

Comunicação científica e ontologias: uma pesquisa no Library and Information Science Abstracts

Scientific communication and ontologies: a research at Library and Information Science Abstracts

Gleisy Regina Bóries FACHIN¹

Raimundo Nonato Macedo dos SANTOS²

Rosângela Schwarz RODRIGUES²

RESUMO

Este artigo relata análise de busca realizada na base de dados da *Library and Information Science Abstracts*, com os termos comunicação científica e as ontologias. O objetivo foi analisar bases de dados referenciais, com ênfase na padronização dos metadados e a interoperabilidade entre sistemas. A metodologia é exploratória e descritiva, documental e quantiquantitativa, usando os softwares bibliométricos Infotrans e Dataview para tratamento dos 129 resultados obtidos no levantamento bibliográfico para tabulação e análise dos dados. Conclui que as bases de dados diferem significativamente entre si, o que implica a necessidade de padronização e de rigor na inserção de metadados para viabilizar resultados eficientes e precisos nas buscas realizadas e o uso de recursos bibliométricos para a análise e discussão dos documentos obtidos, refinando os indicadores quantitativos da produção científica.

Palavras-chave: Comunicação científica. Ontologia. Ferramentas bibliométricas.

ABSTRACT

This article reports a search's analysis at the database *Library and Information Science Abstracts*, using the terms scientific communication and ontology. The objective was to analyze search's possibilities in reference databases with emphasis on metadata standards and interoperability among systems. The methodology is exploratory and descriptive, documental and quant-qualitative; it employs the bibliometric soft wares Infotrans and Dataview for the analysis of the 129 obtained results in the bibliographic survey of data tabulation and analyses. It concludes that the databases and soft wares differ significantly, fact which demands strict standards of metadata to enable efficient and accurate results in searches as well as the utilization of bibliometric tools for analysis and discussion of obtained documents, in order to improve the quantitative indicators of scientific production.

Keywords: Scientific communication. Ontology. Bibliometric tools.

¹ Doutoranda, Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento e Professora da Universidade Federal de Santa Catarina, Departamento de Ciência da Informação. Campus Universitário, Trindade, 88010-970, Florianópolis, SC, Brasil. Correspondência para/Correspondence to: G.R.B. FACHIN. E-mail: <gleisy@cin.ufsc.br>.

² Professores Doutores, Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Ciência da Informação, Departamento de Ciência da Informação. Florianópolis, SC, Brasil.
Recebido em 14/12/2009 e aceito para publicação em 20/4/2010.

INTRODUÇÃO

A divulgação da informação científica evolui com a utilização de redes digitais - a Web - impulsionando o crescimento dos periódicos científicos eletrônicos. A procura por informações científicas no formato *on-line* é crescente nas áreas técnica, científica e acadêmica, em todos os países. Pesquisadores, professores e alunos das universidades e centros de pesquisas, bem como de empresas e organizações de áreas produtivas, comerciais e governamentais necessitam publicar os resultados de suas pesquisas, de seus desenvolvimentos, como meio de disseminação de suas descobertas, contribuindo na construção do saber científico. Já na área acadêmica, é necessária a disseminação da produção científica como critério de avaliação para as universidades e centros de pesquisas, supervisionada pelas instituições de fomento em cada país.

Para Fachin e Hillesheim (2006), Gomes (2006) e Oliveira (2008), a informação científica, em formato eletrônico, contribui na solução das exigências de usuários de informação. Nesse formato digital, os periódicos científicos apresentam rapidez e eficiência na troca e recuperação de informações relevantes e, em muitos casos, oficialmente publicados, indexados e citados entre os pares, contribuindo com o acelerado crescimento da ciência e da tecnologia. Da mesma forma, a expansão de informações científicas no formato *on-line* requer a adoção de recursos técnicos dos ambientes de redes eletrônicas e sua constante atualização. Não só de *softwares*, mas de todos os dispositivos que nela se abrigam para a produção, o armazenamento e a disseminação de informações, as quais servem como alavancas para a evolução do conhecimento científico (Meadows, 2000).

Com a facilidade de acesso e de publicação de informações na Web, muitos recursos foram utilizados de forma personalizada, criando um emaranhado de tipos, normas e padrões, dificultando a recuperação eficiente de informações relevantes. Questões como veracidade, confiabilidade, qualidade e autoria são recorrentes nas informações recuperadas na Web. A credibilidade atribuída a páginas, sites, portais e até mesmo às bases de dados são questionáveis, impulsionando pesquisas, instituições e centros de pesquisas a estudar formas de propiciar uma recuperação eficaz na Web, como discutido no estudo feito por Santos *et al.* (2006, p. 12), quando apresentam

o resultado da análise cienciométrica da produção científica de duas áreas científicas brasileiras “a produção de indicadores necessita de procedimentos prévios de consistência, de modo a adequá-los às análises bibliométricas. Indicou, por consequência, a necessidade de investimento na formação de recursos humanos aptos a produzir bases de dados para as diferentes finalidades acima especificadas”.

Desse modo, este estudo volta-se para a comunicação científica e a necessidade de padronização nos metadados - elementos essenciais - para a recuperação de informações relevantes e eficazes. Apresenta informações científicas de ambientes que necessitam ser organizados e padronizados visando à recuperação rápida, precisa e eficaz e, mais ainda, interoperáveis entre si. Apresenta também os procedimentos de acesso, utilização dos recursos de busca para seleção dos descritores autorizados, recuperação e constituição do *corpus* objeto do estudo, definição e formatação dos metadados e obtenção do arquivo com os registros bibliográficos da base de dados *Library and Information Science Abstracts* (LISA). Conclui-se o estudo com a análise bibliométrica dos registros bibliográficos por meio de *softwares* especializados - Infotrans e Dataview -; com as interpretações das representações gráficas dos resultados para as quais foram utilizados *softwares* estatísticos de Visualização de Informação.

INFORMAÇÕES CIENTÍFICAS E SEUS AMBIENTES

No contexto das Tecnologias de Informação e Comunicação (TIC) voltadas ao processo de disseminação da produção científica, disponibilizada pelas bases de dados tradicionais a partir da segunda metade do século passado; da explosão da informação em suporte digital; dos recursos tecnológicos como os sites especializados, bibliotecas digitais, repositórios e portais institucionais ou temáticos, bem como, de maneira mais informal, as listas de discussão e os blogs, percebe-se, com raras exceções, que os recursos de recuperação de informação científica propostos pelas fontes ditas “especializadas” de informação constituem-se interface de oferta precária de documento, não obedecem aos padrões de representação dos metadados, tanto do ponto de vista da representação descritiva e, pior ainda, quanto do ponto de vista da representação semântica. Tamanho descaso

compromete a eficácia econômica desses sistemas e impossibilita a recuperação e a disponibilização de informações relevantes e exaustivas para aqueles que delas necessitam para a realização de estudos de revisão bibliográfica, de compreensão de domínios de conhecimento, de produção de mapas conceituais, particularmente requeridas para fins de formulação de indicadores, de gestão e de elaboração de políticas institucionais e públicas em ciência e tecnologias.

