

SISTEMA DE RECUPERAÇÃO DA INFORMAÇÃO: UMA ABORDAGEM BASEADA EM ONTOLOGIAS

Resumo: Com a expansão da grande rede mundial de computadores e o crescimento da popularidade da web, um amplo volume de dados é gerado e publicado em inúmeras páginas da internet nas quais se encontram dados das mais diversas áreas do conhecimento. Em consequência disso, surgiu uma grande demanda pela busca de informações relevantes neste emaranhado de dados, fazendo aflorar importantes estudos relacionados à Recuperação da Informação. Nesse cenário, a presente pesquisa evidencia-se como uma análise descritiva, baseada em revisão de literatura, sobre os Sistemas de Recuperação da Informação que utilizam as ontologias para melhorar os resultados da pesquisa informacional. E tem como objetivo descrever as características básicas dos Sistemas de Recuperação da Informação, apresentar considerações sobre ontologias e destacar como esses conceitos podem ser utilizados para melhorar a relevância das buscas. Com base nos trabalhos estudados, verificou-se que os sistemas que fazem uso de ontologias apresentam melhor precisão nos resultados das consultas se comparados com os modelos baseados em palavras-chave e que não utilizam recursos capazes de capturar as conceptualizações associadas às necessidades de conteúdo dos usuários.

Palavras-chave: Ontologia. Recuperação da Informação. Sistema de Recuperação da Informação.

Lúcia Helena de Magalhães
Universidade Federal de Minas Gerais
Doutoranda em Ciência da Informação pela
UFMG. Professora efetiva no Instituto
Federal de Educação Tecnológica - Campus
São João Del Rei.
lhmag@yahoo.com.br

Renato Rocha Souza
Fundação Getúlio Vargas
Doutor em Ciência da Informação pela
Universidade Federal de Minas Gerais e
pós-doutor em Tecnologias Semânticas para
Recuperação de Informação - University of
South Wales, UK. Professor e pesquisador
da Escola de Matemática Aplicada (EMAp)
da Fundação Getúlio Vargas
rsouza.fgv@gmail.com

INFORMATION RETRIEVAL SYSTEM: AN ONTOLOGY-BASED APPROACH

Abstract: Along with the broad growth of the World Wide Web and the web popularity, a large amount of data is generated and published on many Internet pages, in which different data, from various knowledge areas can be found. As a result, a large demand of relevant data search emerged in this tangle of data, producing important studies related to Data Recovery. In this scenario, this paper stands out as a descriptive analysis, based in a literature review, about the Information Retrieval Systems that handle ontologies to enhance the results from the informational research. Its purpose is to describe the basic characteristics of the Information Retrieval Systems, present considerations regarding ontologies and to highlight how these concepts could be used to improve the searches relevance. Based on the studied papers, it stands out that the systems adopting ontologies present better precision on consult results than the models based on keywords and that do not employ features capable of capturing the conceptualizations associated to content needs of users.

Keywords: Ontology. Information Retrieval. Information Retrieval System.

1 INTRODUÇÃO

O avanço tecnológico possibilitou o surgimento de computadores mais potentes, com grande capacidade de processamento e armazenamento. E com a expansão da grande rede mundial de computadores e o crescimento da popularidade da web, um amplo volume de informação é gerado e publicado em inúmeras páginas da internet nas quais se encontram as mais diversas áreas do conhecimento. Isso facilitou a coleta, disseminação, produção e junção de uma grande quantidade de informação em um ambiente digital. Além disso, as pessoas passaram a produzir mais dados e, ao mesmo tempo, coletar mais informações, para que elas ficassem disponíveis e de fácil acesso para auxiliar na tomada de decisão de forma rápida e segura. Em consequência disso, surgiu uma grande demanda pela busca de informações relevantes neste emaranhado de dados, fazendo aflorar importantes pesquisas relacionadas à Recuperação da Informação (RI).

Nesse contexto, os Sistemas de Recuperação da Informação (SRI) também passaram a ser de extrema importância, pois eles auxiliam os usuários a encontrar a informação relevante a partir da grande quantidade de dados disponíveis. Assim, esta pesquisa tem como objetivo descrever as características básicas dos SRI, bem como apresentar as considerações sobre as ontologias e destacar como esses conceitos podem ser utilizados para melhorar a precisão dos resultados da consulta. Assim, este artigo dedica-se à procura das respostas para o seguinte questionamento: Como as ontologias são utilizadas pelos Sistemas de Recuperação da Informação de forma a colaborar para que os resultados da busca sejam mais relevantes, precisos e eficazes?

Como estrutura de desenvolvimento deste artigo, a presente pesquisa evidencia-se como um estudo descritivo, baseado em revisão de literatura, cujos textos selecionados ao tema da pesquisa foram sistematizados. A revisão da literatura permite traçar um recorte sobre como as ontologias têm sido usadas pelos Sistemas de Recuperação da Informação de forma a melhorar os resultados da consulta. Ressalta-se que este trabalho é apenas um prefácio que poderá subsidiar uma pesquisa acadêmica mais aprofundada. Portanto, não esgota o estudo sobre o tema, mas contribui para a construção de uma linha de raciocínio teórica sobre o assunto.

