UM LEVANTAMENTO BIBLIOGRÁFICO SOBRE A UTILIZAÇÃO DE ONTOLOGIAS EM INTERFACES DE BUSCA

A BIBLIOGRAPHICAL SURVEY ON THE USE OF ONTOLOGIES IN SEARCH INTERFACES

Marcia Cristina dos Reis[1]
Edberto Ferneda[2]

RESUMO
Os buscadores disponíveis atualmente na Web, a maioria baseados nas caixas de texto, tiveram mudanças pouco significativas nos últimos anos em relação à forma de interação com o usuário. Apesar de terem se tornado populares em função da sua simplicidade, ainda apresentam inúmeros problemas relacionados principalmente à dependência de conhecimento do pesquisador em relação ao assunto investigado, às dificuldades em expressar uma necessidade de informação por meio de palavras-chave, à problemas de ambiguidade da linguagem e à composição inadequada de termos de busca. Nesse contexto, esta pesquisa tem como objetivo analisar a literatura sobre interfaces de recuperação de informação e levantar conceitos e características que possam subsidiar a proposta de um modelo de interface baseado em ontologias, com recursos capazes de auxiliar o usuário durante a busca e minimizar os problemas observados. Para alcançar os objetivos propostos, a metodologia adotada neste estudo contou com pesquisa de caráter exploratório e bibliográfico, baseada em autores relevantes da área da Ciência da Informação. Como resultados, pode-se perceber que as ontologias têm sido amplamente discutidas e adotadas em diversos experimentos relacionados com o desenvolvimento de interfaces de busca mais amigáveis, proporcionando melhorias significativas no processo de recuperação de informação. Tal recurso pode possibilitar ao usuário uma interação direta com os conceitos do domínio pesquisado e por meio deles, permitindo a escolha de termos que melhor representam a sua necessidade de informação. Além das contribuições e perspectivas, também foi possível identificar os principais desafios que envolvem a área pesquisada.
Palavras-chave: Recuperação de Informação. Interfaces de Busca. Ontologia.

ABSTRACT
Search engines currently available on the Web, most based on the text box, have had little significant changes in the last years regarding the form of interaction with the user. Although they have become popular because of their simplicity, still present numerous problems related mainly to the researcher's dependence on knowledge in relation to the investigated subject, to difficulties in expressing a need for information through keywords, to the problems of ambiguity of language and to the inadequate composition of search terms. In this context, this research has as objective to analyze the literature on information retrieval interfaces and to raise concepts and characteristics that can subsidize the proposal of an interface model based on ontologies, with resources able to assist the user during the search and to minimize the problems observed. In order to reach the proposed objectives, the methodology adopted in this study had exploratory and bibliographic research, based on relevant authors of the area of Information Science. As results, it can be seen that ontologies have been widely discussed and adopted in several experiments related to the development of search interfaces more friendly and easy to be used, providing significant improvements in the information retrieval process. Such a resource can allow the user a direct interaction with the concepts of the researched domain and through them, allowing the choice of terms that best represent your need for information. In addition to the contributions and perspectives, it was also possible to identify the main challenges involved in the researched area.
Keywords: Information Retrieval. Search Interfaces. Ontology.

Artigo recebido em 2/02/2019 e aceito para publicação em 24/04/2019.

1 Doutoranda no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho, Brasil. Professora do Instituto Federal do Paraná - Campus Jacarezinho, Brasil. E-mail: marcia.reis@ifpr.edu.br.
2 Doutor em Ciências da Comunicação pela Universidade de São Paulo, Brasil. Docente permanente do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho, Brasil. E-mail: ferneda@marilia.unesp.br.

1 INTRODUÇÃO

Os sistemas voltados para a recuperação de informação (RI) ganharam bastante destaque nos últimos anos, principalmente no contexto da Web, que possui um grande volume de informações disponível para a consulta. Tornaram-se populares por oferecem inúmeros benefícios na representação, organização, disseminação e recuperação de conteúdos, sendo amplamente utilizados por usuários com diferentes interesses, que buscam informações relevantes para atenderem suas necessidades.

Apesar da sua grande relevância, estudos realizados na área relatam que os mecanismos de busca atuais ainda estão distantes de atender às expectativas dos usuários, uma vez que é necessário despender um tempo significativo na tentativa de localizar uma informação desejada (ALLAN; CARTERETTE; LEWIS, 2005; HOGAN *et al.*, 2011; LAI; SOH, 2004). A eficiência do processo de recuperação de informação está relacionada às coincidências teminológicas entre os termos utilizados na expressão de busca e os termos de indexação utilizados para representar os documentos. Assim, tal eficiência depende do conhecimento que usuário possui da terminogia da área ou assunto que está pesquisando, permitindo-lhe expressar com mais precisão a sua necessidade de informação. Quanto maior o seu conhecimento do vocabulário relacionado ao domínio de interesse, mais opções de vocabulário o usuário terá à sua disposição para enriquecer as suas buscas (ANTONIOU; HARMELEN, 2004; BEPPLER, 2008).

O objetivo desta pesquisa é analisar conceitos e características sobre as interfaces de busca projetadas a partir do uso de ontologias e como elas contribuem para a redução dos problemas de interação e para a formulação de expressões de busca mais eficientes. Tem como foco subsidiar o desenvolvimento de uma interface diferenciada, com novos estilos de interação e que possa minimizar as dificuldades apresentadas pelas ferramentas de recuperação de informação atuais.

A metodologia adotada contou com pesquisa bibliográfica, de caráter exploratório, fundamentada na literatura da Ciência da Informação, mais especificamente em Recuperação de Informação, Interfaces de Busca e Ontologias, realizada em livros, artigos científicos, dissertações, teses e publicações de eventos. Além disso, foram selecionados autores que desenvolveram protótipos de ferramentas de busca baseadas em ontologias, como forma de subsidiar a proposta deste estudo e identificar as principais contribuições e os desafios observados por eles.