Tendo em vista a necessidade premente para reversão desse quadro, a comunidade da área da comunicação científica desenvolve pesquisas que visam à padronização e disseminação eficaz das informações e, mais que isso, à busca da interoperabilidade entre os sistemas. Assim, domínios do conhecimento como Ciência da Informação - de linguagens documentárias, de estudos bibliométricos, padronizações - e da Ciência da Computação - sistemas de informação, inteligência artificial e recuperação da informação e desenvolvimento de ontologias - são exemplos de trabalhos e busca de integração entre as próprias áreas e pesquisas.

Rodrigues *et al.* (2007) comentam essa forma inovadora de disseminar as produções científicas: sejam repositórios ou portais institucionais como sistemas de informação que armazenam, preservam, divulgam e dão acesso à produção intelectual de comunidades universitárias e/ou de grupos de pesquisa. Esses repositórios intervêm em duas questões estratégicas: contribuem para o aumento da visibilidade e do valor público das instituições, servindo como indicador tangível da sua qualidade, e contribuem para a expansão do sistema de comunicação científica, ampliando e flexibilizando o acesso aos resultados das investigações e reassumindo o controle acadêmico sobre a publicação científica.

Na visão de Baroni (2005), o uso de portais permite a integração, colaboração e personalização baseadas na utilização de recursos de TIC e da Web. Esse autor destaca o conceito de portal como sendo uma plataforma tecnológica que permite o acesso e compartilhamento de informações por trabalhadores do conhecimento e possibilita a realização de ações independentemente da sua localização física, do formato da informação e do local em que ela está armazenada. Assim, o portal diz respeito à integração de diversos sistemas, exigindo infraestrutura tecnológica,

equipe especializada, consultores eventuais e apoio da alta direção.

Para Bailey (2005), os repositórios e os portais institucionais ganharam força devido ao movimento de acesso livre gerando mudança na indústria de publicação científica. Desde suas primeiras implementações, tanto os repositórios como os portais têm sido voltados para a organização e disseminação da informação científica.

Diante da evolução constante das TIC, dos mecanismos de tratamento, armazenagem e disseminação da informação científica, e em especial da utilização dos recursos da Web, os periódicos científicos - considerados veículos imprescindíveis da divulgação, evolução e consolidação das ciências - estão migrando para o suporte eletrônico. Em muitos casos, isso se realiza sem os ajustes necessários para que ocorra a organização, padronização e disseminação eficazes e interoperáveis das informações (Mendonça *et al.*, 2006). Experiências têm demonstrado que a recuperação exaustiva e precisa das informações só se torna possível quando autores, editores e instituições responsáveis atendem os requisitos de padronização dos pontos de acesso essenciais - os metadados - para que os agentes inteligentes (buscadores) possam recuperar informações relevantes com índices aceitáveis de precisão e de revocação (Fachin *et al.*, 2006).

A mudança do sistema de comunicação científica é favorecida com o movimento de Acesso Livre aos repositórios institucionais, ampliando a criticidade, melhorando o acesso aos resultados de pesquisas, rentabilizando o processo de investigação, reafirmando o controle dos registros do conhecimento, aumentando a competição e atenuando o monopólio dos editores comerciais em relação aos periódicos (Bailey, 2005)

Além disso, tanto os repositórios como os portais podem ser considerados como indicadores de qualidade de uma universidade e/ou uma instituição, demonstrando a relevância científica, social e econômica das pesquisas realizadas por seu corpo docente e discente. Reenen (2006) destaca dois objetivos principais para manter um repositório institucional, também aplicados aos portais: a) promover acesso livre para pesquisa institucional por meio do arquivamento livre (*self-archiving*) e; b) armazenar e preservar recursos institucionais digitais, incluindo publicações alternativas

ou literatura sem depósito obrigatório (ex: teses e relatórios técnicos).

Nesse aspecto, enfatiza-se a necessidade do estudo de linguagens documentárias, tesouros, taxonomias e, em especial, ontologias como uma solução para a recuperação eficaz de informações científicas. O próprio *World Wide Web Consortium* (W3C), que incentiva e dissemina pesquisas, ferramentas e publicações sobre a Web semântica e a construção de ontologias, destaca a importância do tratamento semântico dos conteúdos disponibilizados na Web, objetivando a recuperação de informações relevantes. Para isso conta com o arcabouço teórico e técnico da Biblioteconomia e da Ciência da Informação, que têm em sua essência o estudo das linguagens para o tratamento, organização e recuperação das informações, bem como os estudos bibliométricos e cienciométricos (Santos *et al.*, 2006).

Cabe às instituições federais de ensino, pesquisa e fomento, promover formas de coletar, organizar, tratar e disseminar suas produções científicas. No Brasil, algumas universidades públicas aderem ao movimento de Acesso Livre, publicando seus estoques de conhecimento em conformidade com as recomendações das instituições de controle e fomento do país, como o uso de softwares em arquivos abertos.

Saracevic (1999), Ferneda (2003), Robredo (2003) e Sales e Vieira (2007) estudam muitos dos recursos utilizados em mecanismos de busca que são atualizados e programados para atender a Web Semântica, como os modelos: Difuso, Fuzzi, Booleano Estendido, Especial Vetorial Generalizado, Indexação Semântica Latente, Redes Neurais e Recuperação Textual Estruturada. Além desses modelos, outros recursos são constantemente disponibilizados na Web e/ou apresentados em estudos e pesquisas, voltados para um melhor desempenho dos mecanismos de busca, objetivando uma recuperação precisa e eficaz da informação. Na visão dos autores citados, destacam-se três sistemas de busca na Web:

a) Diretórios - as informações são organizadas e classificadas hierarquicamente em categorias temáticas pelos editores responsáveis: classificação por assunto, ou melhor, temas específicos, sempre partindo das categorias mais amplas para as mais específicas;

b) Motores de busca - conhecidos também como programas de busca, mecanismo de procura, ferramenta de busca são programas que têm três funções básicas: identificar páginas da web, indexar essas páginas em um banco de dados e um mecanismo de pesquisa como interface. A maioria dos resultados é classificada pela relevância, e

c) Metabuscadores - que são ferramentas de busca que realizam pesquisas em vários motores de busca de forma simultânea. Não possuem robôs indexadores, nem base de dados própria. Esse tipo de ferramenta busca em outros motores de busca, geralmente com os quais tem parceria.

Assim, o objetivo é discutir e compreender as inúmeras ferramentas disponibilizadas para pesquisa na Internet, entre elas as supracitadas, as quais são propostas como recurso de recuperação da informação preciso e relevante, e podem ser, consequentemente, entendidas como um sistema de informação, enquadrado na dimensão conceitual de base de dados que, segundo Rowley (1994, p.66), pode ser definida como “uma coleção de registros similares entre si e que contêm determinadas relações entre esses registros”.

As bases de dados podem ser classificadas em dois tipos: base de dados de referência e base de dados fontes. As bases de dados de referência encaminham ou apontam o usuário para as bases de dados fontes, que podem ser de texto completo, de repositórios institucionais, de “quem é quem”, a fim de obter informações adicionais a respeito de algo ou de alguém ou, mesmo, conseguir o texto integral de um documento. Essas bases referenciais subdividem-se em três tipos: bibliográficas, catalográficas e referenciais (Rowley, 1994).

