

RECUPERAÇÃO INTELIGENTE DA INFORMAÇÃO E ONTOLOGIAS: um levantamento na área da Ciência da Informação

GLEISY REGINA BORIES FACHIN*

RESUMO

Este artigo identifica a aplicação de mecanismos de Recuperação Inteligente da Informação que use Ontologia como recurso na recuperação precisa e eficaz da informação. Discorre sobre a recuperação da informação como recurso possível e concreto do encontro entre uma pergunta formulada, uma informação armazenada e o retorno preciso e eficaz ao usuário. Igualmente, disserta sobre ontologia como a representação de uma conceituação, de um conjunto de conceitos estudados, analisados e especificados sobre uma determinada área de domínio. Efetua-se uma investigação no Portal de Periódicos da CAPES, visando a uma expressão de busca “recuperação inteligente da informação e ontologia” no contexto da Ciência da Informação, apresentando resultados e discussão. Concluiu-se que a recuperação foi positiva em quatro artigos, confirmando a necessidade de implementação e atualização dos mecanismos de buscas em portais, diretórios e base de dados e futuros estudos.

PALAVRAS-CHAVE: Recuperação Inteligente da informação; Ontologia; Recuperação da informação; Ciência da Informação – pesquisa; Ciência da Computação – pesquisa.

ABSTRACT:

This article identifies the application of mechanisms of Intelligent Information Retrieval that uses Ontology as resource in the necessary and efficient information retrieval. It discourses on the information retrieval as possible and concrete resource of the meeting between a formulated question, stored information and the necessary and efficient return to the user. Equally, it discourses on ontology as the representation of a conception, a set of concepts studied, analyzed and specified on a given domain. An inquiry in the “*Portal de Periódicos da CAPES*”, investigating an expression of search “intelligent information retrieval and ontology” in the context of the Science of Information is conducted, presenting results and discussion. It concludes that the retrieval was positive in four articles, confirming the necessity of implementation and update of the mechanisms of search in “*portais*”, directories and database and further studies.

KEY WORDS: Intelligent information retrieval; Ontology; Information retrieval; Science of Information – research; Computer Science – research.

1 – INTRODUÇÃO

Desde as primeiras inscrições e desenhos rupestres feitos pelo homem, a necessidade de localizar, entender e disseminar está presente na evolução da civilização humana. Em cada época, a história é representada de alguma forma, possuindo características próprias, tanto de criação como de recuperação e de conservação de suas informações e evoluções. Da mesma maneira, a organização de documentos e de informação acompanha a evolução da sociedade e esteve centrada na área da Biblioteconomia, da Ciência da Informação e da Computação, desde as primeiras bibliotecas até as mais recentes bases de dados.

A necessidade de recuperar a informação de forma rápida, precisa e relevante remonta à própria história da humanidade. Vannevar Bush, já em 1945, lançava a idéia da automatização da informação e de instituições que atuam na organização da informação. O período de 1945 a 1970 foi marcado pela necessidade de investir em tecnologias que visando à automatização da informação e das unidades de informação. Em 1930, Shiyali Ramamrita Ranganathan divulga um sistema de classificação facetado, baseado em suas pesquisas e experiências na organização de documentos e repasse de informações. Essas pesquisas e publicações têm sido largamente discutidas na academia como uma solução para a organização do conhecimento, em decorrência de suas potencialidades de acompanhar as mudanças e a evolução do conhecimento. Nas décadas de 80 e 90, com as tecnologias de informação e comunicação surgindo como ferramentas eficazes no tratamento, disseminação e recuperação de informação, possibilitaram um grande avanço na área da Computação, Biblioteconomia e da Ciência da Informação. Muitos termos e expressões têm surgido e retratam nada mais do que a teoria da classificação, em especial a facetada, em que Ranganathan (1965) conceituava o conhecimento como “a totalidade das idéias conservadas pelo ser humano” por meio da observação das coisas, fatos e processos do mundo que o cerca e vivencia.

Muitos autores que serão citados ao longo do trabalho apresentam e discutem a recuperação da informação como essencial para a evolução das ciências, reafirmando sua importância na pesquisa e na comunicação científica em todas as áreas do conhecimento. Destaca-se que um dos principais problemas ou desafios dos Sistemas de Recuperação da Informação (SRI) é recuperar somente documentos importantes para o usuário, ou seja, que tenham “relevância” e, infelizmente, esses SRIs possuem relevância parcial.

Cada mecanismo de busca atende a uma recuperação específica, para a qual foi programado. Assim, na maioria dos SRIs e mais

recentemente nos mecanismos de Recuperação Inteligente da Informação (RII) devido à utilização de agentes inteligentes, a programação atende à escolha entre várias opções, cabendo aos criadores investigar, analisar e utilizar recursos nos mecanismos que atendam aos objetivos da área e do público a que se destinam. Com isso, são inúmeros sistemas e recursos disponibilizados para que ocorra, de forma precisa e relevante, uma recuperação inteligente da informação.

Este artigo visa a identificar, baseando-se na literatura, a existência ou não de aplicações de mecanismos de Recuperação Inteligente da Informação que usem ontologia como recurso na recuperação precisa e eficaz da informação. Apresenta-se a seguir uma conceituação sobre RII e ontologia; a identificação de exemplos de RII usando ontologia e sua explanação, acompanhada das considerações e referências.

2 – RECUPERAÇÃO INTELIGENTE DA INFORMAÇÃO

Numa visão global, recuperação da informação é tornar possível e concreto o encontro entre uma pergunta formulada, a informação armazenada e o retorno positivo ao usuário solicitante, quer de forma manual ou automatizada/digital. Em sistemas digitais, os usuários recuperam informações através de *browsing* (navegação através de *links* de um documento ou de um *site*) ou através de *searching* (consulta a um banco de dados). Quando a pesquisa é por meio de acesso ao banco de dados, o usuário tem uma necessidade específica, já sabe o que precisa recuperar. Em um *browsing*, nem sempre o usuário sabe o que precisa; ele acessa um documento, com vários *links*, e estes o remete a outros documentos. Nessa navegação e investigação ele pode recuperar informações mais relevantes das encontradas no primeiro momento, e assim sucessivamente.

Os mecanismos de recuperação de informação são inúmeros e variados, disponibilizados em rede, muitos de acesso livre e arquivo aberto. Mas, cada mecanismo atende a uma recuperação específica, para a qual foi programado. Atualmente (2008), os mecanismos de recuperação da informação evoluíram em desempenho, devido à utilização de agentes inteligentes. As programações desses mecanismos, utilizando agentes, atendem a escolha entre várias opções, cabendo aos criadores investigar, analisar e utilizar recursos que atendam aos objetivos do mecanismo e ao público a que se destina.

O processo de recuperação da informação está baseado em coleta, representação, armazenamento, organização e acesso por parte dos usuários. De modo geral, o processo de um sistema de informação detém aspectos linguísticos e objetos textuais, portanto necessita de

interpretação correta dos elementos envolvidos, o que garante uma recuperação com qualidade. De forma simplificada, representa-se na Figura 1, abaixo, um modelo básico de sistema de recuperação de informação.

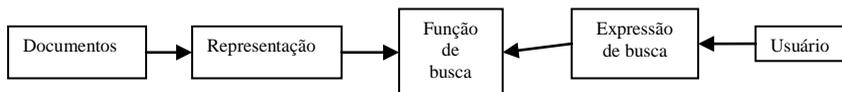


FIGURA 1 – Modelo simplificado de um sistema de recuperação de informação.
Fonte: Baseado em Ferneda (2003).