2 RECUPERAÇÃO DA INFORMAÇÃO

Recuperação de Informação, segundo Salton (1968), é a área de pesquisa que se preocupa com a estrutura, análise, organização, armazenamento, recuperação e busca de informação. É o processo de pesquisar uma coleção de documentos a fim de identificar aqueles textos que tratam de um determinado assunto (LANCASTER, 1993). “É uma área abrangente da Ciência da Computação que se concentra principalmente em prover aos usuários o acesso fácil às informações de seu interesse” (BAEZA-YATES; RIBEIRO-NETO (2013, p. 1). Para Saracevic (1999), a RI pode ser considerada a vertente tecnológica da Ciência da Informação e é resultado da relação desta com a Ciência da Computação.

A Recuperação da Informação trata da representação, armazenamento, organização e acesso a itens de informação, como documentos, páginas Web, catálogos online, registros estruturados e semiestruturados, objetos multimídia, etc. A representação e a organização dos itens de informação devem fornecer aos usuários facilidade de acesso às informações de seu interesse (BAEZA-YATES; RIBEIRO-NETO, 2013, p. 1).

De acordo com Fachin (2009), o processo de RI está baseado em coleta, representação, armazenamento, organização e acesso por parte dos utilizadores. Recuperar 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.” (FACHIN, 2009, p. 3). Assim, para recuperar as informações que realmente são relevantes, a partir da totalidade de materiais disponíveis, algumas ferramentas se fazem necessárias para facilitar esse processo. Nesse contexto, os SRI oferecem recursos para auxiliar os usuários nessa tarefa.

3 SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO

A quantidade de informação disponível em ambiente digital, atualmente, é gigantesca e essa massa de dados tem aumentado cada vez mais, pois “todos os dias, quintilhões de bytes de dados são gerados provenientes do uso de mídia social e interações digitais” (DAS, 2017). Desse modo, encontrar a informação certa é uma tarefa desafiadora. Por isso, a importância de Sistemas de Recuperação de Informação eficientes para auxiliar o usuário na busca de informação relevante a partir de uma grande coleção de documentos.

Kowalski (1997) cita que o principal objetivo de um SRI é minimizar a dificuldade do usuário em localizar a informação desejada.

Os Sistemas de Recuperação da Informação são “... sistemas de comunicação que, entre outras funções, visam dar acesso às informações neles registradas.” (ARAÚJO, 1995, p. 15). É um conjunto de operações consecutivas executadas para localizar, dentro da totalidade de informações disponíveis, aquelas realmente relevantes (CESARINO, 1985). De acordo com Ferneda (2013), um SRI é um ambiente linguístico intermediário entre um acervo de informação e seus solicitantes. “Sua eficiência depende de um controle adequado da linguagem de representação dos itens de informação e das requisições de seus usuários.” (FERNEDA, 2013, p. 5). Assim, um dos principais problemas ou desafios dessas técnicas é recuperar somente informações que são interessantes para o usuário. Porém, infelizmente, esses sistemas apresentam relevância parcial em seus resultados de consultas devido às limitações dos modelos baseados em palavras-chave. Além disso, muitos dos SRI não utilizam recursos capazes de capturar as conceptualizações associadas às necessidades de conteúdo dos usuários. Dessa forma, métodos eficazes, capazes de recuperar informações relevantes e em tempo real, são de suma importância e imprescindível para o usuário que, com o auxílio desses recursos, não mais necessitará de fazer uma procura exaustiva para encontrar a informação de seu interesse, o que significa um considerável ganho de tempo em que o mesmo poderá dedicar-se a outras atividades.

A partir desse posicionamento, faz-se necessário um aprimoramento dessas tecnologias. Por isso, o interesse em um estudo sobre Sistemas de Recuperação da Informação, baseados em ontologia, orientado para a exploração de bases de conhecimento de domínio, que sejam capazes de suportar pesquisa semântica em grandes repositórios de documentos.

4 ONTOLOGIAS

Ontologia, na visão filosófica, é a ciência que estuda o que pode existir, ou seja, os tipos e estruturas de objetos, propriedades, eventos, processos e relações referentes a uma única realidade de cada área (SMITH, 2003). Segundo Almeida (2014), ontologia é um assunto que tem sido estudado em diferentes campos de pesquisa, tais como, Filosofia, Ciência da Computação e Ciência da Informação e no âmbito de vários domínios do

conhecimento como, por exemplo, Medicina, Biologia, Direito e Geografia. “Apesar do uso e difusão do termo, não é uma tarefa simples entender o que significa ‘ontologia’” (ALMEIDA, 2014, p. 242).

Campos (2010, p. 221) define uma ontologia como “um vocabulário comum para uma comunidade que precisa compartilhar informação em um determinado domínio. Inclui definições de conceitos básicos no domínio e as relações entre estes de forma que sejam interpretáveis por máquina”.

Para Lima e Maculan (2017), nas áreas da Ciência da Informação e da Ciência da Computação, a ontologia é estudada como instrumento de representação do conhecimento, em geral, como um artefato computacional. Além disso, as autoras afirmam que a ontologia é uma lista de conceitos ou entidades dentro de um domínio específico, os quais podem ser estruturados de forma hierárquica, por meio de relações semânticas explicitadas formalmente, em meio informatizado. Nesse contexto, “uma ontologia descreve formalmente termos e relações que podem existir entre eles dentro de determinado domínio, inteligível pelo computador.” (LIMA; MACULAN, 2017, p. 65).