Por outro lado, o registro é a informação que a base contém e que diz respeito a um documento ou um item. Esse registro deveria ser padronizado e internacionalmente reconhecido, o que pode ser um problema, visto que as configurações de cada sistema, cada software ou base de dados utilizados é individualizado. A não padronização dos metadados impossibilita a interoperabilidade entre os inúmeros recursos para a recuperação da informação em âmbito global.

PROCEDIMENTOS METODOLÓGICOS

Segundo Kobashi e Santos (2006, p.2), a pesquisa bibliométrica requer que indicadores de atividade científica sejam decorrentes da

Análise de artigos publicados em revistas de prestígio internacional. Essa metodologia é, porém, recorrentemente criticada, pois, há um trabalho intelectual altamente produtivo que antecede, substitui ou é concomitante à publicação de artigos. Com efeito, cada comunidade científica tem modos próprios de dar a conhecer a sua produção.

Assim, é necessária a compreensão das ferramentas a serem utilizadas nos estudos, o entendimento do universo a ser trabalhado e dos recursos de busca a serem utilizados. No caso desta pesquisa, restringiu-se o estudo na base de dados *Library and Information Science Abstracts* (LISA), acessada por meio do Portal de Periódicos da Coordenação de Pessoal de Nível Superior (CAPES). A escolha da LISA deve-se a sua representatividade para a área da Ciência da Informação, além da interoperabilidade com os recursos de busca que oferece para a identificação, acesso, seleção, formatação de metadados e obtenção/transferência de registros bibliográficos. Os mapas conceituais e demais representações gráficas obtidas oferecem um modelo do que pode ser considerada uma metáfora aproximada da memória científica coletiva dos pesquisadores que estudam a temática objeto deste trabalho (Sayão, 1996).

A base LISA é uma base de dados de referência. Após análise dos recursos disponíveis, iniciou a investigação no tesauro, verificando os descritores

adotados: 'comunicação científica' e 'ontologias'. O primeiro assunto não consta na base LISA, e o segundo termo aparece no singular e no plural. A análise da estrutura hierárquica do tesauro da temática gerou os termos apresentados no quadro 1, a seguir.

Percebe-se que o descritor 'comunicação científica', tão largamente utilizado por inúmeras áreas do conhecimento, em seus respectivos idiomas e bases, é um termo não autorizado na base LISA. Nessa base, o termo autorizado para representar o conceito é '*information communication*'.

Observações dessa natureza corroboram o entendimento de que as bases de dados não são neutras e, portanto, faz-se necessário avaliar e estudar as especificidades das bases de dados, pois cada uma tem as suas particularidades e aí está um dos maiores problemas para recuperação eficaz da informação. Segundo Saracevic (1999), Robredo (2003) e Sales e Vieira (2007) o processo de recuperação da informação é essencial e requer atenção para a evolução da ciência.

Ferneda (2003, p.110), ao analisar a importância da interação entre as áreas da Ciência da Computação e da Ciência da Informação, consideradas como alicerce na construção de estruturas de recuperação de informações dotadas de semântica, corrobora esse princípio, argumentando: "procura-se aumentar a eficiência dos mecanismos de busca e de outras ferramentas de processamento automático de documentos através da utilização de linguagens que permitam definir dados e regras para o raciocínio sobre estes dados".

Dentro dessa perspectiva, a pesquisa enfatiza a necessidade da interação entre os recursos de busca, além da padronização dos metadados, o que justifica

Quadro 1. Relação dos descritores - Base de dados LISA.

Assunto/descritor	Subject/descriptor	Tesauro LISA
Comunicação científica	<i>Scientific communication</i>	<i>Information communication</i>
Produção científica	<i>Scientific production</i>	<i>Scientific production</i>
Informação científica	<i>Scientific information</i>	<i>Scientific information</i>
Periódicos científicos ou	<i>Scientific journals</i>	<i>Electronic publishing</i>
Revistas científicas	<i>Scientific periodicals</i>	<i>Electronic periodicals</i>
Ontologia	<i>Ontology</i>	<i>Ontology</i> <i>Ontologies</i>

Fonte: Baseado no tesauro da base de dados LISA.

observar a constatação feita quando da análise dos termos autorizados de busca: o uso desses termos no singular ou plural, como o caso do descritor “ontology e ontologies” implica particularidades relevantes; a não observância desse aspecto pode tornar irrecuperáveis informações importantes ou requerer ajustes na hora de estruturar as sintaxes de busca.

Considerações e constatações dessa ordem foram preponderantes para a formulação da estratégia de busca pela opção *advanced search*, dando ênfase aos elementos da lógica booleana utilizada pelo sistema de recuperação da LISA, e desenvolvida conforme segue: “Search Query DE= (“ontologies” or “ontology” and “electronic periodicals” or “electronic publishing” and “information communication”)”. O resultado obtido é apresentado no Anexo 1.

Foram obtidas 218 referências no total. O período de cobertura da busca compreende o ano da criação da base (1969) até o presente, como forma de garantir exaustividade do *corpus* sobre a temática objeto deste estudo. Ainda para a composição final do estudo, utilizou-se de mais um recurso da base, ou seja, selecionar somente material cujo artigo que lhe deu origem foi submetido a revisão pelos pares, restando assim, dos 218 documentos inicialmente obtidos, 129 referências.

A utilização exaustiva e precisa dos recursos de busca “search tools” colocados à disposição dos usuários da LISA pelo Portal CAPES, tais como: tesouros, histórico de busca, período de cobertura, revisão por pares permite assegurar que as informações obtidas são relevantes, precisas, legitimadas pelos pares. Assim, ressalvadas as devidas proporções do contexto do interesse comercial da LISA, são essas possibilidades oferecidas pela base que conformam a representação aproximada da memória coletiva do conhecimento científico da comunidade de pesquisa da temática, segundo o conceito de Halbwachs (1990) citado por Sayão (1996).

Para responder aos questionamentos propostos neste estudo, os metadados essenciais para recuperação e análise das referências foram exportados da base LISA em formato texto.

Conforme destacam Santos *et al.* (2006, p.3)

A produção de indicadores da ciência, por métodos bibliométricos, requer um conjunto de dados padronizados, sistematizados e consistentes, em princípio encontráveis nas bases de dados bibliográficas. Essas fontes favorecem as abordagens bibliométricas em razão da forma em que os dados se encontram estruturados. No entanto, na prática, as inconsistências encontradas impedem sua utilização direta, sendo necessário realizar um trabalho prévio de consistência e reestruturação das bases. Para isso, recorre-se a procedimentos e softwares bibliométricos criados especificamente para promover a consistência requerida.