Parafraseando autores como Sales e Vieira (2007), Robredo (2003) e Saracevic (1999), os quais apresentam e discutem a recuperação da informação como essencial para a evolução das ciências, reafirmando sua importância na pesquisa e na comunicação científica em todas as áreas de conhecimento, complementa-se que os modelos mais utilizados são: Booleano, Vetorial, Probabilístico, *Clustering* e o *Feedback*, apresentados abaixo. Estes autores destacam, ainda, que um dos principais problemas ou desafios dos Sistemas de Recuperação da Informação (SRI) é recuperar somente documentos importantes para o usuário, ou seja, que tenham “relevância”, e infelizmente esses SRIs possuem uma relevância parcial. Cabe discutir, também, que cada SRI é desenvolvido e se especializa num tipo de recuperação, ou seja, adapta-se a um modelo específico.

a) Modelo Booleano: baseia-se na teoria e álgebra de Boole ou Booleana. A recuperação ocorre por meio das expressões de busca: *and* (e), *or* (ou) e *not* (não) e os documentos recuperados deveriam satisfazer as necessidades dos usuários. Apresenta como vantagens o baixo custo, armazenamento por índice e alta velocidade de resposta.

b) Modelo Vetorial: nesse modelo o documento é representado como um vetor de termos, sendo que cada termo (palavra) recebe um valor associado que indica o grau de importância (relevância) deste no documento ou na coleção toda. Para calcular a relevância, o sistema contabiliza o número de vezes que a palavra aparece em cada documento e o número de vezes que aparece no restante do documento ou da coleção. Dessa maneira, se a frequência é alta, significa que a palavra é relevante. Este é um dos modelos mais utilizados em SRI, embora a relevância seja em relação ao sistema e não ao usuário.

c) Modelo Probabilístico: nesse modelo a ordenação dos documentos é calculada pesando dinamicamente os termos da consulta relativa aos documentos e baseia-se no princípio da ordenação probabilística. O cálculo da probabilidade ocorre por meio dos

documentos que são relevantes para a consulta e os que não são. A relevância é calculada com base nas consultas dos usuários, ou seja, por retroalimentação.

d) Modelo *Clustering*: funciona pelo agrupamento de documentos similares entre si. Considera-se como um processo de organizar “documentos” em grupos cujos membros são, de algum modo, semelhantes ou similares em alguns aspectos. Como resultado, os documentos são classificados de acordo com um agrupamento de algoritmos, pelo sistema, o que não significa que a classificação seja por assuntos ou por áreas do conhecimento.

e) Modelo *Feedback* (retroalimentação): é uma informação recebida por um SRI que o informa sobre seu próprio comportamento. Sua principal característica é essa interação do usuário com o sistema, por meio da formulação e reformulação das buscas. Um sistema de *feedback* pode ser definido por meio de entradas e saídas, ou seja, consome entradas e produz saídas, e o *feedback* é uma saída que se transforma em entrada, quando inserido pelo usuário.

É constante a evolução das tecnologias de informação e comunicação e, em especial, a evolução das ciências, como a inteligência artificial, ciências da computação e da informação, conforme destacam Martínez Usero e Beltrán Orenes (2005), em seu artigo, explorando a evolução dos mecanismos de recuperação da informação. Também Souza (2006), Bundy (1997) e Mendel (1995) discutem e apresentam novos recursos e modelos, alguns revistos, atualizados e aperfeiçoados como soluções para uma recuperação mais eficiente dos mecanismos de busca na atual miscelânea de informações em que se transformou a *web*.

Cabe destacar que os mecanismos de busca que utilizam a lógica tradicional ou clássica para a recuperação de informações assumiam valores lógicos de 0 ou 1 (sim/não ou positivo/negativo) que são considerados valores matemáticos. Nas versões atuais, utilizando abordagens mais recentes, os mecanismos de busca incorporam a lógica difusa, que vem permitir uma abordagem mais ampla e abrangente, permitindo assumir valores entre 0 e 1. Assim, o universo abordado é mais amplo e passa a utilizar valores linguísticos – o uso de significados, de semântica. Nesta concepção, vários outros aspectos devem ser considerados no estudo e incorporação de mecanismos de busca: o elemento passa a ter um grau de pertinência variando entre 0 e 1. O uso de função de pertinência (calcula o grau de pertinência) e o próprio grau de pertinência calculam se o termo indexado é relevante ou não e em que escala se apresenta dentro do sistema de recuperação de informação. Destaca-se que essas evoluções são rápidas – a cada dia se implementa algum recurso mais eficaz. As revisões na área da

tecnologia devem ser delimitantes em um determinado período, pois a cada dia têm-se novas descobertas.

Muitos dos recursos utilizados em mecanismos de busca 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, definidos sucintamente logo a seguir.

a) Modelo Difuso: essa lógica permite intervalos ou valores relativos entre 0 e 1. Possibilita quantificar valores de forma linguística, como pequeno, médio e grande e é muito utilizado, devido à simplicidade. Em aplicação de busca, calcula a similaridade entre documento e consulta, medida pela função de pertinência (construída a partir do conceito de correlação entre termos). Assim, a lógica difusa é caracterizada como mapeamentos não-lineares de um vetor de entrada (dados) em uma escala de saída, formando uma coleção de sistemas independentes de múltiplas entradas e uma única saída (Mendel, 1995).

b) Modelo Fuzzi: similar ao modelo difuso, nesse modelo os documentos (banco de dados) e consultas (perguntas/*queries*) são representados pelo uso de palavras-chave, as quais são relacionadas aos conteúdos semânticos reais. Segundo Baeza-Yates e Ribeiro Neto (1999), a união dos documentos com os termos da consulta é aproximada (até 1) ou parcial (até 0). Isso pode ser aceito, considerando que cada termo da *query* define um conjunto Fuzzy e cada documento possui um grau de similaridade (usualmente menor que 1) neste conjunto. Esta interpretação de processo de recuperação é a fundamentação básica dos vários modelos de conjuntos Fuzzy para recuperação de informação, largamente utilizado nos últimos anos. A aplicação da lógica difusa ou Fuzzy, em que são tratados também os valores entre 0 e 1, permite melhor representação do universo linguístico humano. Entre o “falso” (0) ou “verdadeiro” (1) é permitido atribuir conceitos como “alto”, “médio” ou “baixo” que podem representar pertinência ou relevância (GIMÉNEZ LUGO; ANDRADE; SICHMAN, 2000).

c) Modelo Booleano Estendido: ao contrário do tradicional booleano, esse modelo permite a busca com a intersecção entre duas expressões ou termos. Dessa maneira, o sistema que utiliza essa lógica desconsidera os documentos que não apresentam nenhum dos termos atribuídos na busca e igualmente os documentos que tenham pelo menos um termo da busca. Com base em Souza (2006), afirma-se que esse modelo atribui pesos aos termos e busca superar o problema das decisões binárias do modelo clássico, aproximando-se do modelo vetorial. O problema binário, citado pelo autor, refere-se à capacidade de o modelo clássico possuir apenas duas respostas: 0 ou 1, enquanto o modelo

booleano estendido é representado por valores relativos entre 0 e 1.