De acordo com 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. O autor define uma ontologia como uma especificação explícita de uma conceituação e ainda relata que:

O termo é tomado por empréstimo da filosofia, onde uma ontologia é um relato sistemático da existência. Para sistemas baseados em conhecimento, o que ‘existe’ é exatamente aquilo que pode ser representado. Quando o conhecimento de um domínio é representado em um formalismo declarativo, o conjunto de objetos que podem ser representados é chamado o universo do discurso. Este conjunto de objetos, e as relações formalizadas entre eles, são refletidos no vocabulário representativo com o qual um programa baseado em conhecimento representa conhecimento. Portanto, nós podemos descrever a ontologia de um programa através da definição de um conjunto de termos representativos. Em tal ontologia, definições associam os nomes de entidades no universo do discurso (classes, relações, funções ou outros objetos) com texto legível para humanos, descrevendo o que os nomes significam e axiomas formais que restringem a interpretação e o uso bem formado destes termos (GRUBER, 1993, p. 199).

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. Uma ontologia pode ser considerada como um vocabulário de representação, geralmente

especializado em algum domínio ou assunto, qualificado por conceituações de tipos de objetos e suas relações no mundo. Ou seja, é um corpo de conhecimento que descreve algum domínio, utilizando um vocabulário de representação (CHANDRASEKARAN; JOSEPHSON; BENJAMIN, 1999).

Chandrasekaran e Josephson e Benjamin (1999) afirmam que uma ontologia pode ser considerada o coração de um sistema baseado em conhecimento, pois é por intermédio dela que o conhecimento de um domínio pode ser representado. Quanto à utilização das ontologias, de acordo com Sowa (2002), elas possibilitam o desenvolvimento de sistemas mais inteligentes, pois o conhecimento é representado formalmente de modo a suportar extração explícita e, mais importante, extração implícita de conhecimento.

Para Chandrasekaran e Josephson (1999), as ontologias são desenvolvidas com o objetivo de fornecer um nível semântico à informação de um dado domínio de forma a torná-la processável por máquinas e comunicável entre diferentes agentes, bem como softwares e pessoas. De acordo com 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. Segundo Fachin (2009), as ontologias podem ser vistas como soluções para a recuperação relevante na Web Semântica. São correlatas diretas dos tesauros (lista de palavras importantes de um dado domínio de conhecimento), 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 (FACHIN, 2009).

Bergman (2010) afirma que uma boa ontologia oferece um conjunto de benefícios que não estão disponíveis nas taxonomias, esquemas de base de dados relacionais ou outras formas padrão de estrutura de informação, como os tesauros. Entre os benefícios, o autor destaca a capacidade de representar qualquer tipo de informação, incluindo dados não estruturados, semiestruturados e estruturados. Dessa forma, o uso das ontologias pode facilitar a organização e recuperação das informações, ajudando a resolver os problemas de volume, variedade e veracidade (BRANQUINHO, BARACHO e ALMEIDA, 2015). Em virtude disso, acredita-se que as ontologias podem ser utilizadas no âmbito dos Sistemas de Recuperação da Informação com o objetivo de melhorar os resultados das pesquisas, pois a existência de relações hierárquicas sobre a semântica da informação poderá potencializar a eficiência dos mecanismos de busca (SAIAS, 2003).

Por conseguinte, uma pesquisa sobre os trabalhos que já foram publicados sobre o tema possibilitará um mapeamento de quem já escreveu e o que já foi escrito sobre o assunto.

Além disso, obter informações sobre a situação atual do problema pesquisado, de forma a conhecer publicações existentes e os aspectos que já foram abordados, é de grande valia para o avanço de novas pesquisas nesta área.

5 PROCEDIMENTOS METODOLÓGICOS

O objetivo do estado da arte em uma pesquisa é realizar uma investigação dos trabalhos existentes na área em que se deseja pesquisar, buscando identificar e avaliar toda pesquisa relevante relacionada ao tema para, assim, constatar o que esses trabalhos apresentam em comum, além de contribuir para uma compreensão mais abrangente do desenvolvimento da área. Assim, para esse estudo, realizou-se uma busca sistemática no Google Scholar, na biblioteca eletrônica Scielo, no Portal Capes e em Bancos de Teses e Dissertações, no período de outubro de 2017 a agosto de 2018. O Portal Capes facilita o acesso a várias bases de dados, porém, nesta pesquisa optou-se pelas bases Lisa, Ebsco, Web of Science e IEEE Xplore. A Lisa foi escolhida por ser direcionada aos profissionais da Biblioteconomia e Ciência da Informação. A Ebsco e a Web of Science, por abranger uma vasta área de conhecimento e permitir selecionar a área de conhecimento que se deseja pesquisar, além de apresentar boas opções de filtros. A IEEE Xplore é uma biblioteca digital que oferece acesso ao conteúdo científico e técnico publicado pelo IEEE (Instituto de Engenheiros Elétricos e Eletrônicos) e seus parceiros editoriais. Além das opções de filtro que essa poderosa ferramenta proporciona, ela exibe a quantidade de vezes que os artigos recuperados foram citados, o que facilita a identificação dos artigos de maior relevância.

Assim, como o objetivo de encontrar trabalhos relacionados com o tema “Sistemas de Recuperação da Informação baseados em Ontologias”, iniciou-se a pesquisa com a base EBSCO, utilizando como critério de inclusão os descritores “*Recuperação da Informação e Ontologias*” com a finalidade de recuperar os artigos publicados nos últimos dez anos. Nesse caso, apenas um artigo recuperado foi considerado relevante. Para tanto, optou-se por realizar novamente a consulta usando os mesmos descritores, porém no idioma inglês. Nesse caso, a Ebsco retornou 237 artigos. Para reduzir o número de artigos encontrados, foi reduzido o período de publicação para os últimos oito anos, aplicando os filtros para textos completos e idioma inglês. Nessa nova pesquisa, 154 documentos foram recuperados. Para selecionar os artigos principais, foi realizada uma leitura parcial nos resumos, dando atenção especial para

os artigos em que as palavras-chave estivessem mais relacionadas como o tema desta pesquisa e que foram mencionados um número maior de vezes.