Em atendimento aos requisitos acima recomendados, os metadados essenciais: AB = *Abstract* (resumo); AU = *Author* (autor); DE = *Descriptor* (descritor); KW = *Keywords* (palavra-chave); LA = *Language* (idioma); PT = *Publication Type* (tipo de publicação); PY = *Publication Year* (ano de publicação); SO = *Source* (fonte) e TI = *Title* (título) importados da Base LISA foram submetidos, previamente, a softwares de reformatagem (INFOTRANS³) e tratamento de texto (WORD) a fim de se ter assegurada a consistência dos dados, para posteriores estudos bibliométricos.

Com base no trabalho de Santos *et al.* (2006) e Kobashi e Santos (2006), desenvolveu-se a análise dos dados pesquisados na base LISA e tratados por meio dos softwares utilizados pelos autores:

a) Dataview, software desenvolvido pelo Centre de Recherche Retrospective de Marseille da Université de Aix-Marseille III (CRRM), na França, e vem sendo utilizando em parceria com outras universidades e empresas e objetiva o preparo dos dados selecionados e salvos para o tratamento estatístico, ou seja, a constituição de arquivos de listas e de matrizes de frequência, de correlações e com interfaces com softwares de tratamento de texto e de planilha eletrônica. Esse software apóia-se em métodos bibliométricos, gerando elementos para uma análise estatística

³ Infotrans é um software desenvolvido por pesquisadores alemães da IuK – *Information und Kommunikation* - utilizado para a reformatação e conversão de formatos e de dados. Foi desenvolvido e trabalha em ambiente DOS e seus arquivos são trabalhos em formato ASCII puro, ou seja, no formato de arquivos .txt (texto). Apresenta como potencialidades reconhecer cadeias de caracteres pelo seu tipo de estrutura; realiza conversão específica por registro e por campo; possibilita a conversão de dados por meio de listas externas, bem como permite a visualização do andamento da conversão, o que proporciona total controle sobre o processo.

(quantitativa). Assim, para cada unidade ou elemento bibliográfico, pode-se considerar a ocorrência de três maneiras específicas: estado primário-simples (localização das ocorrências, presença ou ausência dos elementos da referência); estado condensado (desdobramentos dessas ocorrências ou frequências) e a co-ocorrência, que representa a combinação do estado primário e do estado condensado (Kobashi, 2008).

b) MS-Excel 2003, software desenvolvido para criação de planilhas eletrônicas, em que se pode organizar informações numéricas em forma de tabelas, matrizes, realizar cálculos simples ou complexos, criar gráficos;

c) Statistic.

RESULTADOS E DISCUSSÃO

Nesta pesquisa foi possível extrair as 129 referências sobre “ontologias e comunicação científica”. Com o uso do software Infotrans, formatou-se o arquivo para texto, efetuando as organização dos dados, que demonstra o Anexo 2.

O arquivo obtido e organizado no Infotrans permite trabalhar com o software Dataview, que apresenta recursos para gerar arquivos com resultados quantitativos, conforme o Anexo 3.

No Dataview foram gerados arquivos em formato texto e planilha com análise de ocorrência e

co-ocorrência dos descritores, autores e idioma. Esses resultados foram exportados para softwares estatísticos, em que foram gerados gráficos e tabelas relativos aos assuntos do trabalho.

A figura 1 apresenta a distribuição bibliométrica dos descritores autorizados que constituem o núcleo semântico da temática ontologia. Indica que o descritor autorizado “ontologies” está presente em 126 referências; “knowledge representation” em 26, “computer applications” em 25, destacando-se como os de maior incidência e os demais como se apresentam.

Conforme se apresenta, o resultado abaixo pode estar sinalizando que os avanços tecnológicos dessa temática estão concentrados na questão das ontologias, resultado que vai ao encontro do que observou Feitosa (2005, p.209):

Desde o início dos anos 90, as ontologias tornaram-se um tópico de pesquisa popular nas áreas de inteligência artificial, engenharia do conhecimento, processamento da linguagem natural e representação do conhecimento. A razão de tal popularidade é acreditar-se que, por seu intermédio, um entendimento comum e compartilhado sobre um determinado domínio de conhecimento poderá ser comunicado tanto entre pessoas como entre computadores.

Assumindo-se ontologia como um processo, do ponto de vista da representação do conhecimento, ela não pode ser concebida apenas como um vocabulário

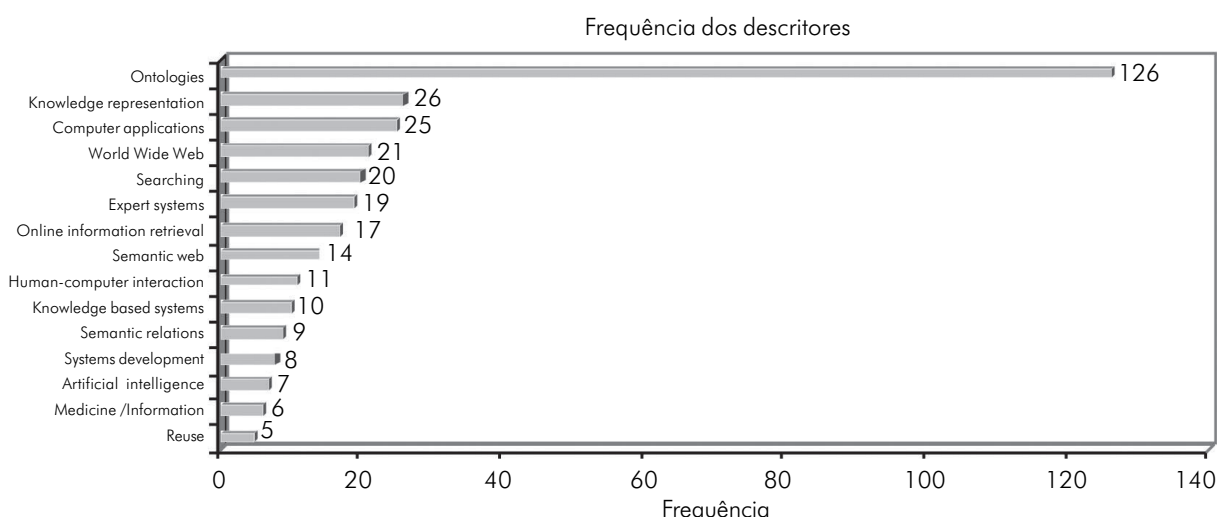


Figura 1. Descritores recuperados por frequência na base de dados LISA.

informal, ou mesmo como uma linguagem de termos estruturados, como um tesouro ou uma taxonomia, mas requer uma interpretação algorítmica dos seus significados e, por conseguinte, uma representação em uma linguagem formal, cujo processamento dos significados pode ser realizado por máquinas (Feitosa, 2005, p. 212).

Desse modo, ainda no Anexo 1, verifica-se que a tendência mundial da temática vai também ao encontro dos aspectos tecnológicos, comprovadamente com a utilização das ontologias em todas as áreas do conhecimento, pois, Lassila (1997, p.2) já comentava que:

[...] uma solução possível para o problema da busca e o uso de “agentes automatizados” na Web semântica é a criação de um mecanismo que permita uma descrição mais precisa e conceitual das coisas que estão na Web. Isto, por sua vez, poderia elevar o status da Web de legível-por-máquina para algo que poderia chamar de compreensível-por-máquina.