d) Modelo Especial Vetorial Generalizado: caracteriza-se por técnicas de indexação automática de termos índice ou palavras-chave. A importância do termo “índice” ocorre por meio da atribuição de pesos. O peso quantifica a importância do termo “índice” para descrever o conteúdo do documento. Nesse modelo, dois vetores de termos “índices” podem não ser ortogonais. Isso significa que os vetores de termos “índices” não são vistos como vetores ortogonais que compõem a base do espaço, da base. Para Souza (2006), esse modelo questiona a independência dos termos do “índice”, dos modelos clássicos, possibilitando relacionar certos termos com conceitos. Uma das formas de determinar relações entre termos é examinar a co-ocorrência destes no texto de cada documento, além de examinar as relações semânticas estabelecidas pelo uso de um tesauro.

e) Modelo Indexação Semântica Latente: nesse modelo a recuperação de documentos é realizada pelo relacionamento aos termos da consulta que não precisam estar necessariamente descritos nos índices. Assim, o sistema relaciona os termos com conceitos próximos retirados dos documentos, partindo do pressuposto de que os termos da consulta não descrevem significativamente o assunto de um documento. Mas esse modelo pode apresentar baixo desempenho em sistemas de recuperação de informação, devido a dois fatores: muitos documentos não relacionados podem ser incluídos no conjunto de respostas, e documentos relevantes que não estão indexados por nenhuma das palavras-chave da consulta não são recuperados. A principal razão desses dois efeitos é a inerente incerteza associada com o processo de recuperação, o qual é baseado em conjunto de palavras-chave.

f) Modelo de Redes Neurais: é baseado na estrutura dos neurônios do cérebro humano, que é representada de forma artificial e aplicada na recuperação da informação. Cada nó representa um neurônio e este por sua vez representa um termo ou um documento da coleção. Esse modelo caracteriza-se por sua capacidade de “aprender”, ou seja, pelo uso de algoritmos de aprendizado (parecidos com neurônios humanos). Dessa maneira, independente dos termos aplicados na busca, a rede neural infere, por meio de algoritmos, quais são os termos mais próximos aos submetidos na consulta e que fazem algum sentido para a pesquisa. Enfatiza-se que, quanto mais camadas a rede possuir, maior será a sua capacidade de aprendizado. As redes neurais estão crescendo muito nos últimos anos, levando-se em conta o aumento da qualidade e eficácia dos *hardwares* e *softwares*.

g) Modelo Recuperação Textual Estruturada: nesse modelo, a recuperação da informação ocorre por meio da estrutura textual dos

documentos indexados, ou seja, as partes mais representativas de um documento, que são as figuras, tabelas e gráficos (ligados à lembrança visual do usuário), são utilizadas como ponto de acesso, facilitando as respostas relevantes aos usuários.

Além desses modelos, inúmeros outros recursos são constantemente disponibilizados na *web* e/ou apresentados em estudos e pesquisas, sempre em busca de um melhor desempenho dos mecanismos de busca, objetivando recuperação precisa e eficaz da informação. Resumidamente, citam-se três sistemas de busca *web*: 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; Motores de busca – conhecidos também como programas de busca, mecanismo de procura e ferramenta de busca, e são programas que têm três funções básicas – identificar páginas da *web*, indexar estas páginas em um banco de dados e um mecanismo de pesquisa como interface. A maioria dos resultados é classificada pela relevância; finalizando, os Metabuscaadores – 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, com os quais geralmente tem parceria.

A parte da ciência que estuda e desenvolve recursos para a recuperação da informação está em constante atualização e criação. Atualmente se discute a Linguística Computacional, que pode ser entendida, segundo Vieira e Lima (2001) como a área do conhecimento que explora as relações entre linguística e informática, tornando possível a construção de sistemas com capacidade de reconhecer e produzir informação apresentada em linguagem natural. Envolve várias áreas de pesquisa tradicionais, como: Linguística Teórica e Aplicada, Sintaxe, Semântica, Fonética e Fonologia, Pragmática, Análise do Discurso, entre outros. Isso tudo para tentar processar (leia-se "compreender e produzir") as línguas naturais e dominar o conhecimento linguístico humano e passá-lo para as máquinas – os agentes inteligentes – que possam localizar, filtrar e devolver as consultas de acordo com as necessidades de cada usuário.

Para que ocorra uma recuperação eficiente é necessário que as informações disponibilizadas na *web* recebam um tratamento adequado e padronizado. Destacam-se como recursos para o tratamento da informação: construção de tesouros (representar relacionamentos conceituais entre as palavras); uso de *clustering* (agrupamento de documentos relacionados); uso da semântica (compreensão de textos);

uso de criptografia (segurança); padronização da gramática e o uso de algarismos compostos por números, sinais e letras podem significar uma sigla, ou um período datado. Para Godoy Viera (2008), o estudo de mais atualizado para a preparação e recuperação da informação pode ser resumido em Análise Léxica, Eliminação de *stopwords*, *Stemming* das palavras restantes, Seleção de termos de indexação, Construção de estruturas de categorização de termos e a Criação de arquivo invertido.

Cabe destacar que o estudo da área da recuperação da informação é imenso, complexo e desafiador, necessitando de dedicação e revisões específicas. Nesta pesquisa foram apenas revisados alguns conceitos básicos, essenciais ao levantamento pretendido, como a questão do uso da recuperação inteligente da informação e ontologia.

A Web é hoje a maior fonte de informação eletrônica disponível e de livre acesso. Entretanto, por causa da sua natureza dinâmica, a tarefa de encontrar informações relevantes se torna cada dia mais frustrante, desgastante e ineficaz. Muitos esforços de pesquisa têm sido feitos no sentido de sanar esse problema e entre resultados têm-se os agentes inteligentes, a mineração de dados na *web* e as ontologias, interesse desta pesquisa, que se apresenta a seguir.

3 – ONTOLOGIA

Parafraseando Gruber (1993), uma ontologia na *web* é a representação de uma conceituação, um conjunto de conceitos estudados e especificados sobre uma determinada área de domínio. Guarino (1998) esclarece o papel de uma ontologia na *web*, o 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.

Segundo Morshed e Singh (2005), a ontologia pode ser usada 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 da *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.

Uma das características da *Web Semântica* está no compartilhamento, significado e reuso de termos que identificam, padronizam e automatizam as informações e/ou um conjunto de

informação da *web*. São os processadores de máquinas – os agentes inteligentes – que localizam os metadados e os relacionam, proporcionando ao usuário final uma resposta complexa, com dados extraídos e direcionados, atendendo a demanda solicitada, de forma precisa e eficaz. Ou seja, o usuário deixa de analisar cada retorno de uma pesquisa e passa a incluir uma questão que descreve sua necessidade e os agentes inteligentes buscam pela melhor resposta, melhor opção para a pergunta do usuário.

Noy e McGuinness (2005) apresentam um manual para desenvolvimento de ontologias. Abordam o assunto repassando um passo-a-passo, facilitando o entendimento sobre o assunto. Destacam a importância das perguntas às quais a ontologia desenvolvida deverá responder. Essas autoras, que por sua vez se basearam em diversos outros autores, destacam algumas motivações para o desenvolvimento de ontologias, sendo o compartilhamento de um entendimento comum da estrutura da informação entre humanos e agentes inteligentes; permitir a reutilização de domínio do conhecimento; tornar explícitos os pressupostos de um domínio; separar o conhecimento, e analisar o conhecimento sobre o domínio. Portanto, é necessário um trabalho multidisciplinar de pesquisadores na construção de ontologias, pois é necessário o aprofundamento de questões semânticas, axiomas e sinônimos de uma área específica.