O mesmo procedimento foi realizado nas demais fontes de pesquisa. Com base nos resultados das buscas, observou-se que existem trabalhos relacionados Sistemas de Recuperação da Informação baseados em ontologias, porém ainda há necessidade de novas pesquisas de forma a aprimorar as tecnologias já existentes e abranger novas áreas. Por fim, com os trabalhos selecionados, uma revisão sistemática foi realizada e apresentada, a seguir, por ordem cronológica.

6 SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO BASEADOS EM ONTOLOGIAS: ESTADO DA ARTE

O objetivo do estado da arte é realizar uma investigação nos trabalhos existentes na área em que se deseja pesquisar, buscando identificar e avaliar toda pesquisa relevante relacionada ao tema para, assim, constatar o que esses trabalhos apresentam em comum, além de contribuir para uma compreensão mais abrangente do desenvolvimento da área. Por conseguinte, os trabalhos estudados serão apresentados por ordem cronológica.

Saias (2003), em sua dissertação de mestrado, intitulada uma metodologia para a construção automática de ontologia e sua aplicação em Sistemas de Recuperação de Informação, afirma que um SRI não deve limitar-se à pesquisa direta de palavras em documentos. Assim, se faz necessário desenvolver um mecanismo capaz de extrair e processar a informação semântica de um texto. Desta forma, o autor sugere o uso de ontologias para representação semântica dos documentos num ambiente semântico web. O autor propõe a automatização deste processo, permitindo o ajuste dinâmico da ontologia à semântica encontrada nos documentos. E toma como ponto de partida o Sistema de Recuperação de Informação da Procuradoria Geral da República Portuguesa, já existente, e também a sua base documental e tenta conceber um sistema paralelo baseado em ontologias.

Para isso, foi construído um protótipo que analisa a base de textos em busca de elementos a incluir na ontologia e que representará, parcialmente, a semântica expressa nesses documentos.

A metodologia proposta por Saias (2003) faz uso de técnicas de Processamento de Língua Natural. Primeiramente, aplica-se um analisador sintático para obter uma árvore

sintática representativa das frases de cada texto. Em seguida, faz-se uma análise semântica, de onde se extraem triplos (sujeito, verbo, complemento) que dão origem a instâncias de entidades ou conceitos e ações envolvendo esses conceitos. E a ontologia de conceitos gerada é representada em OWL (*Ontology Web Language*) e associada aos documentos. E para possibilitar a resposta a interrogações sobre o conteúdo dos textos, foi implementado um motor de inferência baseado em regras de programação em lógica (SAIAS, 2003, p. 117). De acordo com o pesquisador, o sistema proposto utiliza a ontologia nos processos de dedução para responder a interrogações, visto que é a ontologia que representa a semântica parcial do documento. Para Saias, o objetivo do trabalho foi atingido, porém apresenta algumas limitações que precisam ser aprimoradas em trabalhos futuros, tais como, captar mais detalhes para a representação semântica da frase para que a interpretação pragmática tenha melhores resultados, também é necessário diversificar a base de textos, fazer experiências com textos de natureza diferente e comparar o comportamento do sistema. Além disso, o protótipo necessita ser testado e avaliado por utilizadores e o *feedback* dessa experiência usado para tornar o sistema mais robusto.

Airio *et al.* (2004) apresentam o sistema CIRI (*Concept-based Information Retrieval Interface*), que utiliza ontologias na indexação de documentos, criação e expansão de consultas. Inicialmente, o usuário escolhe a ontologia relacionada ao seu interesse de busca e seleciona os termos em uma representação hierárquica e visual dos conceitos da ontologia selecionada. A partir de um conjunto inicial de termos escolhidos pelo usuário, o sistema expande automaticamente a consulta, considerando os relacionamentos entre os conceitos da ontologia. Para isso, o CIRI apresenta uma interface para recuperação de informação baseada em conceito. E a representação de conceitos é baseada em três modelos de nível de abstração: conceitos, expressões nomeando o conceito e as ocorrências representando as expressões das consultas. Os conceitos são representados como hierarquias em árvore, onde os usuários podem navegar para obter uma ideia das relações conceituais em determinada área de assunto, e a partir do qual eles podem escolher os conceitos para suas buscas. Quando um Sistema de Recuperação é selecionado, uma consulta é formulada automaticamente a partir dos conceitos escolhidos. O usuário não precisa procurar nomes alternativos para conceitos nem se familiarizar com a sintaxe das linguagens de consulta. Por fim, como no momento da publicação do artigo o CIRI estava em fase de testes, o autor não fez uma análise em relação ao desempenho do sistema.

Castells, Fernández e Vallet (2007) propõem um sistema de recuperação que usa ontologias para anotação semiautomática de documentos. Os autores propuseram uma adaptação no modelo clássico de espaço vetorial para recuperação de informação e exploram bases de conhecimento baseadas em ontologia para melhorar a pesquisa em grandes repositórios. O modelo consiste em um esquema de anotação semiautomática nos documentos e em um método de recuperação de informação que utiliza o Modelo de Espaço Vetorial, incluindo um algoritmo capaz de lidar com diferentes valores de relevância para os termos, resultando em um sistema com buscas semânticas que apresenta os resultados ordenados por relevância. Segundo os pesquisadores, os experimentos realizados em um corpus relativamente grande mostraram que a abordagem apresenta melhorias nas consultas em relação aos sistemas baseados em palavras-chave (CASTELLS; FERNÁNDEZ; VALLET, 2007).