Tal observação é reeditada dez anos depois, quando Baptista e Ferreira (2006) afirmam que o uso das ferramentas disponíveis na Web, voltadas às comunicações acadêmicas, facilitam o contato entre pesquisadores e entre áreas do conhecimento, tendo que a própria Web aparecido em um ambiente acadêmico, de investigação.

A explicitação das manifestações acima apresentadas demonstra que a adoção das ontologias está presente em todos os campos do saber. Aplicadas com conceitos amplos e relacionáveis para as informações contidas na Web, permite a interoperabilidade entre os diversos sistemas de informação e os demais recursos tecnológicos disponíveis.

Com a criação da Web Semântica e o uso de ontologias, várias áreas do conhecimento, entre as quais a Computação e a Ciência da Informação, voltam-se para pesquisas e desenvolvimentos de ontologia e o aperfeiçoamento e adequação de linguagens documentárias, entres as quais as classificações, indexações, tesouros e taxonomias, possibilitando o uso da semântica, visando a uma recuperação eficaz pelos agentes automatizados.

Parafraseando Gruber (1993), uma ontologia na Internet é a representação de uma conceitualização,

um conjunto de conceitos estudados e especificados de uma determinada área de domínio. Guarino (1998), esclarece-se o papel de uma ontologia na Internet, a qual se pode considerar como um conjunto de axiomas lógicos, concebido para ter em conta o significado de um vocabulário específico, ou seja, destinado a uma área única do conhecimento.

Já para Morshed e Singh (2005), a ontologia é essencial para compartilhar conhecimento usando o vocabulário, a semântica e os relacionamentos similares entre conceitos de um domínio. Complementa-se, ainda, que podem ser usados também os tesouros e as taxonomias, abordados semanticamente e migrados para as linguagens específicas para a Web Semântica, como XML (*Extensible Markup Language*); RDF (*Resource Description Framework*), DAML+OIL (*DARPA Agent Markup Language + Ontology Interchange Language*), sendo a mais referenciada atualmente a OWL (*Web Ontology Language*), entre outras linguagens.

Essas afirmações nos levam a crer que os recursos tecnológicos são parte integrante de todas as áreas do conhecimento humano, assim como a necessidade de quebras de paradigmas e de domínios – em que as pessoas necessitam de uma visão maior das evoluções em suas áreas de atuação, bem como do domínio de novos recursos e idiomas -, apontando para o uso da língua inglesa, reconhecida como universal na academia. Das 129 referências recuperadas, 109 são de artigos redigidos em inglês e 20 em chinês (destacado junto aos títulos dos artigos com a expressão [original em chinês]), o que demonstra uma presença muito significativa desse idioma na evolução do assunto “ontologias e comunicação científica”, na base analisada.

Analisando o item autoria dos trabalhos, observa-se que os 129 artigos foram produzidos em coautoria, ou seja, com mais de um autor. A Figura 2 apresenta a distribuição bibliométrica dos autores, destacando-se HUI, Dong, com 3 artigos, todos em coautoria, como o mais profícuo.

No estudo de correlação (Tabela 1) autores versus descritores autorizados -, verificam-se as frequências de correlação idênticas às dos autores mais profícuos, sinalizando o foco dentro de suas temáticas.

Analisando essas correlações, observa-se que um autor teve três publicações com o descritor ‘ontologies’; 31 autores tiveram duas ocorrências

relacionados com descritores que envolvem recursos tecnológicos associados a comunicação científica. Esse tipo de resultado corrobora os estudos apresentados por Barradas e Targino (2008, p. 18), onde se verifica que a “expansão da ciberescritura, a proliferação da

autoria compartilhada e a adesão crescente ao sistema *copyleft* são alguns dos muitos elementos responsáveis por mudanças radicais no processo de editoração, independentemente do suporte impresso ou eletrônico.”

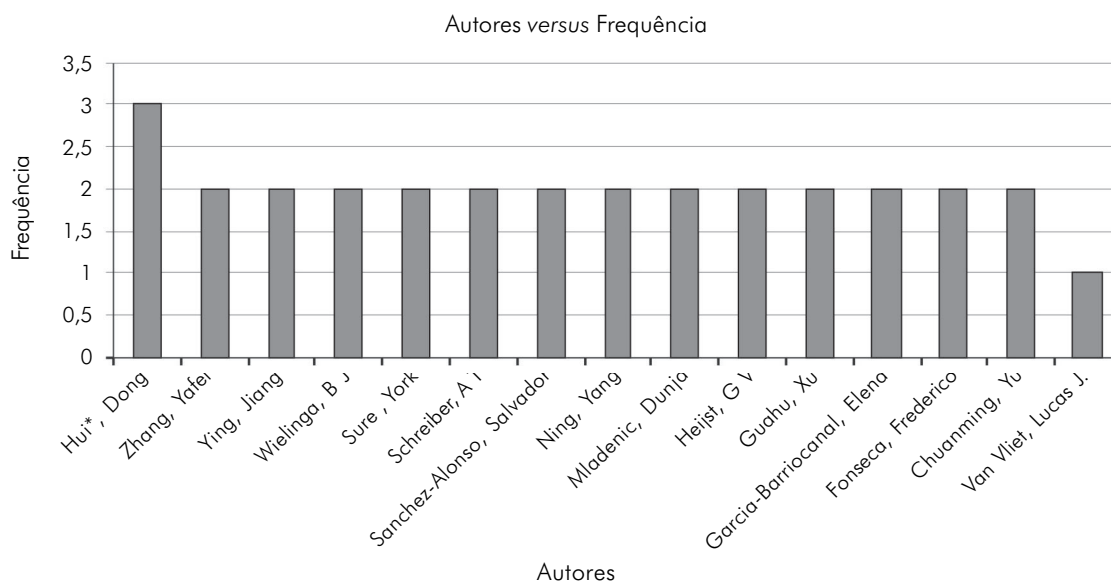


Figura 2. Frequência de autores e artigos publicados na LISA.

Tabela 1. Frequência de autor e descritor.

Pair	Descriptor	Frequency
Hui*, Dong	<i>Ontologies</i>	3
Zhang, Yafei	<i>Ontologies</i>	2
Ying, Jiang	<i>Searching, Ontologies, Digital libraries</i>	2
Wielinga, B J	<i>Systems development, Ontologies, Knowledge based systems</i>	2
Sure, York	<i>Ontologies, Knowledge management</i>	2
Schreiber, A T	<i>Systems development, Ontologies, Knowledge based systems</i>	2
Sanchez-Alonso, Salvador	<i>Ontologies</i>	2
Ning, Yang	<i>Searching, Ontologies, Digital libraries</i>	2
Mladenic, Dunja	<i>Ontologies</i>	2
Hui*, Dong	<i>Searching, Digital libraries</i>	2
Heijst, G V	<i>Systems development, Ontologies, Knowledge based systems</i>	2
Guohu, Xu	<i>Searching, Ontologies, Digital libraries</i>	2
Garcia-Barriocanal, Elena	<i>Ontologies, Computer applications</i>	2
Fonseca, Frederico	<i>Ontologies</i>	2
Chuanming, Yu	<i>Searching, Ontologies, Digital libraries</i>	2

Fonte: os autores.