Ontologias descrevem categorias em um determinado domínio do conhecimento. Estas são criadas a partir do tratamento de cada conteúdo a ser inserido nesse segmento, obedecendo aos critérios linguísticos, semânticos e representativos. Conforme Kobashi (2007), ontologias dão “suporte ao raciocínio sobre as coisas de um domínio, a representação do conhecimento deve descrever o comportamento das coisas e as suas interações”.

A área da Ciência da Informação atua na organização, tratamento e dissemina a informação desde os tempos mais remotos da humanidade, portanto se caracteriza pelo acompanhamento histórico das evoluções no tratamento e recuperação da informação. Assim, as ontologias – vistas como soluções para a recuperação relevante na *Web Semântica* – são correlatas diretas dos tesouros, pois envolvem a teoria da classificação, indexação, tratamento linguístico/semântico de cada termo a ser incorporado e são desenvolvidas para cada área específica do conhecimento. Kobashi (2007) diz que “as ontologias pretendem ir além dos tesouros documentários. Essas abordagens ampliam as possibilidades de organizar e ter acesso à informação em sistemas”. Na atualidade, os mecanismos de recuperação de informação utilizam agentes inteligentes ou multiagentes, que interpretam (linguagem

artificial), localizam, codificam e armazenam informações, proporcionando resposta mais rápida e completa aos usuários. San Segundo e Béltran (2003), citados por Martínez Usero e Beltrán Orenes (2005), definem ontologias como “instrumento ou estrutura de tradução tecnológica que pode ser usado para a representação do conhecimento em tesouros, glossários, lista de cabeçalho de assunto, banco de dados e catálogos na *web* ou outros dispositivos de controle de vocabulário”.

Ramalho, Vidotti e Fujita (2007) contribuem de maneira marcante para a necessidade de trabalho multidisciplinar na área do tratamento e recuperação de informação na *web*, em especial na construção e utilização de ontologias, quando destacam:

Portanto, torna-se evidente o caráter interdisciplinar que delinea o *corpus* teórico do projeto *Web Semântica*, englobando essencialmente áreas como a Ciência da Informação e a Ciência da Computação, entre outras, apresentando-se como um campo fértil para pesquisas. Assim, esforços interdisciplinares são necessários para possibilitar o desenvolvimento de soluções multidisciplinares, respeitando as especificidades de cada área do conhecimento e tendo como objetivo comum auxiliar na evolução do conhecimento humano de forma integral.

Assim, com essa inovação proposta pela área da ciência da computação, que resgata teorias e conhecimentos antigos, ficam claras as parcerias entre ciências, para a busca da concretização destes recursos tecnológicos em prol da precisão, relevância e eficácia dos sistemas de recuperação de informação e a satisfação de usuários.

4 – ASPECTOS METODOLÓGICOS

Esta pesquisa caracteriza-se como um estudo bibliográfico e descritivo e foi desenvolvida no primeiro trimestre de 2008, como pesquisa final da disciplina Recuperação Inteligente da Informação (PCI3214), do Programa de Pós-Graduação em Ciência da Informação (PGCIN), da Universidade Federal de Santa Catarina (UFSC).

Objetivando identificar a existência ou não de aplicações de mecanismos de Recuperação Inteligente da Informação que utilizem ontologias como recurso na recuperação eficaz da informação, delimitou-se que esta pesquisa usaria o Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), com textos completos, na área de Ciências Sociais Aplicadas: Ciência da Informação e que tenham inserida em seus títulos a expressão “ciência da informação” e seu equivalente em inglês “*information science*”; portanto, os idiomas definidos foram português e

inglês. Delimitou-se o período de janeiro de 2007 a maio de 2008 como período a ser investigado. A pesquisa foi realizada nos dias 18 e 19 de maio de 2008 e conferida nos dias 23 e 24 do mesmo mês. Justifica-se esta escolha pelo fato de que a disciplina cursada é do PGCIN, ou seja, dentro da área da Biblioteconomia e Ciência da Informação.

A base de dados do Portal de Periódicos da CAPES apresenta em seu sítio o total de 12.365 títulos com textos completos. Seleccionando no Portal “textos completos”, área de “Ciências Sociais Aplicadas” e a subárea de “Ciência da Informação”, resultou em 355 títulos. Foram pesquisadas as expressões “ciência da informação” e “*information science*”, conforme demonstrado nas figuras 2 e 3, a seguir.

Resultados 0 - 7 de 7 na pesquisa por "ciencia da informação".

Título	Mostrar informações detalhadas
+ Ciência da Informação	
+ Ciência da Informação	
+ DataGramaZero : Revista de Ciência da Informação	
+ Encontros Bibli : Revista Eletrônica de Biblioteconomia e Ciência da Informação	
+ Perspectivas em Ciência da Informação	
+ Perspectivas em Ciência da Informação	
+ Revista Digital de Biblioteconomia e Ciência da Informação	

topo

FIGURA 2 – Lista de títulos recuperados com a expressão “ciência da informação”.
Fonte: Portal de Periódicos CAPES.

Resultados 0 - 19 de 19 na pesquisa por "information science".

Título	Mostrar informações detalhadas
+ African Journal of Library Archives and Information Science	
+ African Journal of Library Archives and Information Science	
+ Cartography and Geographic Information Science	
+ Information Science and Technology Abstracts : ISTA	
+ Information Sciences = Informacijos Mokslai	
+ Information Sciences. Applications	
+ Information Sciences. Informatics and Computer Science Intelligent Systems Applications	
+ Journal of Education for Library and Information Science	
+ Journal of Information Science	
+ Journal of Librarianship and Information Science	
+ Library and Information Science Abstracts : LISA	
+ Library and Information Science Research	
+ Library Literature and Information Science Full Text	
+ Library, Information Science and Technology Abstracts : LISTA	
+ Malaysian Journal of Library and Information Science	
+ Malaysian Journal of Library and Information Science	
+ Pakistan Journal of Library and Information Science	
+ Pakistan Library and Information Science Journal	
+ Surveying and Land Information Science	

topo

FIGURA 3 – Lista de títulos recuperados com a expressão “*information science*”.
Fonte: Portal de Periódicos CAPES.

Efetou-se o acesso ao Portal de Periódicos da CAPES por meio do acesso via UFSC, obtendo pleno acesso às bases de dados. Em cada um dos 26 títulos encontrados foi realizada a pesquisa com a expressão “Recuperação Inteligente da Informação (RII) e Ontologia” e seu equivalente em inglês, “*Intelligent Information Retrieval (IIR) and Ontology*”. Optou-se por aplicar a expressão em sua forma exata, utilizando os recursos booleanos e/ou alternando os termos, procurando recuperar informações relevantes. A seguir, apresentam-se os resultados obtidos.