Silva (2008) apresenta uma arquitetura que se baseia no uso de ontologias para responder consultas avançadas. Para o autor, as ontologias possuem grande capacidade em representar conhecimento, provendo uma fonte muito rica para responder a requisições solicitadas pelo usuário. A arquitetura proposta por Silva (2008) utiliza um repositório central que concentra todos os documentos a serem pesquisados e recuperados. Além disso, existe uma tabela que referencia todos os termos aos documentos que cada qual se relaciona. Isso é similar às ferramentas de busca existentes atualmente. Porém, neste modelo existem duas fontes de dados, ou seja, os documentos e a ontologias. Com isso, não apenas as informações explícitas na busca serão recuperadas, mas também outras informações implícitas que somente poderão ser alcançadas através do uso de ontologias. Isso se deve ao fato das ontologias descreverem conhecimentos sobre uma determinada área do conhecimento.

A associação de uma ontologia a um documento resulta em um metadado do documento denominado anotação semântica. A anotação é uma descrição que pode trazer informações, tais quais: do que se trata o documento, quem é o autor, entre outros descritivos. (SILVA, 2008, p. 12).

Assim, na arquitetura proposta por Silva (2008), o módulo de recuperação, ao efetuar uma busca a partir da consulta de um usuário, irá relacionar um registro da tabela da base de dados a algum termo, instância ou conceitos de uma determinada ontologia. Segundo o autor, a recuperação de informação apoiada pela web semântica torna mais simples a tarefa de descoberta de informação, por parte do usuário. Além disso, com o suporte semântico, o

sistema apresenta resultados mais relevantes se comparado com a grande maioria das ferramentas de recuperação de informação que trabalha com dados textuais.

Para concluir, Silva (2008) afirma que a utilização de um repositório para documentos e outro para ontologias agregou maior poder organizacional das informações. Porém, apresenta algumas limitações que precisam ser revistas em trabalhos futuros, pois as ontologias que forem alteradas fora da ferramenta não têm como serem atualizadas dentro do repositório, se não por intervenção de algum usuário administrador. Outra limitação da arquitetura é permitir manipular somente documentos do pacote de escritório *OpenOffice*. Sendo assim, melhorias precisam ser realizadas para que o sistema suporte outros tipos de documentos.

Brisaboa *et al.* (2010) apresentaram uma arquitetura de um Sistema de Recuperação de Informação que leva em conta não somente os textos, mas também as referências geográficas incluídas nos documentos. Para isso, os autores usaram uma estrutura de índice que combina um índice textual (índice que associa cada palavra a uma lista de documentos que a contém) e um índice espacial com base em uma ontologia do espaço geográfico. Segundo os pesquisadores, poucas estruturas de índice ou algoritmos de Recuperação da Informação levam em conta a natureza espacial das referências geográficas incorporadas nos documentos. As técnicas textuais puras concentram-se apenas nos aspectos linguísticos e as técnicas espaciais puras concentram-se apenas nos aspectos geográficos dos documentos. Assim, os autores propõem uma estrutura de índice que faz uso de uma ontologia do espaço geográfico que descreve os conceitos desse domínio e as relações que os mantêm. Dessa forma, há melhoras no resultado da pesquisa, visto que a consulta pode ser facilmente expandida através das referências geográficas, porque a estrutura do índice é construída a partir de uma ontologia desse espaço.

Sousa (2011) desenvolveu uma arquitetura de um Sistema de Recuperação de Informação disponibilizada em vídeos, testado em palestras do domínio da Doutrina Espírita. Além disso, criou uma ontologia de domínio contendo principalmente os termos relacionados à Doutrina Espírita, onde os conceitos são usados para expandir as consultas avançadas realizadas pelos usuários. Nos experimentos, o autor observou um aumento significativo da precisão nas configurações que utilizaram a ontologia para Recuperação da Informação disponibilizada em vídeos por meio de indexação multimodal (áudio e imagem), visto que, normalmente, a Recuperação da Informação baseia-se na análise individual de um dos descritores disponíveis, ou seja, áudio, imagem ou texto.

Ferneda (2013), em sua pesquisa pós-doutoral, apresenta um modelo de recuperação de informação que utiliza ontologias de domínio como um elemento normalizador e unificador da linguagem de representação dos documentos e das buscas dos usuários.

Os vetores dos documentos são criados durante o processo de indexação automática, no qual as ontologias fornecem novos termos a fim de enriquecer semanticamente a representação dos documentos. O vetor de busca é criado a partir de um processo de expansão de consulta, no qual novos termos são inseridos na expressão de busca inicialmente formulada pelo usuário a partir de inferências realizadas em uma ontologia. (FERNEDA, 2013, p. 6).

Para Ferneda (2013), as ontologias são vistas como ferramentas de padronização terminológica das representações dos documentos e das buscas dos usuários de um SRI. Assim, o autor propôs um Sistema de Recuperação de Informação baseado em ontologia denominado OntoSmart, que tem por objetivo melhorar a precisão do processo de recuperação de informação. Porém, como o sistema, no momento que o trabalho foi publicado, ainda estava em desenvolvimento, não foi possível analisar de forma consistente o seu desempenho.