Os resultados indicam que há avanços significativos na área de recuperação de informação e que a utilização de ontologias é promissora e está trazendo grandes contribuições para as pesquisas e aplicações desenvolvidas. A partir de um determinado domínio, as ontologias se propoem a classificar

itens de informação em categorias, sendo empregada para organizar o conhecimento, descrever estruturas conceituais e tornar viável o uso de um vocabulário compartilhado (VITAL; CAFÉ, 2011). Desta forma, ao serem inseridas no contexto das interfaces de busca, possuem potencial para o desenvolvimento de recursos capazes de minimizar parte dos problemas da busca, principalmente os relacionados ao uso de terminologia e vocabulário adequados.

2 SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÃO NA WEB E OS DESAFIOS DO PROCESSO DE BUSCA

A dificuldade de encontrar uma informação relevante e de interesse do usuário, principalmente na Web, é um dos grandes problemas observados e discutidos no âmbito da Ciência da Informação (SARACEVIC, 1996). A Recuperação de Informação (RI) "é a área de pesquisa que se preocupa com a estrutura, análise, organização, armazenamento, recuperação e busca de informação" (SELTON, 1968 apud BEPPLER, 2008, p.15). Em função da quantidade de informações disponibilizadas na Web e da consequente necessidade de recuperá-las, a Recuperação de Informação é um campo que tem progredido significativamente nos últimos anos e para a qual pesquisadores têm dedicado esforços consideráveis (LIN; DEMNER-FUSHMAN, 2006).

> O objetivo é estudar e entender os processos de Recuperação da Informação para projetar, construir e testar sistemas de recuperação que possam facilitar a comunicação efetiva de informação desejada entre o gerador humano e o usuário humano (INGWERSEN, 2002, p.49, tradução nossa).

Os sistemas de recuperação de informação na Web têm como principal objetivo "minimizar as dificuldades do usuário em localizar a informação requisitada, ou seja, diminuir o tempo gasto em um processo de busca até que a informação desejada possa ser acessada" (CRISTOVÃO; DUQUE; SERQUEIRA, 2012, p.5-6). São de grande importância para a sociedade contemporânea e organizações em geral, que dependem de documentos e informações relevantes para a tomada de decisão e resolução de problemas (BEPPLER, 2008).

Segundo Huang (1999), apesar da recuperação de informação na Web ser uma das atividades mais utilizadas pelos usuários, os sistemas com este propósito requerem atenção especial sobre alguns aspectos importantes: a) a quantidade de informações disponibilizada é muito ampla, heterogênea e se modifica com muita frequência; b) existe alta variabilidade de formatos (textual, imagens, vídeos, etc.) e idiomas (inglês, português, etc.); c) a quantidade de documentos duplicados é volumosa, ou seja, a mesma informação pode ser encontrada em diversas fontes e elas, na maioria das vezes, não

estão correlacionadas; d) cada documento pode ter *links* para outros documentos, e estes para outros; e) existe uma estimativa de que 85% dos usuários que utilizam sites de busca não olham a segunda página dos resultados recuperados. Na maioria das vezes, estes fatores resultam na necessidade de despender tempo considerável na tentativa de localizar a informação requerida.

A recuperação de informação é um processo de comunicação complexo que ocorre entre o usuário e um acervo documental, realizado por meio de uma interface, com o objetivo principal de influenciar o estado cognitivo e satisfazer as necessidades informacionais do usuário (ARAUJO JUNIOR, 2007; SARACEVIC, 1997). No entanto, encontrar uma informação parcialmente ou totalmente relevante ao contexto pesquisado dependente do uso eficiente das ferramentas de busca. Para isso, o usuário precisa estabelecer um diálogo com o sistema na tentativa de expressar suas necessidades, que geralmente são traduzidas no formato de palavras-chave ou expressões de busca. O sucesso do resultado de uma busca depende da coincidência entre os termos fornecidos pelo usuário em sua expressão de busca e os termos utilizados na indexação dos documentos.

Por ser um ambiente fortemente influenciado pelos problemas linguísticos, semânticos e sintáticos, os usuários frequentemente apresentam dificuldades para traduzir suas necessidades de informação em um conjunto de termos por meio da interface disponibilizada pelo sistema. Entre os problemas observados, destacam-se: a falta de interatividade que as ferramentas de busca atuais oferecem; o uso de poucas palavras para formular consultas (em média 2,2 termos); a rara utilização de expressões booleanas; a composição inadequada de termos e palavras-chave e a obrigatoriedade de conhecimento do usuário sobre o domínio pesquisado, ou seja, quanto mais ele souber sobre o contexto de interesse, mais opções de terminologia e vocabulário terá à sua disposição para enriquecer as buscas, ao mesmo tempo em que, quanto menos conhecimento possuir, mais genéricas serão as expressões utilizadas (ANTONIOU; HARMELEN, 2004; BEPPLER, 2008).

Outro obstáculo de grande influência nos sistemas de RI refere-se às variações linguísticas e ambiguidade de ordem semântica e sintática da linguagem, que possibilitam interpretações diferentes para palavras ou termos (a palavra "letra", por exemplo, está relacionada à vários contextos diferenciados: elemento básico do alfabeto, o texto de uma canção ou a caligrafia de um determinado indivíduo).

> [A língua pode ser considerada um sistema] de uma complexidade extrema: compreende regras (de pronúncia, de formação de palavras, de formação de frases, de relacionamento das formas com os significados), itens léxicos (palavras e morfemas, com suas propriedades gramaticais e seus significados), expressões idiomáticas (como *pisar na bola* ou *mãe de santo*) e clichês (como *ficar sem fala* e *tomar café*) (PERINI, 2010, p.1).