5 – RESULTADOS E DISCUSSÃO

O Portal de Periódicos da CAPES é destinado a “professores, pesquisadores, alunos e funcionários de 191 instituições de ensino superior e de pesquisa em todo o País”. Oferece acesso aos textos completos de artigos de mais de 12.365 revistas internacionais, nacionais e estrangeiras e 126 bases de dados com resumos de documentos em todas as áreas do conhecimento. Inclui também uma seleção de importantes fontes de informação acadêmica com acesso gratuito na *web*. “O uso do Portal é livre e gratuito para os usuários das instituições participantes” e o acesso pode ser realizado a partir de qualquer terminal ligado à *web*, localizado nas instituições ou por elas autorizado. Encontra-se disponível no endereço <<http://www.periodicos.capes.gov.br/>> (CAPES, 2008).

Selecionando “Textos Completos – área Ciências Sociais e Aplicadas – subárea Ciência da Informação” do Portal, obteve-se uma lista de 355 títulos. Nesse universo, foram pesquisadas as expressões “ciência da informação” e “*information science*”. Foram recuperados sete títulos para a expressão em português e 19 títulos em inglês, perfazendo um universo de 26 títulos.

Dos sete títulos recuperados com a expressão “ciência da informação”, dois títulos eram duplos: revista *Ciência da Informação* (Scielo e *WilsonWeb Journal Directory*) e a revista *Perspectivas em Ciência da Informação* (*site* direto e Scielo). Quanto aos títulos recuperados com a expressão em inglês, constatou-se também a duplicidade de dois títulos: *African Journal of Library Archives and Information Science* e *Malaysian Journal of Library and Information Science*, ambos disponibilizados nas bases de dados EBSCO e *WilsonWeb Journal Directory*. Como as bases de dados são diferentes e o objetivo é verificar os mecanismos de busca e, conseqüentemente, se a recuperação é eficaz, manteve-se o universo de 26 títulos a serem pesquisados. A seguir apresenta-se o Quadro 1, demonstrando os resultados obtidos no portal de periódicos, perfazendo o universo da pesquisa.

Quadro 1 – Resultados da coleta de dados no Portal de Periódico – Universo da pesquisa.

TÍTULO	ISSN	BASE	RESULTADO
NACIONAIS = ciência da informação			
Ciência da Informação	0100-1965 (print)	SCIELO	Só 'ontologia' 2004
Ciência da Informação	0100-1965	WilsonWeb Journal Directory	Nenhum
DataGramaZero: Revista de Ciência da Informação	1517-3801		Só 'ontologia' = 3 artigos (sem mecanismo de busca - usado o 'localizar' nesta página)
Encontros Bibli : Revista Eletrônica de Biblioteconomia e Ciência da Informação	1518-2924		Expressão de busca completa = 5 artigos - 2 de 2007
Perspectivas em Ciência da Informação	não encontrado		Expressão de busca completa = 4 artigos - só 1 de 2007
Perspectivas em Ciência da Informação	1413-9936 (print)	SCIELO	Só 'ontologia' = 1 / 'RI' = 1
Revista Digital de Biblioteconomia e Ciência da Informação	1678-765X		Só 'ontologia' = 1 / 'RI' = 1
INTERNACIONAIS = information science			
African Journal of Library Archives and Information Science	0795-4778	EBSCO Research Databases	Nenhum. Remete pesquisa p/ portal africano de periódicos.
African Journal of Library Archives and Information Science	0795-4778	WilsonWeb Journal Directory	Nenhum. Remete pesquisa p/ portal africano de periódicos.
Cartography and Geographic Information Science	1523-0406	GALE Cengage Learning	Nenhum. Várias tentativas p/ conseguir acesso.
Information Science and Technology Abstracts : ISTA	Diretório/base	EBSCO Research Databases	Retirado termo 'intelligent' = 16 artigos
Information Sciences = Informacijos Mokslai	1392-1487	EBSCO Research Databases	Nenhum
Information Sciences. Applications	Falta	ScienceDirect	Expressão de busca completa = 1 artigo
Information Sciences. Informatics and Computer Science Intelligent Systems Applications	0020-0255	Elsevier	Retirado termo 'intelligent' = 4 artigos

TÍTULO	ISSN	BASE	RESULTADO
Journal of Education for Library and Information Science	0748-5786	WilsonWeb Journal Directory	Nenhum
Journal of Information Science	e1741-6485 / 0165-5515	SAGE Journals	Retirado termo 'intelligent' = 4 artigos
Journal of Librarianship and Information Science		SAGE Journals	Nenhum
Library and Information Science Abstracts: LISA	e1741-6477 / 0961-0006	CSA / ILLUMINA	Retirado termo 'intelligent' = 2 artigos
Library and Information Science Research	0740-8188	Science Direct	Retirado termo 'intelligent' = 4 artigos
Library Literature and Information Science Full Text	Diretório/base	WilsonWeb Journal Directory	Retirado termo 'intelligent' = 4 artigos
Library, Information Science and Technology Abstracts : LISTA	Diretório/base	EBSCO Research Databases	Nenhum. Contém o título Information Science and Technology Abstracts: ISTA
Malaysian Journal of Library and Information Science	1394-6234	EBSCO Research Databases	Nenhum
Malaysian Journal of Library and Information Science	1394-6234	WilsonWeb Journal Directory	Nenhum
Pakistan Journal of Library and Information Science	1680-4465	EBSCO Research Databases	Nenhum. Periódico inteiro de 2007 em pdf.
Pakistan Library and Information Science Journal	1994-3059	EBSCO Research Databases	Nenhum
Surveying and Land Information Science	1538-1242	WilsonWeb Journal Directory	Nenhum

Fonte: a autora.

Ao efetuar a pesquisa no *Library, Information Science and Technology Abstracts (LISTA)* constatou-se que é um diretório contendo vários periódicos na área da Ciência da Informação. Dentre eles se encontra o *Information Science and Technology Abstracts (ISTA)*, já pesquisado anteriormente, configurando, assim, duplicidade de informações. Mas, ao efetuar-se a pesquisa no diretório *LISTA*, não foi recuperado nenhum artigo. Isso indica que a recuperação de informações efetuadas no diretório *LISTA* não foi eficiente nesta pesquisa, já que consultando o *ISTA* foram encontrados artigos, apesar de ter efetuado alterações na expressão de busca.

Analisando-se o Quadro 1 acima, foram encontrados 43 artigos

em 26 títulos pesquisados no Portal da CAPES. Desses 26 títulos, 14 títulos não apresentaram nenhum artigo com a expressão de busca. Nove títulos do universo pesquisado trouxeram resultados após a alteração da expressão de busca (retirado o termo 'inteligente/*intelligent*', recuperaram-se seis títulos; pesquisando somente o termo 'ontologia/*ontology*' foram recuperados mais três títulos). Apenas três títulos apresentaram resultado afirmativo para a expressão de busca "Recuperação Inteligente da Informação e Ontologia" e seu equivalente em inglês "*Intelligent Information Retrieval and Ontology*".

No Quadro 2 apresenta-se a totalização de dados recuperados nesta pesquisa.

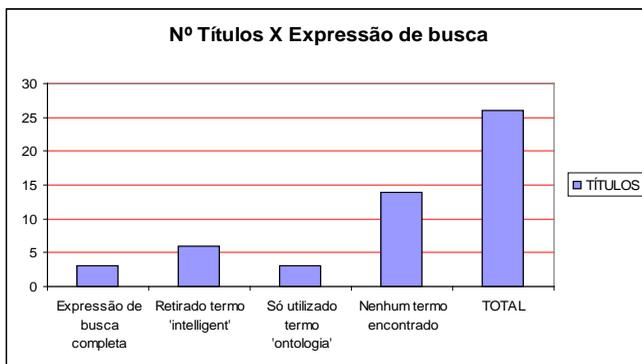
Quadro 2 – Resultados da pesquisa no Portal de Periódicos da CAPES

PESQUISA PORTAL CAPES	TÍTULOS	ARTIGOS
Expressão de busca completa	3	4
Retirado termo ' <i>intelligent</i> '	6	34
Só utilizado termo 'ontologia'	3	5
Nenhum termo encontrado	14	0
TOTAL	26	43

Fonte: a autora.

