta sistemática possibilita estabelecer a adequação de um termo t levando-se em conta a utilização de duas variáveis, em vez de uma, conforme as soluções anteriormente descritas de Bordogna e Pasi (1995) e de Lalmas e Ruthven (1998).

Acredita-se que a utilização de duas variáveis - os fatores de crença e descrença – permitirá o desenvolvimento de uma sistemática de indexação que poderá ser comparada tanto com as soluções clássicas de indexação automática quanto com as soluções que se utilizam de outras teorias para o tratamento da incerteza, aqui descritas.

Uma proposta para utilização da Lógica paraconsistente no contexto da recuperação de informação foi efetuada por Sucolotti (2001) que sugeriu sua utilização para expansão de consultas efetuadas a um sistema de recuperação de informações, com a utilização de tesouros. Esta proposta considerava a utilização dos índices atribuídos aos documentos e sua comparação com o conteúdo do tesouro.



No entanto, consideramos o potencial de utilização da lógica paraconsistente tanto na geração de índices quanto no seu refinamento, uma vez utilizada alguma metodologia já conhecida. Acredita-se que tal procedimento possibilite um aumento na efetividade da recuperação de informação.

6. Considerações finais

A importância da indexação para os processos de recuperação, organização e visualização, assim como as dificuldades inerentes aos problemas associados a ela justificam a realização de pesquisas visando à obtenção de processos, procedimentos e teorias que garantam sua efetividade.

A complexidade dos problemas associados à indexação estimulou o desenvolvimento de trabalhos multidisciplinares. Ganhos de efetividade com esta abordagem incentivam a utilização de teorias até então inexploradas. Nesse sentido destacam-se os trabalhos que buscam o tratamento da incerteza, imprecisão e vagueza inerentes à indexação.

Nesse contexto, configura-se o potencial de utilização da Lógica paraconsistente, apta a trabalhar, formalmente, as situações de incerteza e imprecisão. Ao expandir os dois estados lógicos tradicionais para quatro (inconsistente, verdadeiro, falso, indeterminado), e possibilitar o estabelecimento de graduações entre tais estados, essa lógica estimula sua aplicação no tratamento dos complexos problemas associados à indexação. Acredita-se que tal abordagem potencializará os procedimentos e soluções associados à organização, recuperação e visualização de informações.

Alem disso, os sistemas de visualização de informações, quando alimentados com coleções de documentos, podem utilizar-se dos índices estabelecidos para cada documento da coleção ou realizar cálculos para o estabelecimentos destes índices, da mesma forma como fazem os sistemas de recuperação de informações. Assim, os sistemas de visualização apresentam potencial para testes com diferentes abordagens de indexação, fornecendo um ambiente propício para a verificação dos efeitos dos diferentes modos de realizar a indexação.

Nesse sentido, encontra-se em desenvolvimento a preparação de testes da proposta aqui delineada utilizando-se o sistema de visualização de informações Projection Explorer. A idéia básica dos testes pode ser resumida da seguinte sequencia: (1) gerar a visualização de uma coleção de documentos utilizando-se o ferramental de indexação embutido originalmente no PEx; (2) utilizar uma heurística desenvolvida no âmbito da lógica paraconsistente para redefinir e/ou refinar os índices estabelecidos para os documentos da mesma coleção; (3) obter uma visualização da coleção à partir destes novos índices; (4) comparar a visualização obtida em (2) com a visualização obtida em (1). Cabe acrescentar que existem diferentes métodos que possibilitam mensurar a



qualidade de uma dada visualização, bem como alguns conjuntos de coleções disponibilizadas com o PEx. Dessa forma, a idéia é verificar se a utilização do procedimento (2) alterou a qualidade da visualização gerada pelo sistema e, desta forma, inferir sobre a qualidade do procedimento de indexação utilizado em (2), quando comparado com os métodos de indexação utilizados em (1).

ABSTRACT

It can be argued that the processes of organization, retrieval and visualization of information are conected, presenting, at least, one point in common: the indexing of texts or documents. Indexing, an essential component of text analysis, is important to information retrieval and information visualization. **In this context, we highlight solutions that aim to automate the indexing process.** The research that proposes solutions for automatic indexing are developed based on different theoretical assumptions as, for example: Statistics, Linguistics and controlled vocabularies (Leiva 1999). Most of these solutions develop hybrid models combining two or three of these theoretical assumptions. Other approaches to the problem of automatic indexing use, besides the aforementioned assumptions, theories that allow the treatment of uncertainty, imprecision and vagueness. The aim of this paper is to argue the theoretical potential of paraconsistent logic, a nonclassical logic, capable of dealing with situations involving uncertainty, imprecision and vagueness, in automatic indexing.