Para tentar minimizar os problemas observados, técnicas importantes estão sendo estudadas, trazendo contribuições valiosas para as pesquisas e aplicações propostas em RI. Um recurso

importante que vem sendo analisado refere-se à utilização de ontologias no processo de recuperação de informação. Entre as diversas aplicações observadas, este recurso pode ser útil para padronizar a terminologia utilizada na comunicação do usuário com o sistema (na representação dos documentos do *corpus* e na representação da expressão de busca). As ontologias também pode ser utilizadas na indexação automática de documentos, na expansão de consultas, em funções de processamento de texto (classificação, extração e busca) e, por último, como interfaces de busca, com o intuito de reduzir os problemas linguísticos e auxiliar o usuário na elaboração de expressões de busca mais significativas e melhor contextualizadas (busca guiada por uma estrutura terminológica).

Apesar das inúmeras aplicabilidades de ontologias no contexto da recuperação de informação, este trabalho tem como objetivo discutir apenas a utilização de ontologias em interfaces de busca.

3 ONTOLOGIAS: DEFINIÇÕES, CARACTERÍSTICAS E APLICABILIDADE

O conceito de ontologias vem sendo amplamente discutido e utilizado em áreas distintas que englobam principalmente a filosofia, a ciência da computação e a ciência da informação, "tendo em vista a possibilidade de melhorar significativamente a representação de um domínio de conhecimento" (FERNEDA, 2003, p.26). Para Branco Neto (2006, p.74), "as ontologias representam o ponto mais elevado já atingido em termos de representação, compartilhamento e reutilização do conhecimento".

> As ontologias se colocam como um novo instrumento a ser incorporado ao arsenal teórico e prático da Ciência da Informação. A aprendizagem de novos conceitos e novos recursos oferecidos pelas ontologias é um desafio para os profissionais da informação, mas que pode ser facilmente enfrentado utilizando toda bagagem teórica acumulada durante a história da Ciência da Informação (FERNEDA, 2013, p.41).

Gruber (1993, p.199) define ontologia como sendo "uma especificação formal e explícita de uma conceitualização compartilhada". Segundo o autor, a conceitualização refere-se ao conjunto de conceitos, relações, objetos e restrições que são definidos para um modelo semântico de um domínio de interesse. As ontologias expressam o formalismo dos conceitos e das relações acerca de um domínio, ou seja, "define os termos usados para descrever e representar uma área de conhecimento" (OWL, 2003). Em outras palavras, elas codificam o conhecimento do domínio e também o conhecimento que se entende do domínio, tornando-o reutilizável. Segundo Gruber (1996), Noy e Mcguinness (2001), a ontologia formaliza o conhecimento através da utilização de classes (organizadas em uma taxonomia), relações (representam o tipo de interação entre os conceitos de um domínio), axiomas (usados para

modelar sentenças sempre verdadeiras) e instâncias (utilizadas para representar elementos específicos, ou seja, os próprios dados).

Deste modo, pode-se dizer que uma ontologia visa (em alguns aspectos) desenvolver um conjunto de regras que possibilitem a abstração da semântica das informações de um determinado domínio disponibilizadas na *Web*. Segundo Fensel (2003), isso oferece vantagens, tais como: possibilitar o compartilhamento e a interoperabilidade do conhecimento entre os domínios, estruturá-los de forma que permita sua compreensão com maior clareza e objetividade e permitir a reutilização dos conceitos abordados.

Branco Neto (2006) e Ferneda (2013) destacam algumas aplicabilidades proporcionadas pela utilização de ontologias, relacionadas principalmente à recuperação de informação: **a) checagem de consistência**: as informações obtidas em uma busca podem ser associadas às definições de uma ontologia para verificar se os valores retornados são válidos ou não; **b) complemento de informações**: após obter um pequeno conjunto de informações sobre determinada entidade, as ontologias podem ser utilizadas para gerar novas informações com base nas previamente encontradas; **c) resolução de problemas de linguagem** (ambiguidade, polissemia e sinônimos): se o mesmo termo possui mais de um significado, ele está relacionado à uma classe e subclasse correspondente que pode ajudar a distinguir entre os vários contextos do termo; **d) indexação automática**: a representação de um documento é realizada automaticamente por termos derivados de uma ontologia; **e) expansão de consulta**: adição de novos termos à consulta do usuário provenientes de uma ontologia; **f) sistemas de recuperação de informação semânticos**: os documentos são previamente anotados (marcados) de acordo com uma ontologia de domínio; e **g) interfaces de busca**: os conceitos de uma ontologia são apresentados ao usuário que seleciona aqueles que serão utilizados como termos de busca.

Apesar das diversas possibilidades de aplicação de ontologias na recuperação de informação, o tema ainda é considerado incipiente e apresenta grande potencial para o desenvolvimento de estudos e aplicações na área. As pesquisas atuais apontam que a principal utilização das ontologias está relacionada à representação dos documentos, mais especificamente o processo de indexação automática, extração de informação e expansão de consultas (FERNEDA, 2013).

> Ontologias são instrumentos de controle terminológico utilizados na recuperação da informação de domínios específicos do conhecimento. No tocante a sistemas informatizados, significam hoje o que há de mais avançado na área de representação do conhecimento e na promoção da interoperabilidade entre sistemas processados por máquinas (DIAS; COSTA, 2011, p.481).

No contexto deste estudo, as ontologias serão abordadas como artefatos de construção e desenvolvimento de interfaces de busca, com o objetivo de apresentar alternativas diferenciadas para a melhoria das estratégias de consulta e, consequentemente, do processo de recuperação de informação na Web.

4 CONCEITOS E CARACTERÍSTICAS DAS INTERFACES DE BUSCA BASEADAS EM ONTOLOGIAS

Para tentar minimizar os problemas observados no processo de recuperação de informação, vários estudos estão sendo direcionados para a utilização de ontologias em interfaces de busca, uma vez que este recurso possibilita a construção de componentes interativos, com condições de auxiliar a composição de expressões de busca, permitindo que o usuário de um sistema de recuperação utilize conceitos por definidos na ontologia para traduzir linguisticamente sua necessidade de informação (GARCIA; SICÍLIA, 2003).