Pereira, Gonçalves e Santos (2013) apresentam uma metodologia de pré-processamento para investigar e recuperar informação em textos do século XIX. O Sistema de Busca de Textos Históricos (SBTH) recebe como entrada os documentos históricos em formato PDF e devolve como saída, um arquivo XML Semântico que possibilitará a busca semântica de informação. Uma ontologia de domínio também alimenta o sistema principal possibilitando assim as anotações semânticas. Segundo os pesquisadores, na etapa de pré-processamento, os documentos são convertidos em formato de texto puro, e, em seguida, recebem uma anotação nos conceitos de acordo com a ontologia desenvolvida junto com o especialista do domínio. A saída dessa etapa é um documento no formato TXT anotado de acordo com os termos presentes na ontologia de domínio. Assim, a busca por um termo permite também encontrar os outros termos presentes na ontologia e os relacionamentos entre eles.

Valbuena e Londoño (2014) apresentam um sistema de busca suportado por um sistema de indexação ontológico. Nesse caso, o sistema seleciona previamente algumas ontologias tais como Ontologia de Tipo de Célula (CL) e Ontologia de Proteína (PO), ambas pertencentes ao corpus CRAFT¹. Segundo os autores, o sistema apresenta melhorias no desempenho e maior precisão nos resultados mostrados aos usuários, apesar de exigir maior

¹ <https://sourceforge.net/projects/bionlp-corpora/files/CRAFT/v0.9/>

capacidade computacional e grandes quantidades de memória, a proposta de fazer uso de ontologias, para garantir que os resultados realmente correspondam ao domínio de busca, é satisfatória.

Coneglian, Fusco e Santarém (2015) apresentam uma arquitetura semântica de Recuperação da Informação. A pesquisa tem como objeto propor uma arquitetura que realize a recuperação de dados não estruturados de modo a permitir a interoperabilidade entre outros sistemas, podendo, assim, trazer informações com alto valor agregado às organizações. Utilizam-se dos conceitos e tecnologias da web semântica e de metadados para possibilitar que a recuperação das informações ocorra de modo mais semântico e inteligente, observando o contexto no qual as informações estão inseridas ao invés de basear apenas em palavras-chave.

A pesquisa procura resolver o problema da extração semântica e automática dos ambientes informacionais com o desenvolvimento de um agente semântico de extração de dados. O agente comunica com os espaços informacionais internos e externos de *Big Data* e baseia suas buscas em regras ontológicas fundamentadas num padrão de metadados para realizar a extração semântica do domínio proposto (CONEGLIAN; FUSCO; SANTARÉM, 2015).

Oshaiba, Houbay e Salah (2015) propõem no artigo “*Semantic Annotation for Biological Information Retrieval System*” uma estrutura que ajuda os pesquisadores da área biológica a recuperar documentos significativos que são relacionados com os termos declarados em uma Ontologia Genética (OG). O sistema utiliza a OG que é decomposta semanticamente em três subontologias (componente celular, processo biológico e função molecular). O usuário tem a flexibilidade de escolher os termos da pesquisa e expandir a consulta a partir da OG, que é usada para inferir termos relacionados semanticamente às palavras pesquisadas. Segundo os autores, adicionando sinônimos e as relações dos termos da ontologia genética à consulta, os resultados da pesquisa melhoram significativamente.

Ferneda e Neto (2016) apresentam um método que utiliza ontologia como um elemento normalizador e unificador da linguagem de representação dos documentos e das buscas dos usuários a fim de melhorar a precisão dos resultados alcançados no processo de recuperação de informação. Nesse trabalho, a ontologia é utilizada na expansão das consultas formuladas pelo usuário, por meio da inserção de novos termos derivados dos relacionamentos entre os conceitos da ontologia. Para isso, utilizam-se o Modelo Espaço Vetorial como estrutura formal para a representação dos documentos e das buscas dos

usuários. Os vetores dos documentos são criados durante o processo de indexação automática no qual uma ontologia fornece novos termos, além daqueles extraídos do texto, enriquecendo a representação do documento. O vetor de busca é criado com base em um processo de expansão de consulta no qual, a partir de inferências a uma ontologia, novos termos são inseridos na expressão de busca inicialmente formulada pelo usuário. Almeja-se, com isso, padronizar as representações dos documentos e das buscas, aperfeiçoando, deste modo, a precisão do processo de recuperação de informação (FERNEDA; NETO, 2016).

Na pesquisa de Ferneda e Neto (2016), os elementos linguísticos que formam uma ontologia são considerados termos de um vocabulário de domínio. O usuário define o seu domínio de interesse por meio da seleção de uma ontologia, que será utilizada para agregar novos termos à expressão de busca inicialmente formulada. “O Modelo Vetorial fornece a estrutura formal de representação tanto para os documentos como para as buscas, o que permite fornecer como resultado uma lista de documentos ordenados pelo grau de similaridade/relevância” (FERNEDA; NETO, 2016, p. 96).

Constâncio (2017) afirma que os motores de busca semântica, que fazem uso de ontologias para identificar os conceitos e ideias presentes nas decisões judiciais, podem minimizar os efeitos nocivos do chamado problema da Recuperação da Informação. Com esse entendimento, o autor desenvolveu uma ontologia para a construção de um motor de busca semântica. Primeiramente, realizou a conversão do Vocabulário Jurídico Controlado mantido pelo Superior Tribunal de Justiça do Brasil em uma ontologia denominada Vocabulario-2016-06. Em seguida, fez a combinação da ontologia Vocabulario-2016-06 com a ontologia JurisTJPR de Molinari, originando a ontologia chamada OntoLegis e, posteriormente, construiu um motor de busca semântica baseado na ontologia OntoLegis. Por fim, o pesquisador fez uma comparação do motor de busca semântico com o sistema de busca jurisprudencial do Tribunal de Justiça do Paraná. Segundo Constâncio (2017), os experimentos mostraram que o potencial do motor de busca semântico, no domínio da recuperação jurisprudencial, é real, pois as relações semânticas estabelecidas pela ontologia efetivamente permitem a seleção mais criteriosa de documentos e melhora a precisão dos resultados. Porém, segundo o autor, a metodologia ainda necessita de aperfeiçoamentos.