Apenas três títulos da lista selecionada no Portal da CAPES atenderam a expressão de busca "recuperação inteligente da informação e ontologia/*intelligent information retrieval and ontology*", sendo que destes, dois estão em português e um em inglês, como se observa no Gráfico 1 a seguir. É surpreendente este resultado, visto que a maioria de publicações está em inglês, principalmente quando trata de novos assuntos criados e/ou em desenvolvimento, como a questão da recuperação inteligente da informação, pois nestes últimos anos vem tomando grande impulso.

GRÁFICO 1 – Relação do número de títulos recuperados e a expressão de busca utilizada.

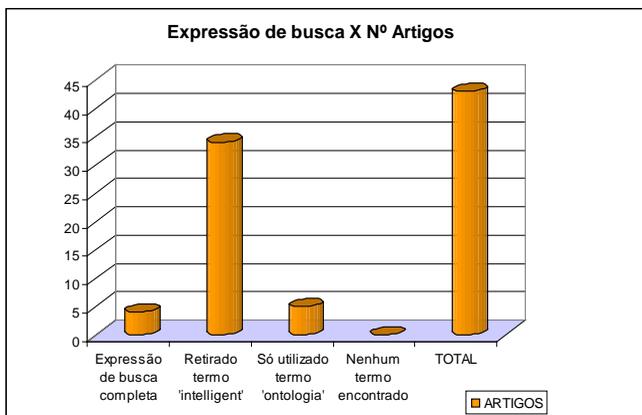


Fonte: a autora.

Outro aspecto a ser considerado é que a área pesquisada foi a Ciência da Informação. Acredita-se que tal pesquisa, se aplicada nas áreas de Ciência da Computação e Engenharias, os resultados poderão ser bem diferenciados.

No Gráfico 2 demonstra-se o número de artigos recuperados com a pesquisa realizada utilizando a expressão de busca anteriormente definida.

GRÁFICO 2 – Relação da expressão de busca 'RII e ontologia/IIR and ontology' e o número de artigos recuperados.



Fonte: a autora.

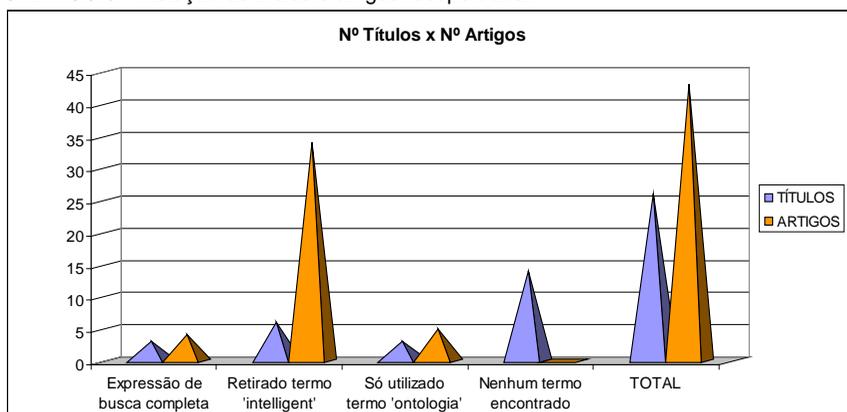
Considerando que as pesquisas voltam-se quase totalmente à questão da recuperação precisa e relevante na *web*, e que têm sido feitos esforços em prol do funcionamento da *Web Semântica*, constata-se que apenas quatro artigos atenderam plenamente a expressão de busca adotada, o que demonstra que na área da Ciência da Informação as publicações sobre o tema “RII e ontologia/IIR and ontology” são incipientes.

Segundo Souza (2006, p. 171), inúmeras pesquisas têm se voltado ao uso de estruturas da linguagem natural, como os sintagmas verbais e nominais, para indexação e recuperação de informações, e de “ferramentas de representação de relacionamentos semânticos e conceituais, como os tesouros, para ampliar a gama de informações recuperadas e aferição de contextos, além de outras estratégias derivadas da linguística e da ciência da informação”. Complementa-se aí a ênfase dada ao desenvolvimento de ontologias.

Ainda para Souza (2006, p. 172), “outras metodologias similares implantadas em SRIs permitem a busca de expressões regulares, ou mesmo analisam a proximidade da ocorrência de alguns termos, expandindo o conceito de palavra-chave para frases ou outras hierarquias lexicais”.

No Gráfico 3 a seguir, apresenta-se a relação entre os títulos definidos como universo da pesquisa e o número de artigos recuperados.

GRÁFICO 3 – Relação de títulos e artigos recuperados.



Fonte: a autora.

Observa-se que a expressão de busca adotada obteve quatro artigos em três títulos da amostra pesquisada, o que indica que se

encontra em pleno desenvolvimento a junção desses dois assuntos: recuperação inteligente da informação e ontologia, da mesma forma que toda a interação com a *Web Semântica*.

Da mesma maneira, chama a atenção o fato de 14 títulos apresentarem recuperação igual a zero da amostragem pesquisada. Isso condiz com a necessidade de que os mecanismos de busca sejam adequados a cada base de dados, diretório ou *sites*. Na visão de Godoy Vieira (2008), os mecanismos de buscas possuem regras e procedimentos (algoritmos) que realizam diferentes representações lógicas e tipos de buscas, onde refinam/filtram a informação que o usuário deseja. Mas, é necessário que o usuário saiba utilizar o sistema de recuperação, ou seja, que conheça os recursos deste e o que deseja procurar. Assim, os sistemas de recuperação da informação podem utilizar campos determinados e/ou todos os campos, ou recursos para o acesso à base, a *Web*, sendo que o tamanho dos índices e a capacidade dos equipamentos têm que ser levados em consideração, tanto pelos projetistas que o constroem, como pelos usuários que o utilizam.

Ferneda (2003, p. 110) analisa a importância da interação entre as áreas da Ciência da Computação e a Ciência da Informação, ambas como alicerce na construção de estruturas de recuperação dotadas de semântica, pois “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”. Isso pode ser observado ao comparar-se o número de artigos recuperados (34) em apenas seis títulos da amostra pesquisada, porque o termo “inteligente/intelligent” foi omitido.

Conforme Hamdi (2005), a Recuperação Inteligente de Informação (RII) prevê estruturação e acessibilidade para a informação, permitindo sua recuperação precisa e relevante, pois atua num contexto semântico. A extração, personalização e agentes são vistos como elementos fundamentais de suporte às tarefas cotidianas e possibilitam um melhor desempenho nos resultados obtidos. Já a agilidade e a precisão são os maiores objetivos de todas as áreas do conhecimento – em especial as áreas da Ciência da Computação e a Ciência da Informação – enfim, incrementar os mecanismos de busca com “capacidade decisória” tem sido a maior busca de inúmeros estudiosos, pesquisadores e cientistas (FERNEDA, 2003).