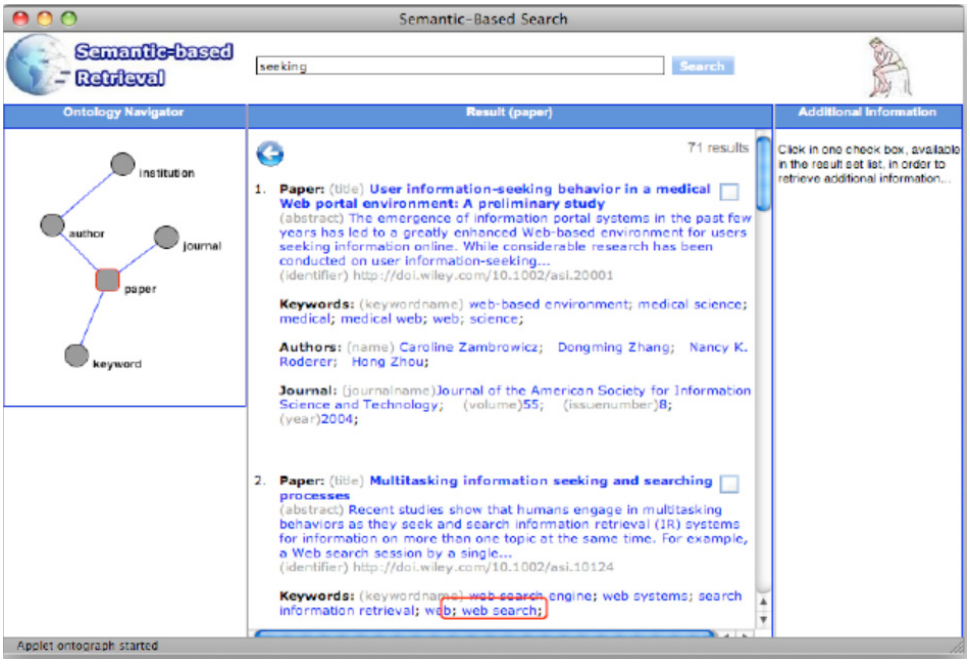
Com o acesso direto aos conceitos de uma ontologia, o usuário pode elaborar expressões de busca mais significativas, contextualizadas e obter um melhor entendimento sobre o domínio pesquisado. Além disso, a interação com documentos recuperados pode servir de base para refinar ou iniciar novas buscas. Tang (2007) afirma ainda que a possibilidade de interação direta com conceitos definidos por uma ontologia pode estimular os usuários a explorarem diferentes aspectos de suas necessidades, além de ser um recurso importante para ajudar na elaboração de expressões de busca, principalmente quando o usuário não sabe ao certo o que precisa ou quando não possui conhecimento sobre o assunto pesquisado.

As interfaces de busca baseadas em ontologias podem ser projetadas para interagir com os usuários de formas diferenciadas. Para Paulheim e Probst (2010), as aplicações podem optar por apresentar o conteúdo nos seguintes formatos: **a) nenhuma representação**: as ontologias são escondidas completamente, não possibilitando a visualização/interação pelo usuário. Neste caso, são utilizadas apenas na estruturação do sistema ou na camada de dados e não como componentes de interface; **b) listas de categorias**: as ontologias e suas relações são apresentadas ao usuário no formato de listas de seleção, taxonomias, relação de conceitos ou blocos de textos em lugares específicos da interface; **c) gráfico**: as relações entre os conceitos são visualizadas como árvores (na maioria das vezes mostrando a taxonomia) ou gráficos (incluindo as relações não taxonômicas). Existem diversas técnicas para visualização de ontologias, abrangendo representações gráficas em 2D e 3D (KATIFORI *et al*., 2007); **d) verbalizado**: a representação textual dos axiomas contidos na ontologia é fornecida ao usuário; **e) código-fonte**: o código-fonte da ontologia é mostrado ao usuário (**OWL ou F-Logic,** por exemplo).

Diversos trabalhos já foram desenvolvidos na tentativa melhorar o processo de recuperação de informação utilizando ontologias como componente de interface de busca. Beppler (2008) apresenta

o Hermeneus, que é um *framework* de recuperação e busca de informação que permite aos usuários navegar nos conceitos e instâncias de uma ontologia de domínio, que é o artefato principal para a criação de módulos e componentes. O sistema apresentado na Figura 1 foi desenvolvimento para ser utilizado em qualquer domínio e é composto por quatro componentes: **a) módulo de indexação**: utiliza uma ontologia de domínio para criar índices semânticos automaticamente; **b) módulo de recuperação**: responsável por recuperar as instâncias da ontologia que estão armazenadas nos índices semânticos em resposta a uma consulta; **c) módulo de apresentação**: utiliza uma ontologia de domínio para disponibilizar um ambiente interativo na interface do usuário. Esse módulo é composto de outros três componentes: navegador de ontologia, instâncias recuperadas e informação adicional; **d) módulo de inferência**: responsável por extrair informação adicional de uma determinada instância recuperada por meio da utilização de técnicas de inferência e regras pré-configuradas sobre a ontologia e sua base de conhecimento.