Por conseguinte, o levantamento dos trabalhos relacionados com a utilização de ontologia no processo de recuperação de informação nos permitiu verificar que há estudos relacionados ao uso de ontologias para melhorar a precisão dos resultados da consulta em ambientes controlados como visto nas pesquisas de Saias (2003), Airio et al. (2004), Castells,

Fernández e Vallet (2007), Silva (2008), Brisaboa et al. (2010), Sousa (2011), Ferneda (2013), Pereira, Gonçalves e Santos (2013), Valbuena e Londoño (2014), Oshaiba, Houbly e Salah (2015), Coneglian, Fusco e Santarém (2015), Ferneda e Neto (2016) e Constâncio (2017). Porém, acredita-se que o estudo de Saias foi bastante audacioso, ao propor um modelo automático para criação de uma ontologia. Na proposta de Saias (2003), a análise semântica está limitada ao interior da frase não levando em consideração o contexto. Além disso, a ontologia é baseada em uma coleção de documentos e não representa um conjunto de conceitos e seus relacionamentos de um domínio de conhecimento específico.

Logo, o levantamento sobre os trabalhos relacionados neste campo de saber foi de extrema importância para documentar o que está sendo feito, pesquisado e publicado atualmente. Além de evidenciar lacunas e oportunidades de novas perspectivas nesta área de pesquisa.

7 CONSIDERAÇÕES FINAIS

Sem a intenção de esgotar a discussão do tema, o estudo nos possibilitou entender que a eficiência de um Sistema de Recuperação de Informação depende da terminologia utilizada pelos usuários nas expressões de busca para representar as suas necessidades de informação. Desta forma, as ontologias assumem um papel importante na padronização terminológica das consultas, por permitirem a definição de um vocabulário específico de um determinado domínio, minimizando ambiguidades semânticas.

Foi possível observar nos trabalhos estudados que as ontologias contribuem para que o resultado da busca seja mais eficiente. Conforme visto na pesquisa de Airio *et al.* (2004), em que os autores apresentaram um Sistema de Recuperação da Informação que utiliza ontologias na indexação de documentos, criação e expansão de consultas. Assim como nos estudos de Castells, Fernández e Vallet (2007) que propuseram um Sistema de Recuperação que usa ontologias para anotação semiautomática de documentos para melhorar a pesquisa em grandes repositórios de textos. Silva (2008) também confirmou essa eficiência ao apresentar uma arquitetura que se baseia no uso de ontologias para responder consultas avançadas e assim recuperar informações que não estão evidentes nos documentos. Brisaboa *et al.* (2010) usaram uma estrutura de índice baseado em uma ontologia do espaço geográfico que descreve os conceitos desse domínio e as suas relações e com isso obtiveram melhoras no resultado da

pesquisa. Sousa (2011) criou uma ontologia de domínio contendo termos disponibilizados em vídeos sobre a Doutrina Espírita, em que os conceitos são usados para expandir as consultas avançadas realizadas pelos usuários e desse modo obteve um aumento significativo da precisão dos resultados. O Sistema de Recuperação de Informação proposto por Ferneda (2013) também faz uso de ontologias de domínio para padronização terminológica das representações dos documentos e das buscas dos usuários e, dessa forma, conseguiu melhor precisão nas respostas. Pereira, Gonçalves e Santos (2013) alimentaram um sistema principal com uma ontologia para anotações semânticas. Assim, os pesquisadores conseguiram melhorar o resultado da consulta, pois ao pesquisar por um termo, outras palavras relacionadas presentes na ontologia também eram encontradas. Valbuena e Londoño (2014) apresentam um sistema de busca suportado por um sistema de indexação ontológico que conseguiu melhorias no desempenho e maior precisão nos resultados mostrados aos usuários. Coneglian, Fusco e Santarém (2015) utilizaram os conceitos e as tecnologias da web semântica e de metadados para possibilitar que a recuperação das informações ocorra de modo mais semântico e inteligente. Oshaiba, Houbay e Salah (2015) propuseram uma metodologia que ajuda os pesquisadores da área biológica a recuperar documentos significativos que são relacionados com os termos declarados em uma Ontologia Genética. De acordo com os estudiosos, ao adicionar sinônimos e as relações dos termos da ontologia à consulta, os resultados da pesquisa melhoram significativamente. Nos estudos de Ferneda e Neto (2016), o usuário define o seu domínio de interesse por meio da seleção de uma ontologia que é utilizada para agregar novos termos à expressão de busca inicialmente formulada e assim conseguem melhores resultados. Constâncio (2017) desenvolveu uma ontologia da área jurídica para a construção de um motor de busca semântico e afirmou que os experimentos mostraram que o potencial do motor de busca semântico, no domínio da recuperação jurisprudencial, é relevante. Por conseguinte, as pesquisas mostraram que os Sistemas de Recuperação da Informação apresentaram melhorias nas consultas se comparados com os sistemas baseados apenas em palavras-chave.