Desse modo, pesquisas como essas apontam a necessidade de estudos e do desenvolvimento de pesquisas e projetos multidisciplinares, aproximando a Ciência da Informação e a Ciência da Computação e suas correlatas, pois o uso de ontologias está presente e constitui solução possível em várias áreas, sejam teóricas ou de aplicações. Por outro lado, podem, ainda, instigar os pesquisadores a buscar por mais recursos tecnológicos que demonstrem as facilidades na troca de saberes, como o uso de softwares estatísticos, nesse caso, o *Matrisme* (C.R.R.M - Universidade Aix-Marseille III e Lepont – Toulon, França), que permite obter relações e indicadores gráficos a partir de listas de frequências e matrizes de vários tipos, também baseado na aplicação de conceitos bibliométricos (Hoffmann *et al.*, 2001).

O dendograma (Figura 3) apresenta o resultado de uma análise da estrutura hierárquica entre os

descritores autorizados que guardam proximidade semântica com o descritor 'ontologia' objeto de outro corpus também recuperado nos mesmos critérios metodológicos aplicados ao descritor "ontologies" e, particular da Base LISA. Essa forma de representação constitui uma construção ordenada de aglomerados semânticos "cluster". Assim, pode-se deduzir que, na configuração apresentada, "ontology" guarda uma proximidade semântica com "artificial intelligence" maior do que ambos com "semantic analysis", ao mesmo tempo em que os três descritores autorizados constituem um aglomerado semântico "cluster".

A representação acima, comparada com uma representação idêntica a partir de descritores autorizados que guardam proximidade semântica com a temática "ontologias", sinaliza que os descritores aqui apresentados guardam relações de similaridade com

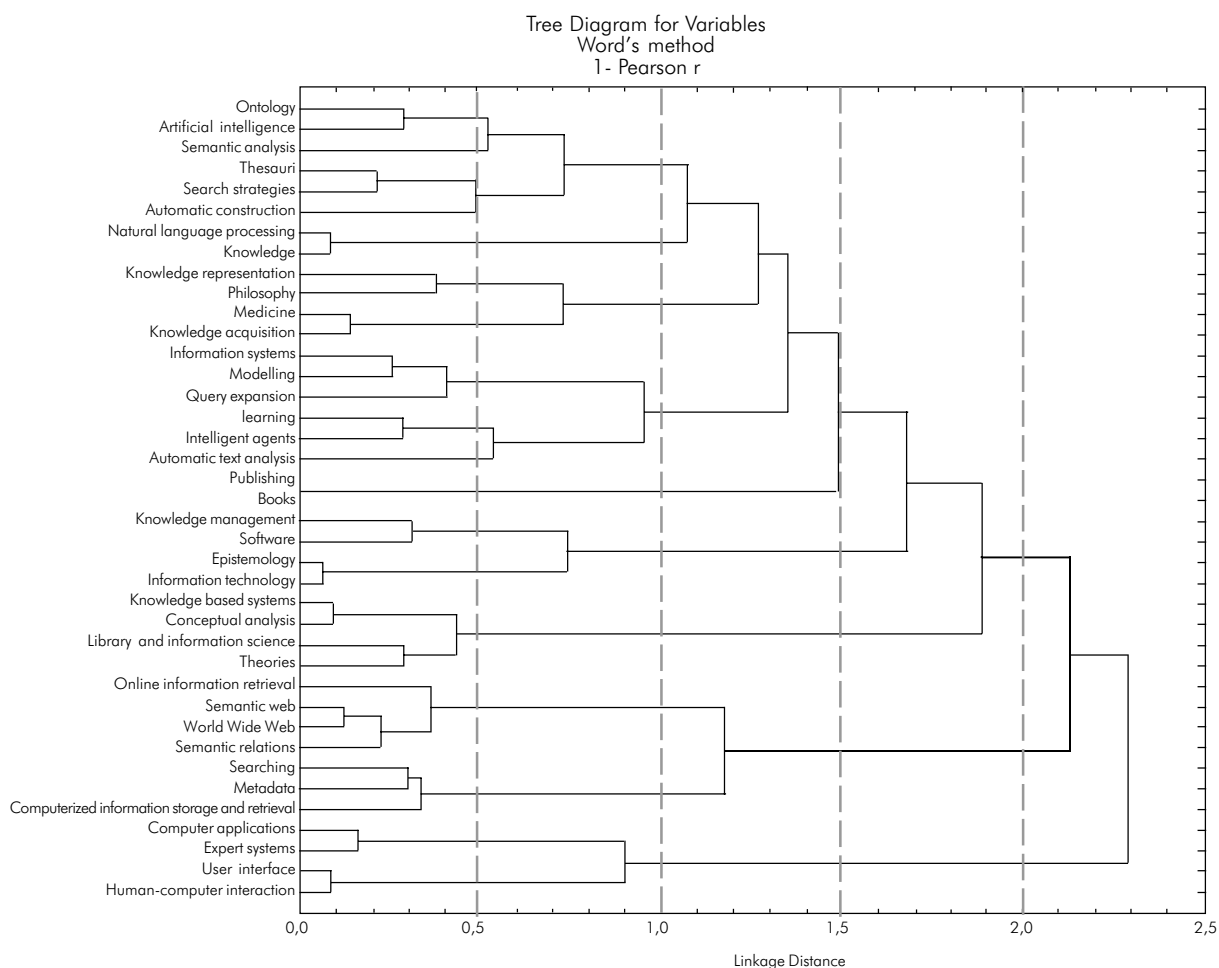


Figura 3. Dendograma que representa os descritores autorizados que guardam proximidade semântica com o descritor "ontologia" em um corpus da base LISA.

temas mais de ordem teórica e da própria epistemologia da ciência ontologia, enquanto, no outro caso, voltam-se mais para questões associadas a aplicações tecnológicas em diversas áreas.

CONSIDERAÇÕES FINAIS

Os resultados aqui obtidos demonstram o quanto se faz necessária a adoção rigorosa de padrões pelas fontes de informação científica disponibilizadas em qualquer tipo de suporte, particularmente pela rede mundial de computadores, em todas as áreas do conhecimento. Isso é relevante para toda a estrutura tecnológica existente e em constante desenvolvimento, pois a falta de integração e interoperabilidade entre os sistemas é fator de descrédito e de ineficiência na recuperação de informações relevantes em todas as áreas científicas nacionais e internacionais, públicas ou privadas.

Dessa forma, informações precisas e eficazes fazem a diferença nos contextos educacionais, científicos, econômicos e sociais e tanto instituições públicas e privadas buscam pela qualidade nos sistemas de informação. E, assim, o uso de recursos computacionais é necessário e, mais que isso, esse uso deve ser compatível com uma estrutura que permita o uso por toda a sociedade, em suas áreas específicas.