Figura 1 – Interface de Busca do Hermeneus



Fonte: Beppler (2008, p.77)

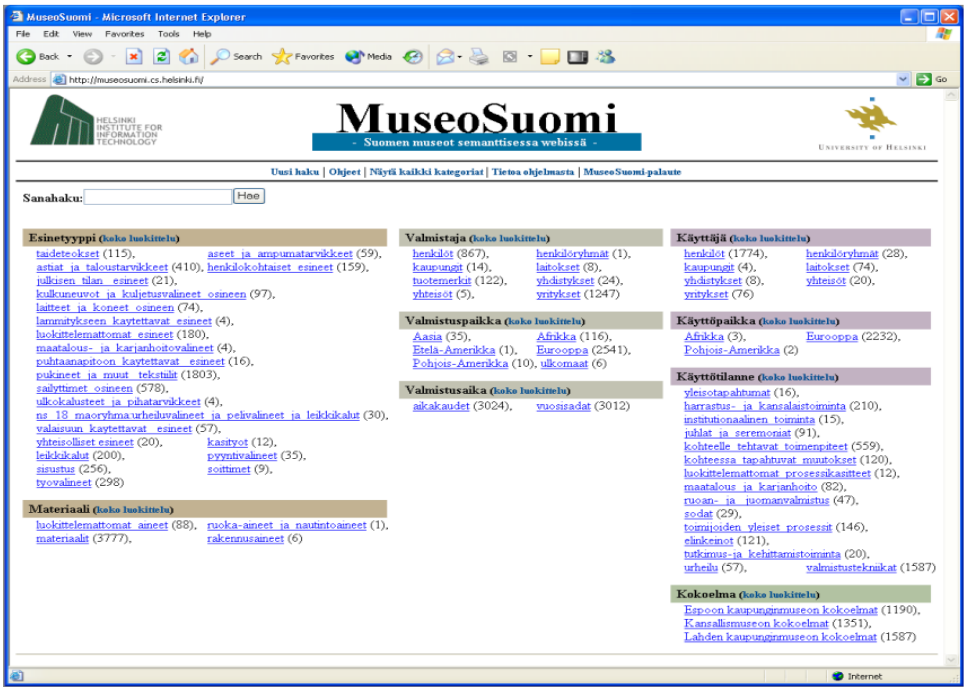
Segundo Beppler (2008, p.86), o objetivo de uma ferramenta gráfica "é facilitar que os seres humanos consigam adquirir percepções mais aguçadas sobre dados". O Hermeneus foi desenvolvido com objetivo de estimular o usuário a explorar a informação disponibilizada, além de auxiliar na construção de uma expressão de busca mais apropriada, mesmo quando não existe conhecimento sobre o domínio pesquisado. Possibilita

ainda a visualização de diferentes perspectivas das informações recuperadas, bem como o resultado de busca contextualizado, que pode servir de base para o desenvolvimento de novas buscas.

Ainda, de acordo com o autor, os testes realizados com o protótipo desenvolvido foram positivos e apresentaram contribuições importantes para a melhoria do processo de RI, tais como: a) elaboração de expressões de busca mais precisas, geradas a partir da interação com uma ontologia por meio de cliques do mouse; b) possibilidade do usuário adquirir, durante a interação com uma ontologia, novas percepções sobre as suas necessidades de informação; c) os usuários podem formular consultas menos ambíguas, uma vez que podem expressar claramente o contexto de cada termo utilizado na consulta; d) com base nos resultados apresentados, os usuários podem ter novas percepções sobre as suas necessidades, além de refinar ou formular novas consultas; e) as ontologias podem ser usadas para produzir diferentes visões do conhecimento, particularmente para propósitos de exploração e navegação e; f) possibilidade do usuário desenvolver uma série de tipos de interações sobre a interface de busca, que vão desde a reformulação de uma consulta, navegação e refinamento, até a verificação, análise de relevância e aprendizagem.

Outro exemplo importante para o contexto de interfaces baseadas em ontologias é o MuseumFinland, que é um portal semântico utilizado para a publicação de coleções culturais heterogêneas de museus, conforme pode ser observado na Figura 2.

Figura 2 – Interface de busca do MuseumFinland



Fonte: Mäkelä (2006, p.21)

O MuseumFinland foi apresentado no trabalho de Hyvönen e Mäkelä *et al*. (2005) e tem como base a utilização de ontologias para garantir a interoperabilidade e possibilitar busca semântica. Além disso, permite ao usuário utilizar os conceitos definidos na ontologia, que são disponibilizados no formato de taxonomia multifacetada, para elaborar consultas, refinar os resultados e navegar nos itens das coleções. Bonino *et al*. (2004) afirmam que o ponto chave para o processo de refinamento de uma busca semântica está na disponibilidade de uma ontologia de domínio, e na capacidade de compreender as relações semânticas entres os conceitos ontológicos.

Segundo Hyvönen e Mäkelä *et al*. (2005), o MuseumFinland se diferencia dos demais portais por apresentar um sistema de busca inteligente, baseado em conceitos ontológicos e não em palavras-chave. Além disso, as relações semânticas implícitas na ontologia possibilitam ao usuário entender o contexto da pesquisa e evitar problemas relacionados com ambiguidade, como também descobrir associações dentro do conteúdo do portal e utilizá-las para enriquecer a navegação.

Os estudos de Hyvönen e Mäkelä *et al*. (2005) destacam algumas contribuições para o processo de busca e recuperação de informação: a) o conteúdo apresentado no formato de taxonomia proporciona ao usuário uma visão geral do tipo de informação disponível (ou não) para a consulta; b) as hierarquias podem orientar o usuário na formulação das consultas, uma vez que disponibilizam vocabulário adequado para realizar pesquisas sob diferentes pontos de vista; c) as hierarquias não apresentam problemas de ambiguidade, relacionados principalmente com homônimos; d) as facetas podem ser utilizadas para auxiliar a navegação do usuário em busca de um conteúdo específico; e) o usuário pode refinar as consultas ou formular outras, por meio da seleção de subcategorias ou de novas categorias.

Assim como o MuseumFinland, o sistema de busca SemSearch proposto por Lei, Uren e Motta (2006) foi projetado para disponibilizar ao usuário a possibilidade de realizar buscas semânticas. De acordo com os autores, o sistema possui uma interface no estilo do Google, conforme pode ser observado na Figura 3, que tenta extrair a semântica dos termos digitados em uma consulta para recuperar documentos melhor contextualizados e mais relevantes. Neste sentindo, tem como objetivo criar um sistema de fácil utilização, capaz de auxiliar usuários a especificar consultas simples (até dois termos) ou complexas (três termos ou mais), de forma eficiente e com resultados mais precisos, mesmo para aqueles que não tenham familiaridade com o domínio pesquisado, escondendo, porém, a complexidade das buscas semânticas. Dessa forma, o sistema possibilita que todos os tipos de usuários (experientes ou não) tenham condições de obter os benefícios das tecnologias da Web Semântica, mesmo sem conhecer linguagens de consulta específicas.