5.1 – Exemplos de RII e Ontologia

Este trabalho surgiu a partir da discussão se existem ou não aplicações de mecanismo de Recuperação Inteligente da Informação que use Ontologia como recurso na recuperação precisa e eficaz da informação. Conforme citado na discussão, apenas quatro artigos científicos foram recuperados neste levantamento, os quais apresentam-se abaixo, acompanhados do resumo e das palavras-chave.

1) SCHIESSL, José M. Ontologia: o termo e a idéia. *Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.*, Florianópolis, n. 24, p. 172-181, 2º sem. 2007.

Aborda o conceito de Ontologia e relaciona alguns problemas que motivam os pesquisadores desta área. Contextualiza a sobrecarga informacional em virtude do crescente volume de texto eletrônico armazenado em bases de dados, em especial na Web, e sua implicação na manutenção do fluxo da informação. Além disso, introduz a idéia da Web Semântica como uma solução para a organização dos conteúdos das páginas nos sítios da Web e o papel preponderante das ontologias, dentro deste contexto, que auxiliam a automatização da interpretação destes conteúdos e otimiza a recuperação da informação.

2) DIAS, Guilherme A.; HENN, Gustavo; SILVA, José W. de Moraes. Tecnologia da informação e serviços de referência eletrônicos: uma proposta de aplicação baseada em *chatbots* e ontologias. *Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.*, Florianópolis, n.23, p. 47-61, 1º sem. 2007.

Discute os principais conceitos relacionados ao desenvolvimento de um sistema de referência eletrônico em unidades de informação através do uso da tecnologia de informação. Introduz o conceito de *chatbots*, apresenta seus principais usos, gerações evolutivas e linguagens utilizadas na sua implementação. Conceitua ontologia de maneira geral e contextualiza seu significado na área da Ciência da Informação. Descreve a arquitetura do sistema DELPHOS, explicitando as etapas da construção da interface do usuário, processador AIML e bases de conhecimentos. Conclui sugerindo estudos futuros com o objetivo de avaliar as dinâmicas emergentes da interação dos usuários com o *chatbot* e a possibilidade de construção de um perfil de usuário a partir das interações com o mesmo.

3) VALENTIM, Marta L. P. Web Semântica: ontologias como ferramentas de representação do conhecimento. *Perspec. Ci. Infor.*, v. 12, n. 1, p. 65-83, 2007.

A Web Semântica é uma extensão da Web que acrescenta

semântica ao atual formato de representação de dados. Para isso foram propostas diversas tecnologias, dentre estas a criação de ontologias, visando atribuir sentido e significado ao conteúdo dos documentos, atuando como ferramenta de representação do conhecimento. Na presente pesquisa, procurou-se verificar se as ontologias seriam utilizadas apenas para representar o assunto de uma página Web através de termos contextualizados ou se tentariam controlar o vocabulário da Web como um Tesouro.

4) KIM, Kyoung-Y et al. Mereotopological assembly joint information representation for collaborative product design. *Robot Comput Integr Manuf* (2008). doi:10.1016/j.rcim.2008.03.010.

Discute-se um método de representação baseado em ontologia para diferenciar articulações (junções) de montagem no projeto de produto inteligente e colaborativo. No que o projeto se torna mais intenso de conhecimento, inteligente e colaborativo, há necessidade de tornar-se mais crítico para os quadros computacionais que permitam o desenvolvimento dos produtos através do efetivo apoio à representação formal, captura, recuperação e reutilização de produtos de conhecimento. Articulação é um elemento fundamental da montagem de modelos que são muitas vezes ambíguos quando ocorre a partilha de modelos. Embora várias articulações podem ter semelhanças, geometrias e topologias, as implicações físicas dos processos selecionados podem variar significativamente. É possível atribuir notas e anotações para entidades geométricas a fim de distinguir articulações. No entanto, essas informações textuais não preparam prontamente o modelo para atividades de baixo fluxo, tais como análise e simulação. De forma ilustrativa, os analistas devem ler e interpretar as anotações, a fim de desenvolver as condições de limites adequadas. Neste trabalho, apresentamos uma ontologia de projeto de montagem que explicitamente representa limitações de montagem, incluindo limitações de articulações, e deduz qualquer remanescente implícito. Ao relacionar conceitos através da tecnologia de ontologias em vez de apenas definir dados de sintaxe, conceitos de montagem e articulação, todos podem ser capturados na sua totalidade ou prorrogados conforme necessário. Ao usar o conhecimento adquirido pela ontologia, articulações e similares podem ser diferenciadas. Para esta pesquisa, foi utilizado a mereotopologia¹ (*mereotopology*), que é uma teoria baseada em regiões para partes, e *Semantic Web Rule Language - SWRL* para representar a

¹ Segundo Campos (2004, p.29) “[vem] da ontologia formal, de teoria todo-parte, também denominada ‘mereotopologia’”.

diferença das articulações, bem como definir os termos de articulação de projeto e seus relacionamentos. Também utilizamos *SWRL* de modo que regras de articulação possam ser fundamentadas para diferenciar articulações de montagem. Por último, através do uso uma ontologia, várias articulações geométrica e topologicamente parecidas são diferenciadas com sucesso de maneira padrão e que possa ser interpretada por máquinas.

6 – CONSIDERAÇÕES FINAIS

Este estudo objetivou identificar, na literatura, a existência ou não de aplicações de mecanismos de “Recuperação Inteligente da Informação” que usem “Ontologia” como recurso na recuperação precisa e eficaz da informação. Foram localizados os quatro artigos referenciados no item anterior, demonstrando que as pesquisas na área da recuperação da informação buscam por melhores recursos que visem uma recuperação precisa e eficaz da informação, neste mundo de caos informacional, em virtude da *web* e sua abertura desenfreada e sem controle.

Partindo do exposto, percebe-se que as áreas de Ciência da Informação e da Computação, principalmente a Inteligência Artificial e a Engenharia do Conhecimento têm muito trabalho a ser desenvolvido, pois tratar e recuperar informações tem característica multidisciplinar e requer um trabalho conjunto envolvendo estas áreas e outras afins.

Diante de problemas encontrados ao efetuar a pesquisa, obedecendo à metodologia proposta (item 4) como a repetição do mesmo periódico científico, dentro do mesmo portal, mas em bases de dados diferentes, significando duplicidade de títulos, ao utilizar os recursos disponíveis para a recuperação de informações, os resultados foram diferentes e, no caso da Ciência da Informação, na segunda base de dados consultada, foi negativo.

Observa-se outro caso, o do *Information Science and Technology Abstracts (ISTA)*, que recuperou o maior número de artigos (Quadro 1), embora com alteração na expressão de busca, mas que aparece em outro diretório – *Library, Information Science and Technology Abstracts (LISTA)*, cujo resultado foi negativo para a pesquisa. Isso indica que a recuperação de informações efetuada pelo diretório *LISTA* não foi eficiente nesta pesquisa, já que a consulta ao *ISTA* recuperou artigos.

Casos como estes indicam a necessidade de programar mecanismos de buscas mais eficientes em portais, diretórios e base de dados e/ou moderniz-los, utilizando os recursos da recuperação inteligente da informação – o uso de agentes inteligentes, multiagentes,

em parceria com tesouros e mais inovador, as ontologias.