O uso de recursos tecnológicos, como os grandes sistemas de informações, bases de dados e softwares específicos para o tratamento, organização, armazenagem a recuperação eficiente e precisa são as necessidades prementes em todas as áreas do conhecimento, especificamente as áreas científicas e tecnológicas. Assim, esta pesquisa iniciada destaca a relevância do uso de recursos tecnológicos bibliográficos, bibliométricos e a necessidade de estudos sobre esse uso além da sua integração para a obtenção de resultados eficazes em toda pesquisa.

Os resultados ora apresentados são apenas quantitativos e demonstram que cada base de dados trata seus conteúdos de forma particularizada e isso requer a análise individual de cada uma. Isso reforça a necessidade da integração e da interoperabilidade entre os sistemas de informação e a adoção de linguagens universais, inclusive entre bases de dados de diferentes áreas do conhecimento.

Conclui-se que, na atual estrutura existente, são necessários o estudo e o aperfeiçoamento em cada fonte de informação, em sua área específica de atuação, mais os recursos de softwares e o conhecimento de sistemas específicos para o trabalho de investigação. A identificação, arquivamento, análise e divulgação de resultados eficazes e com qualidade em determinado assunto permitiriam avanços, adequações e compartilhamento entre as ciências.

REFERÊNCIAS

BAILEY, C.W. Jr. *Open access bibliography: liberating scholarly literature with e-prints and open access journals*. Washington: Association of Research Libraries, 2005. p.17. Available from: <<http://www.escholarlypub.com/oab/oab.htm>>. Cited: 28 May 2006.

BAPTISTA, A. A.; FERREIRA, M. Conversas de repositório. *Brazilian Journal of Information Science*, v. 0, n. 0, p. 50-64, 2006. Disponível em: <<http://www.portalppgci.marilia.unesp.br/bjis/>>. Acesso em: 16 dez. 2007.

BARONI, R. Avaliando os efeitos dos portais corporativos em iniciativas de gestão de conhecimento. In: CONFERÊNCIA ONLINE E WORKSHOP: GESTÃO DE ORGANIZAÇÕES BASEADAS EM CONHECIMENTO, 2005, Lisboa. *Anais eletrônicos...* Disponível em: <<http://www.aquifolium.com/apgc2005/>>. Acesso em: 20 maio 2008.

BARRADAS, M.M.; TARGINO, M.G. Redação de artigo técnico-científico: a pesquisa transformada em texto. In: FERREIRA,

S.M.S.P.; TARGINO, M.G. (Org.). *Mais sobre revistas científicas: em foco a gestão*. São Paulo: SENAC, 2008. p.17-40.

FACHIN, G.R.B.; HILLESHEIM, A.I.A. *Periódico científico: padronização e organização*. Florianópolis: Ed. da UFSC, 2006.

FACHIN, G.R.B et al. Publicação periódica: revendo padrões de publicação e avaliação de artigos. In: CONFERÊNCIA IBEROAMERICANA DE PUBLICAÇÕES ELETRÔNICAS NO CONTEXTO DA COMUNICAÇÃO CIENTÍFICA, 1., 2006, Brasília. *Anais eletrônicos...* Brasília: UnB, 2006. Disponível em: <<http://portal.cid.unb.br/cipecbr>>. Acesso em: 9 abr. 2007.

FEITOSA, A.L.G. *A integração entre sistemas legislativos, terminologia e web semântica na organização e representação da informação legislativa*. 2005. 405 f. Tese (Doutorado em Ciência da Informação) - Departamento de Ciência da Informação e Documentação - Universidade de Brasília, Brasília, 2005.

FERNEDA, E. *Recuperação da informação: análise da contribuição da ciência da computação para a ciência da*

informação. 2003. 147. Tese (Doutorado em Ciência da Informação e Documentação) - Universidade de São Paulo.

GOMES, S.L.R. Os ambientes contemporâneos de informação para a comunicação do conhecimento científico: mudança e crescimento. In: CONFERÊNCIA IBEROAMERICANA DE PUBLICAÇÕES ELETRÔNICAS NO CONTEXTO DA COMUNICAÇÃO CIENTÍFICA, 1., 2006, Brasília. *Anais eletrônicos...* Brasília: UnB, 2006. Disponível em: <<http://portal.cid.unb.br/cipecbr>>. Acesso em: 9 nov. 2007.

GUARINO, N. *Formal ontology and information systems*. 1998. Available from: <<http://www.loa-cnr.it/Papers/FOIS98.pdf>>. Cited: 17 Nov. 2007.

GRUBER, T.R. A translation approach to portable ontologies. *Knowledge Acquisition*, v.5, n.2, p. 199-220, 1993.

HOFFMANN, W.A.M. et al. Monitoramento de tecnologias de materiais utilizados em anéis de pistão de motores automotivos. *RECITEC*, v.5, n.1, p.1-18, 2001.

KOBASHI, N.Y. *Curso sobre produção de informações documentárias: abordagem teórico-metodológica*. Programa de Pós-Graduação em Ciência da Informação. UFSC, 21 a 24 junho, 2008.

KOBASHI, N.Y.; SANTOS, R.N.M. Arqueologia do trabalho imaterial: uma aplicação bibliométrica à análise de dissertações e teses. In: *ENANCIB*, 7., 2006, Marília. *Anais eletrônicos...* Marília: UNESP, 2006. Disponível em: <<http://portalppgi.marilia.unesp.br/enancib/index.php>>. Acesso em: 21 set. 2009.

LASSILA, O. *Introduction to RDF metadata*. 1997. Available from: <<http://www.w3.org/TR/NOTE-rdf-simple-intro>>. Cited: 28 Jul. 2007.

LIBRARY and Information Science Abstracts (LISA). Available from: <<http://www.periodicos.capes.gov.br/portugues/index.jsp>>. Cited: 10 set. 2008.

MEADOWS, A.J. Avaliando o desenvolvimento da comunicação eletrônica. In: MUELLER, P. M.; PASSOS, J. L. (Org.). *Comunicação científica*. Brasília: Universidade de Brasília, 2000. p. 23-34.

MENDONÇA, C.; FACHIN, R. B.; RADOS, J.V. Padronização de periódicos científicos on-line. *Informação & Sociedade: Estudos*, v.16, n.1, 2006. Disponível em: <<http://www.informacaoesociedade.ufpb.br>>. Acesso em: 15 out. 2006.

MORSHED, A.; SINGH, R. Evaluation and ranking of ontology construction tools. Technical Report DIT-05-013. 2005. Available from: <<http://eprints.biblio.unitn.it/archive/00000747/01/013.pdf>>. Cited: 15 Aug. 2007.

OLIVEIRA, É.B.P.M. Periódicos científicos eletrônicos: definições e histórico. *Informação & Sociedade: Estudos*, v.18, n.2, p. 69-77, 2008. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/1701/2111>>. Acesso em: 9 set. 2008.

REENEN, J.V. Open access and connectedness: stimulating unexpected innovation through the use of institutional open archives. *Ciência da Informação*, v.35, n.2, p. 17-26, 2006. Available from: <<http://www.scielo.br/pdf/ci/v35n2/a03v35n2.pdf>>. Cited: 22 May 2007.

ROBREDO, J. *Da ciência da informação revisitada aos sistemas humanos de informação*. Brasília: Ed. Thesaurus, 2003.