Durante o processo de busca, o SemSearch compara cada termo utilizado na consulta com fontes de dados heterogêneas, na tentativa de encontrar conceitos, relações ou instâncias descritas. "Quando há casamento entre um ou mais termos da consulta com os termos presentes na ontologia, a consulta é alterada, tornando-se uma consulta descrita em linguagem formal, ou seja, uma consulta semântica" (BEPPLER, 2008, p.91).

Figura 3 – Interface de busca do SemSearch

Fonte: Lei, Uren e Motta (2006, p.15)

Uma avaliação inicial do SemSearch apresentada por Lei, Uren e Motta (2006) mostrou resultados positivos para o processo de recuperação de informação: a) o sistema possui uma interface simples, que minimiza os problemas de sobrecarga de informação; b) é capaz de produzir respostas precisas para as consultas realizadas pelos usuários; c) possibilita o desenvolvimento de consultas semânticas, mesmo que o usuário não tenha conhecimento das ferramentas da Web Semântica.

Além dos três experimentos analisados, diversas outras aplicações foram desenvolvidas utilizando ontologias com o propósito de minimizar problemas observados no processo de busca e recuperação de informação: CIRI (AIRIO *et al.*, 2004), OntoIR (GARCÍA; SICILIA, 2003), Textpresso

(MÜLLER; KENNY; STERNBERG, 2004), OnAIR (PAZ-TRILLO; WASSERMANN; BRAGA, 2005), OntoSeek (GUARINO; MASOLO; VETERE, 1999), OWLIR (FININ *et al.*, 2005), OntoSmart (FERNEDA, 2013), entre outras.

Apesar dos resultados positivos observados, das contribuições das ontologias no processo de RI e das perspectivas para o desenvolvimento de ferramentas de busca semântica e interfaces diferenciadas na Web, alguns problemas precisam ser analisados cuidadosamente. Segundo Lei, Uren e Motta (2006), as propostas de busca semântica ainda não estão sendo direcionadas para o usuário final, que não estão familiarizados com domínios específicos de dados semânticos, linguagens de consultas e navegação em interfaces guiadas por estruturas terminológicas (ontologias). Baeza-Yates e Ribeiro-Neto (2013) afirmam que não há um padrão muito bem estabelecido para a visualização dos resultados de busca, principalmente em caso de interfaces gráficas, além da dificuldade de representar respostas de grandes dimensões (grande quantidade de resultados), como é o caso no ambiente Web. Outra questão comum apresentada pelos autores é a falta de habilidade dos usuários em utilizar estruturas de navegação (ontologias, listas, categorias), uma vez que eles podem perder o contexto quando mergulham profundamente em um caminho específico. Asseguram ainda que as ferramentas visuais para a seleção e/ou navegação de conteúdo não foram implantadas em toda a Web porque elas ainda têm que demonstrar valor agregado ao usuário.

Katifori *et al.* (2007) também discutem alguns desafios encontrados no desenvolvimento de novas ferramentas de busca utilizando ontologias. Segundo eles, não existem avaliações comparativas relacionadas à eficácia dos métodos de visualização de ontologias para tarefas diferenciadas, tais como a busca. Isto significa que não basta projetar e implementar novos mecanismos de busca e recuperação de informação. É necessário acrescentar às pesquisas, avaliação rigorosa junto aos usuários para analisar quais tipos de recursos e interfaces são realmente eficazes e podem contribuir significativamente para o desempenho e a melhoria do processo de busca de informações na Web.

O quadro 1 apresenta uma síntese das principais contribuições, perspectivas e desafios levantados neste estudo sobre a utilização de ontologias em interfaces de busca na Web.

Quadro 1 – Síntese das principais contribuições, perspectivas e desafios da utilização de ontologias em interfaces de busca

Contribuições
• Permitem a utilização de conceitos da ontologia na elaboração de expressões de busca, minimizando os problemas linguísticos; • Possibilitam a formulação de consultas menos ambíguas, uma vez que os usuários podem expressar claramente o contexto de cada termo utilizado na consulta; • Disponibilizam vocabulário adequado ao usuário para a elaboração de expressões de busca mais significativas e melhor contextualizadas, realizadas a partir dos conceitos definidos na ontologia; • Possibilitam um melhor entendimento sobre os conceitos e instâncias de um domínio pesquisado; • Permitem a interação com documentos recuperados, servindo de base para refinar ou iniciar novas buscas; • Ajudam na elaboração de expressões de busca, principalmente quando o usuário não sabe ao certo o que deseja ou não possui conhecimento sobre o assunto pesquisado; • Estimulam os usuários a explorarem diferentes aspectos de suas necessidades, por meio da interação direta com conceitos definidos na ontologia; • Possibilitam a visualização de diferentes perspectivas das informações recuperadas; • Podem ser implementadas de forma a extraírem o significado semântico do termo digitado pelo usuário e, desta forma, recuperar documentos melhor contextualizados e mais relevantes.
Perspectivas
• Desenvolvimento de ferramentas de busca semântica e interfaces diferenciadas na Web; • Desenvolvimento de interfaces de busca com a possibilidade de oferecer ao usuário diversos tipos de interações, que vão desde a reformulação de uma consulta, navegação e refinamento, até a verificação, análise de relevância e aprendizagem; • Desenvolvimento de interfaces de busca que ofereçam aos usuários menores problemas de linguagem (ambiguidade, polissemia e sinônimos) • Desenvolvimento de ferramentas que ofereçam busca semântica de forma gráfica, substituindo as palavras-chave, tais como: lista de categorias, hierarquias e taxonomia multifacetada; • Desenvolvimento de sistemas com recursos variados que auxiliem o processo de busca e recuperação, tais como: indexação automática, expansão de consultas, checagem de consistência e complemento de informações.
Desafios
• As propostas de busca semântica ainda não estão sendo direcionadas para o usuário final; • Os usuários não estão familiarizados com domínios específicos de dados semânticos, linguagens de consultas e navegação em interfaces guiadas por ontologias; • Não existe um padrão bem estabelecido para a visualização dos resultados de busca, principalmente no caso de interfaces gráficas; • Ainda existem dificuldades para representar respostas com grande quantidade de resultados; • Falta de habilidade dos usuários em utilizar estruturas de navegação (ontologias, listas e categorias); • As ferramentas visuais para a seleção e/ou navegação de conteúdo ainda precisam demonstrar valor agregado ao usuário; • Ainda não existem avaliações comparativas relacionadas à eficácia dos métodos de visualização de ontologias para tarefas diferenciadas, tais como a busca; • A maioria das ferramentas desenvolvidas não foram testadas com usuários finais. O que se têm são apenas opiniões dos desenvolvedores sobre o seu desempenho e sua eficácia; • É necessário acrescentar avaliação rigorosa junto aos usuários para analisar quais tipos de recursos e interfaces são realmente eficazes.