Nesta pesquisa, outras informações foram encontradas, como o trabalho de Lévy (2008) com a arquitetura para uma rede de linguagem semântica (*Architecture of a semantic networking language*). Ou ainda, Ying, Tianjiang e Xueling (2007), que desenvolveram um sistema de recuperação inteligente de informação baseado em ontologia (*Building intelligent information retrieval system based on ontology*), pois consideram que uma ferramenta de recuperação de informação é uma maneira importante para que usuários obtenham o conhecimento e a informação, e ainda mais, de forma precisa e eficiente.

Os resultados apresentados neste artigo instigam novas pesquisas, principalmente o trabalho interdisciplinar entre a Ciência da Informação e a Computação e suas áreas afins.

REFERÊNCIAS

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. *Modern Information Retrieval*. New York: Addison Wesley, 1999.

BARRETO, Aldo de Albuquerque. *Memórias, esquecimento e estoques de informação*. Página pessoal, jan. 2008. Disponível em: <<http://aldoibct.bighost.com.br/MemorEsquecim.pdf>>. Acesso em 10 maio 2007.

BUNDY, A. *Artificial Intelligence Techniques*. Springer-Verlag, 1997.

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Disponível em: <<http://www.capes.gov.br/>>. Acesso em: 15 maio 2008.

_____. Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Disponível em: <<http://www.periodicos.capes.gov.br/portugues/index.jsp>>. Acesso em: 15 maio 2008.

CAMPOS, Maria L. A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. *Ci. Inf.*, Brasília, v. 33, n. 1, p. 22-32, jan./abr. 2004.

FERNEDA, Edberto. *Recuperação da informação: análise da contribuição da ciência da computação para a ciência da informação*. São Paulo, 2003. 147 p. Tese (Doutorado em Ciência da Informação e Documentação) – Universidade de São Paulo.

GIMÉNEZ LUGO, Gustavo A.; ANDRADE, Marco T. C. de; SICHMAN, Jaime S. Recuperação de informação usando computação nebulosa a partir de documentos com estruturas heterogêneas. In: WORKCOMP 2000, São José dos Campos. *Anais eletrônicos...*, São Paulo: ITA, 2000. Disponível em: <http://www.lti.pcs.usp.br/publicacoes/publicacoes2000.html>. Acesso em: 15 mar. 2008.

GODOY VIERA, Angel F. *Disciplina de recuperação inteligente da informação (RII)*. Programa de Pós-Graduação em Ciência da Informação, da Universidade Federal de Santa Catarina. 1. trim. 2008.

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

GUARINO, N. *Formal ontology and information systems*. 1998. Disponível em:

<<http://www.loan-cnr.it/Papers/FOIS98.pdf>>. Acesso em: 17 nov. 2007.

HAMDI, Mohamed Salah. Extracting and Customizing Information Using Multi-Agents. In: SCIME, Anthony (ed.). *Web mining: applications and techniques*. Hershey: Idea Group Publishing, 2005. p. 228-252.

KOBASHI, Nair Yumiko. Fundamentos semânticos e pragmáticos da construção de instrumentos de representação de informação. *DataGramaZero - Revista de Ciência da Informação*, v. 8, n. 6, 2007.

LÉVY, Pierre. *Architecture of a semantic networking language*. CRC, FRSC, University of Ottawa, v. 2, 2008. Disponível em: <<http://www.ieml.org/IMG/pdf/1-SNL-2.pdf>>. Acesso em 12 maio 2008.

MARTÍNEZ USERO, José Angel; BELTRÁN ORENES, Maria Pilar. Ontologies in the Context of Knowledge Organization and Interoperability in e-Government Services. In: *IRFD World Forum 2005: Conference on Digital Divide, Global Development and the Information Society*, Tunis, Tunisia, 2005. Disponível em: <<http://www.irfd.org/events/wf2005>>. Acesso em 11 maio 2008.

MENDEL, J. M. Fuzzy logic systems for engineering: a tutorial. *Proceedings of the IEEE*. v. 83, n. 3, p. 345-377, 1995.

MORSHED, Ahsan-ul; SINGH, Ramanjit. Evaluation and ranking of ontology construction tools. *Technical Report* DIT-05-013. March 2005. Disponível em: <<http://eprints.biblio.unitn.it/archive/00000747/01/013.pdf>>. Acesso em 15 ago. 2007.

NEPOMUCENO, Carlos. As plataformas do conhecimento. *DataGramaZero - Revista de Ciência da Informação*, v. 8, n. 5, 2007.

NOY, Natalya F.; McGUINNESS, Deborah L. *Desarrollo de Ontologías-101: guía para crear tu primera ontología*. Stanford University, 2005. Disponível em: <http://protege.stanford.edu/publications/ontology_development/ontology101.pdf>. Acesso em 03 ago. 2007.

PAES CARDOSO, Olinda N. Recuperação da Informação. *INFOCOM: Journal of Computer Science*, v. 2, n. 1. 2004. Disponível em: <<http://www.dcc.ufla.br/infocomp/artigos/v2.1/olinda.pdf>>. Acesso em: 09 mar. 2008.

PEREIRA, Vasco N. S. S. (2004). *Arquitetura de um motor de busca: exemplo do Google*. Seminário 1, Mestrado em Informática e Sistemas. Departamento de Engenharia Informática, Universidade de Coimbra 2003/2004. Disponível em: <http://eden.dei.uc.pt/~vasco/portfolio/Google_v1.pdf>. Acesso em: 17 mar 2008.

RAMALHO, Rogério A. Sá; VIDOTTI, Silvana A. B. G.; FUJITA, Mariângela S. L.. Web semântica: uma investigação sob o olhar da Ciência da Informação. *DataGramaZero - Rev. Ci. Inf.*, v.8, n.6, 2007.

RANGANATHAN, S. R. *The colon classification*. New Brunswick: The Rutgers State University, 1965. p. 9 - 41 (Rutgers series on systems for the intellectual organization of information, 4)

ROBREDO, J. (2003). *Da ciência da informação revisitada aos sistemas humanos de informação*. Universidade de Brasília, DF: Ed. Thesaurus. 2003. 245 p.

SALES, R. de; VEIRA, A. F. G. (2007). 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: 09 mar. 2008.

SARACEVIC, T. (1999). *Information Science*. Disponível em: <<http://www.scils.rutgers.edu/~tefko/JASIS1999.pdf>>. Acesso em: 09 mar. 2008. [gleysj]

SOUTO, Patricia Nascimento. E-publishing development and changes in the scholarly communication system. *Ci. Inf.*, Brasília, v. 36, n. 1, p. 158-166, jan./abr. 2007.

SOUZA, Renato Rocha. Sistemas de recuperação de informações e mecanismos de busca na *web*: panorama atual e tendências. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 11, n. 2, p. 161-173, maio/ago. 2006. Disponível em: <<http://www.scielo.br/pdf/pci/v11n2/v11n2a02.pdf>>. Acesso em: 20 mar. 2008.

VIEIRA, R.; LIMA, V. L. S. *Linguística computacional: princípios e aplicações*. In: *Escola de Informática da SBC-Su*, 9. Luciana Nedel (Ed.) Passo Fundo, Maringá, São José. SBC-Sul, 2001.

YING, Pan; TIANJIANG, Wang; XUELING, Jiang. Building intelligent information retrieval system based on ontology. In: The Eighth International Conference on Electronic Measurement and Instruments ICEMI'2007. *IEEE*, n. 1, 2007.