RODRIGUES, E. et al. *Repositório: criação e desenvolvimento do Repositório Institucional da Universidade do Minho*. 2007. Disponível em: <http://eprints.rclis.org/archive/00007962/01/BAD_artigo_-_Final.pdf>. Acesso em: 30 maio 2008.

ROWLEY, J. *Informática para bibliotecas*. Brasília: Briquet de Lemos, 1994.

SALES, R.; VEIRA, A.F.G. Grupos e linhas de pesquisa sobre recuperação da informação no Brasil. *Biblios*, ano 8, n. 28, 2007. Disponível em: <http://www.bibliosperu.com/articulos/28/biblios_28_07.doc>. Acesso em: 9 mar. 2008.

SANTOS, R.N.M. et al. Análise cienciométrica de produção científica por meio de dissertações e teses: uma experiência brasileira. In: INTERNATIONAL CONFERENCE ON MULTIDISCIPLINARY INFORMATION SCIENCES AND TECHNOLOGIES, 1., 2006, Merida. *Proceedings...* Badajoz: Open Institute of Knowledge, 2006. v.1, p. 553-556.

SARACEVIC, T. *Information Science*. Available from: <<http://www.scils.rutgers.edu/~tefko/JASIS1999.pdf>> Cited: 9 Mar. 2008.

SAYÃO, L.F. Base de dados: a metáfora da memória científica. *Ciência da Informação*, v. 24, n. 2, p. 232-240, 1996.

ANEXOS

ANEXO 1

RESULTADO DE PESQUISA REALIADA NA BASE DE DADOS LISA

Command Search Results - Internet Explorer fornecido por Dell

http://csaweb102v.csa.com/ids70/results.php?id=2&pub_type=peer&SID=s5t42bfb1puuaeed0ic9ogrb2&mark_id=search%3A2%3A0%2C0%2C10#pubtypes

Google

Google G OK Favoritos 9 bloqueado Verificar Enviar para Configurações

Command Search Results

CSA ILLUMINA
The Death Detectives - the Ultimate Forensic Investigators

Logout Quick Search Advanced Search Search Tools Browse

Results Edit Search

History/Combine Searches Command Search Thesaurus Indexes

218 results found for: DE=(**"ontologies"** or **"ontology"** and **"electronic periodicals ... in LISA? Library and Information Science Abstracts** Alert Me

Published Works 218 Scholars 0

All Publication Types 218 Journals 218 Peer-Reviewed Journals 129 Conferences 1

Mark or Clear all on page | Update Marked List

◀ Previous 1 2 3 4 5 Next ▶

1. **Ontology-based evaluation of Organizational Memory**
Weinberger*, Hadas, Te'eni, Doy, Frank, Ariel J.
Journal of the American Society for Information Science and Technology, vol. 59, no. 9, pp. 1454-1468, July 2008
In this article we offer a new approach to evaluating Organizational Memory (OM). Our proposed evaluation methodology, named KnowledgeEco, is based on an ontology for the domain of OM. Its key steps are: 1) mapping the OM in the evaluated ...
View Record | Document Delivery

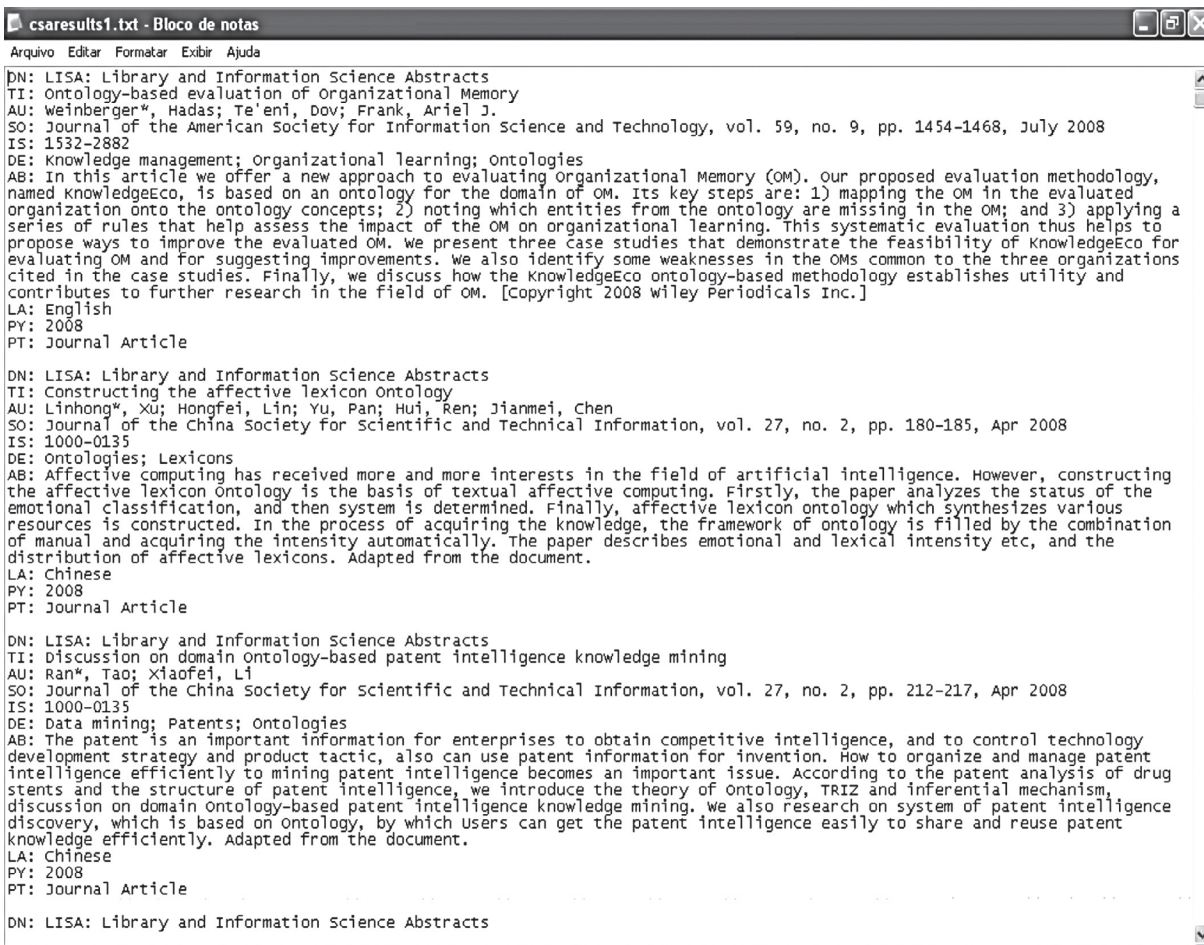
Database:
LISA: Libra
Descriptor
Knowledge
Ontologies

Concluído Internet | Modo Protegido: Ativado 130%

Fonte: Library and Information Science Abstracts (LISA).

ANEXO 2

ARQUIVO EM TEXTO ORGANIZADO UTILIZANDO O INFOTRANS



Fonte: Resultados da busca na base LISA.

DATAVIEW