Fonte: Propría

Como pode ser observado, a utilização de ontologias em interfaces de busca é promissora e apresenta inúmeras contribuições e perspectivas para a área. No entanto, não foram encontrados autores que comentassem sobre os três sistemas discutidos neste trabalho: o Hermeneus, o MuseumFinland e o SemSearch. Os dados apresentados e discutidos referem-se à opinião dos desenvolvedores. Também não foram apresentados resultados de testes exaustivos realizados com usuários finais, impossibilitando, portanto, uma análise mais aprofundada sobre o desempenho e a eficácia das ferramentas desenvolvidas, bem como sua aceitação pelo usuário final e sua contribuição efetiva no processo de recuperação de informação.

5 CONSIDERAÇÕES FINAIS

O desenvolvimento de sistemas de recuperação de informação utilizando ontologias, direcionados principalmente para a Web, vem se consolidando nos últimos anos, devido aos inúmeros benefícios evidenciados na representação, organização, disseminação e recuperação de conteúdos. Os pesquisadores da área têm direcionado esforços para projetar ferramentas que consigam minimizar os problemas observados no processo de busca e possam oferecer resultados mais relevantes aos usuários. Para isto, muitos estudos se concentram nas interfaces de busca baseadas em ontologias, com o propósito de acrescentar novas funcionalidades às interfaces atuais, torná-las mais amigáveis e interativas.

Os resultados apresentados ainda são parciais e referem-se às análises realizadas por meio da revisão da literatura. No entanto, foi possível identificar inúmeras contribuições que podem ser agregadas às interfaces de busca atuais, destacando-se principalmente a possibilidade de disponibilizar ao usuário um vocabulário adequado, baseado nos conceitos, instâncias e relações semânticas de uma ontologia de domínio, para ser utilizado na elaboração de expressões de busca mais significativas, contextualizadas e com a possibilidade de reduzir os problemas relacionados com a subjetividade inerente à língua (ambiguidade semântica e sintática, polissemia e sinônimos), que tanto dificultam o processo de recuperação de informação. Outra contribuição significativa está no fato de que, por meio da interação direta com os conceitos presentes na ontologia, o usuário terá mais condições para elaborar expressões de busca eficientes, mesmo que não tenha conhecimento sobre o contexto pesquisado. Também podem oferecer diversos tipos de interação para a formulação e reformulação de uma consulta, refinamento, análise de relevância e aprendizagem sobre o contexto pesquisado.

Embora tenham sido identificadas inúmeras contribuições sobre o assunto, ainda existem questões em aberto relacionadas à utilização de ontologias em interfaces de busca que precisam ser melhor exploradas. Entre os principais pontos levantados, está o envolvimento efetivo dos usuários nos estudos realizados, uma vez que a maioria dos trabalhos encontrados na literatura apresentam resultados baseados apenas opinião dos autores e desenvolvedores, impossibilitando, portanto, uma análise mais aprofundada sobre o desempenho, a eficácia, a usabilidade e as limitações das ferramentas desenvolvidas, bem como sua aceitação pelo usuário final e sua contribuição efetiva no processo de recuperação de informação.

No entanto, apesar dos problemas levantados, as ontologias possuem grande potencial para o desenvolvimento de aplicações mais sofisticadas e interfaces de busca contendo recursos gráficos que possam torná-las mais fáceis de serem utilizadas e capazes de auxiliar o usuário a expressar de forma mais precisa sua necessidade de informação. Além disso, as ontologias podem ser utilizadas para promover a interoperabilidade dos sistemas e implementar outros recursos importantes para a melhoria dos sistemas de RI, tais como: indexação automática, expansão de consultas, checagem de consistência e complemento de informações.

REFERÊNCIAS

AIRIO, E. *et al.* CIRI: an ontology-based query interface for text retrieval. In: HYVÖNEN, E.; KAUPPINEN T.; SALMINEN, M.; VILJANEN, L.; ALA-SIURU, P. (Eds). **Proceedings of the 11th Finnish Artificial Intelligence Conference**, 2004.

ALLAN, J.; CARTERETTE, B.; LEWIS, J. When will Information Retrieval be “Good Enough?”. In: PROCEEDINGS OF THE 28TH ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL (SIGIR ‘05). ACM, New York, NY, USA, p. 433-440, 2005.

ANTONIOU, G.; HARMELEN, F. Van. **A semantic web primer.** Cambridge: MIT Press, 2004.

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

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de informação**: conceitos e tecnologias das máquinas de busca. 2. ed. Porto Alegre: Bookman, 2013.

BEPPLER, F. D. **Um modelo para recuperação e busca de informação baseado em ontologia e no círculo hermenêutico**. 2008. 123 f. Tese (Doutorado em Engenharia e Gestão do Conhecimento) – Universidade Federal de Santa Catarina, Florianópolis, 2008.