Referências

BAEZA-YATES R.; RIBEIRO-NETO B. *Modern Information Retrieval*, New York: Acm Press, 1999.

BELKIN, N. Interaction with Texts: information retrieval as information-seeking behavior, *Proceedings of the Fifth International Symposium for Information Science*, 1993
<<http://citeseer.ist.psu.edu/belkin93interaction.html>>.

BORDOGNA G.; PASI, G. Controlling information retrieval through a user adaptative representation of documentos, *International journal of approximate reasoning*, 1995, v.12, p. 317-339.

BORNER, K.; CHEN, C.; BOYACK, K. Visualizing Knowledge Domains. *Annual Review of Information Science and Technology*(ARIST), 2003, v. 37, p. 179–255.

CHEN, C. *Information Visualization: Beyond the Horizon*, Springer, 2006.



- CRESTANI F.; PASI G. Handling vagueness, subjectivity, and imprecision in information access: an introduction to the special issue. *Information Processing and Management*, 200339, p. 161-165.
- DA COSTA N. C.; ABE J. M.; DA SILVA FILHO, J. I.; MUROLO A. C.; LEITE C. F. S. *Lógica paraconsistente aplicada*, São Paulo:Atlas,1999.
- FERNEDA E. *Recuperação de informação: Análise sobre a contribuição da Ciência da Computação para Ciência da Informação*. Tese (Doutorado) – Curso de Ciências da Comunicação, Departamento de Biblioteconomia e Documentação, Universidade de São Paulo, São Paulo, 2003
- KOWALSKI G. J.; MAYBURY M. T. *Information storage and retrieval systems – theory and implementation*, Norwell: Kluwer Academic Publishers, 2000.
- LALMAS M.; RUTHEVEN I. Representing and retrieving structured documents using the Dempster-Shafer theory of evidence: modelling and evaluation, *Journal of documentation*, 1998, 54, n. 5, p.529-565.
- LANCASTER F. W. *Indexação e Resumos: teoria e prática*, Brasília: Briquet de Lemos, 2004.
- LEIVA I. G. *La automatización de la indización de documentos*, Gijón: Ediciones Trea, 1999.
- MOLINARI A.; PASI G. A fuzzy representation of HTML documents for information retrieval systems. *Proc. of IEEE International Conference on fuzzy systems*, New Orleans, 1996, p.107-112.
- OGAWA Y.; MORITA T.; KOBAYASHI K. A fuzzy document retrieval system using the keyword connections matrix and a learning method, *Fuzzy sets and systems*, 1991, v. 39, p.163-179.
- PAULOVICH, F.; OLIVEIRA, M. C. F.; MINGHIM, R. The Projection Explorer: A Flexible tool for projection-based multidimensional visualization, *Proc. of XX Brazilian Symposium on Computer Graphics and Image Processing*, 2007, p. 27-34.
- PAULOVICH, F. V.; NONATO, L. G.; MINGHIM, R.; LEVKOWITZ, H. Least Square Projection: A Fast high-precision multidimensional projection technique and its application do documet mapping. *IEE Transactions on Visualization and Computer Graphics*, v. 14. n. 2, March/April, 2008, p. 1-12.
- SALTON G. *Automatic text processing – the transformation, analysis and retrieval of information by computer*, New York: Addison-Wesley Publishing Company, 1989.
- SUCOLOTTI A. A., *Recuperação de informação em bases textuais: uma abordagem baseada em lógica paraconsistente*. Dissertação (Mestrado) – Programa de pós-graduação em Ciência da Computação, Universidade Federal de Santa Catarina, Florianópolis, 2001.
- WILSON T. D. Information behaviour: an interdisciplinary perspective. *Information Processing and Management*, 1997, v.33, p. 551-572.
- ZHU B.; CHEN H. Information Visualization, *Annual Review of Information Science and Technology* (ARIST), 2005, v. 39, p. 139–177.