BONINO, D. *et al*. Ontology driven semantic search. **WSEAS Transaction on Information Science and Application**, Issue 6, v. 1, p. 1597-1605, 2004.

BRANCO NETO, W. C. **Web Semântica na construção de sistemas de aprendizagem adaptativos**. 2006. 219 f. Tese (Doutorado em Ciência da Computação) - Universidade Federal de Santa Catarina, Florianópolis, 2006.

CRISTOVÃO, H. M.; DUQUE, C. G.; SERQUEIRA, L. D. Recuperação de informação: uma aplicação na criação e configuração automáticas de cursos virtuais a distância. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 13., 2012, Rio de Janeiro. **Anais…** Rio de Janeiro: Fiocruz, 2012.

DIAS, E. A. V.; COSTA, H. G. Mapeamento da produção científica no escopo da ontologia. **Sistemas & Gestão**, v. 6, n. 4, p. 481-507, 2011.

FENSEL, D. **Ontologies:** a silver bullet for knowledge management and eletronic commerce. 2. ed. New York: Springer-Verlag. 2003.

FERNEDA, E. **Ontologia como recurso de padronização terminológica de um sistema de recuperação de informação**. 2013. 96 f. Relatório de Pesquisa (Pós-Doutorado em Ciência da Informação) – Universidade Federal da Paraíba, João Pessoa, 2013.

FERNEDA, E. **Recuperação de informação**: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 2003. 137 f. Tese (Doutorado em Ciências da Comunicação) – Universidade de São Paulo, São Paulo, 2003.

FININ, T. *et al*. Information retrieval and the semantic web. In: PROCEEDINGS OF THE PROCEEDINGS OF THE 38TH ANNUAL HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES (HICSS'05). IEEE Computer Society, 2005.

GARCÍA, E.; SICILIA, M. A. Designing ontology-based interactive information retrieval interfaces. In: PROCEEDINGS OF THE WORKSHOP ON HUMAN COMPUTER INTERFACE FOR SEMANTIC WEB AND WEB APPLICATIONS, Springer Lecture Notes in Computer Science 2889, 2003. p. 152-165.

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

GRUBER, T. R. **What is na ontology?** 1996. Disponível em: http://www.ksl.stanford.edu/ kst/what-is-an-ontology.html. Acesso em: 12 nov. 2017.

GUARINO, N.; MASOLO, C.; VETERE, G. Ontoseek: content-based access to the web. **IEEE Intelligent Systems**, v. 14, n. 3, 1999.

HOGAN, A. *et al*. Searching and browsing Linked Data with SWSE: the semantic web search engine. **Web Semantics**: Science, Services and Agents on the World Wide Web, v. 9, n. 4, p. 365-401, dez. 2011.

HUANG, L. **A survey on Web information retrieval technologies**. New York, University of New York, 1999.

HYVÖNEN, E., MÄKELÄ, E. *et al*. MuseumFinland: finnish museums on the semantic web. **Journal of Web Semantics**, v. 3, n. 2-3, p. 224-241, 2005.

INGWERSEN, P. **Information retrieval interaction**. London: Taylor Graham, 2002. Disponível em: www.db.dk/pi/iri. Acesso em: 10 jan. 2018.

KATIFORI, A. *et al*. E. G. Ontology visualization methods: a survey. **ACM Computing. Surveys**, v.39, n.4, 2007.

LAI, J.; SOH, B. Similarity score for information filtering thresholds. ISCIT 2004. In: IEEE INTERNATIONAL SYMPOSIUM ON COMMUNICATIONS AND INFORMATION TECHNOLOGY, p. 216-221, 2004.

LEI, Y.; UREN, V.; MOTTA, Enrico. SemSearch: a search engine for the semantic web. **3rd European Semantic Web Conference (ESWC)**. Montenegro, 2006.

LIN, J.; DEMNER-FUSHMAN, D. The role of knowledge in conceptual retrieval: a study in the domain of clinical medicine. In: PROCEEDINGS OF THE 29th ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL (SIGIR '06). ACM, New York, NY, USA, 2006. p. 99-106.

MÄKELÄ, E. **View-based search interfaces for the semantic web**. Master's thesis, University of Helsinki, 2006.

MÜLLER, H. M.; KENNY, E. E.; STERNBERG, P. W. Textpresso: an ontology-based information retrieval andextraction system for biological literature. **PLoS Biol**, v. 2, n. 11. 2004.

NOY, N. F.; McGUINNESS, D. L. **Ontology development 101**: a guide to creating your first ontology. Stanford University, Stanford, 2001.

OWL. **Web Ontology Language Guide**. 2003. Disponível em: http://www.w3.org/TR/ 2003/CR-owl-guide-20030818/. Acesso em: 10 dez. 2017.

PAULHEIM, H.; PROBST, F. Ontology-enhanced user interfaces: a survey. **International Journal on Semantic Web and Information Systems**, v. 6, n. 2, p. 36-59, 2010.

PAZ-TRILLO, C.; WASSERMANN, R.; BRAGA, P. P. An information retrieval application using ontologies. **Journal of the Brazilian Computer Society**, v. 11, n. 2, 2005.

PERINI, M. A. Sobre língua, linguagem e Linguística: uma entrevista com Mário A. Perini. **ReVEL**, v. 8, n. 14, 2010.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SARACEVIC, T. The stratified model of information retrieval interactions: extension and applications. **American Society for Information Science**, p. 313-327, 1997.

TANG, M. C. Browsing and searching in a faceted information space: a naturalistic study of pubmed users' interaction with a display tool. **Journal of the American Society for Information Science**, v. 58, n. 13, p. 1998-2006, 2007.

VITAL, L. P.; CAFÉ, L. M. A. Ontologias e taxonomias: diferenças. **Perspectivas em Ciência da Informação**, v. 16, n. 2, p. 115-130, 2011.