Portanto, se faz necessário um estudo multidisciplinar para a construção de ontologias, pois é necessário o aprofundamento de questões semânticas, axiomas e sinônimos de determinado domínio do conhecimento. Assim, parcerias entre as áreas de Ciência da Informação e da Ciência da Computação são necessárias de forma a concretizar os recursos tecnológicos em prol da precisão, relevância e eficácia dos Sistemas de Recuperação de Informação e a satisfação dos usuários.

Por conseguinte, o estudo realizado nos possibilitou observar que os Sistemas de Recuperação de Informação, tais como o Google, têm como objetivo recuperar toda a informação disponível na Web. Porém, o fato de não possuírem uma delimitação semântica dos termos com os quais os documentos e as buscas do usuário são representados afetam negativamente na exatidão dos resultados da pesquisa. Desta forma, há necessidade de novas pesquisas nessa área, pois a utilização das tecnologias da web semântica como um todo é até então algo inviável, pois ainda não é possível o desenvolvimento de ontologias tão avançadas que possam cobrir todos os campos do conhecimento existentes, passíveis de classificar todos os termos e conceitos existentes ou que sejam capazes de comunicar-se umas com as outras, de modo a interoperar todos os dados.

Existem diversas pesquisas, mais ou menos coordenadas, para tratar do problema da recuperação das informações. Mas ainda pode-se esperar por melhorias em diversas áreas como a recuperação multilíngue, a recuperação por conteúdo em mídias diversas, como imagens, áudio e vídeo e as buscas em linguagem natural. Assim, pode-se afirmar que a Recuperação da Informação continua sendo uma das áreas de interesse de pesquisas, devido à quantidade de informação disponível e a necessidade de localização rápida do conteúdo relevante para agilizar a tomada de decisão.

Sendo assim, para que um Sistema de Recuperação de Informação tenha a capacidade de disponibilizar as informações relevantes e que estão acessíveis na web, é necessária a anotação semântica nos documentos das mais diversas fontes informacionais. Isso posto, espera-se que, com o uso de ontologia agregada em ambientes de Recuperação de Informação, pessoas e organizações possam lidar com as transformações apresentadas pelo avanço das tecnologias de Informação e Comunicação, recuperando informações significativas e transformando-as em novas possibilidades de criação do conhecimento.

Na realidade, os desafios são inúmeros e não há respostas concretas capazes de solucionar todos os problemas relacionados à Recuperação da Informação, mas muitas outras abordagens ainda podem ser encontradas na literatura devido à relevância do tema. Assim, importantes são os novos desafios intelectuais que apresenta essa área para a comunidade de pesquisa, pois aplicações capazes de analisar, selecionar e recuperar informações com precisão e qualidade, e de acordo com a necessidade de cada usuário, ainda são necessárias. À vista disso, espera-se que, a partir das reflexões expostas neste trabalho, novas pesquisas sejam realizadas sobre a temática, visando ampliar as investigações sobre a aplicação e uso

das ontologias pelos Sistemas de Recuperação da Informação com o intuito de melhorar a precisão das consultas.

REFERÊNCIAS

- AIRIO, E.; JÄRVELIN, K.; SAATSI, P.; KEKÄLÄINEN, J.; SUOMELA, S. **CIRI** – an ontology-based query interface for text retrieval. In: HYVÖNEN, E.; KAUPPINEN T.
- ALMEIDA, M. B. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. **Perspectivas em Ciência da Informação**, v. 19, n. 3, p. 242–258, set. 2014.
- ARAÚJO, V. M. R. H. de. Sistemas de Informação: nova abordagem teórico-conceitual. **Ciência da Informação**, Brasília, v. 24, n. 1, 1995. Disponível em: <http://dici.ibict.br/archive/00000141/01/Ci%5B1%5D.Inf-2004-577.pdf>. Acesso em: 7 jun. 2017.
- BAEZA-YATES, R. RIBEIRO-NETO, B. **Recuperação de Informação: Conceitos e tecnologia das máquinas de busca**. Tradução técnica: Leandro Krug Wives, Viviane Pereira Moreira. 2. ed. Porto Alegre: Bookman, 2013.
- BERGMAN, M. K. **Listing of 185 ontology building tools**. 2010. Disponível em: <http://www.mkbergman.com/904/listing-of-185-ontology-building-tools/>. Acesso em: 09 jul. 2017.
- BRANQUINHO, L. P.; BARACHO, R. M.; ALMEIDA, M. M. **Descoberta de conhecimento com uso de ontologias na mineração de dados**. Abakós, v. 1, n. 1, PUCMINAS. Disponível em: <http://mba.eci.ufmg.br/downloads/10580-39382-1-PB.pdf>. Acesso em: jun. 2017.
- BRISABOIA, N. R.; LUANCES, M. R.; PLACES, A. S.; SECO, D. (2010). Exploiting geographic references of documents in a geographical information retrieval system using an ontology-based index. **Geoinformática** 14 (3), 307–331. Disponível em: <https://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs10707-010-0106-3.pdf>. Acesso em: ago. 2018.
- CAMPOS, M. L. de A. O papel das definições na pesquisa em ontologia. **Perspectiva em Ciência da Informação**, v. 15, n. 1, p. 220-238, jan/abr. 2010. Disponível em: <http://www.scielo.br/pdf/pci/v15n1/13.pdf>. Acesso em: 13 out. 2019.
- CASTELLS, P.; FERNANDEZ, M.; VALLET, D. **An Adaptation of the VectorSpace Model for Ontology-Based Information Retrieval**. IEEE Trans. Knowl. Data Eng., [s.l.], v. 19, n. 2, p. 261-272, fev. 2007. Institute of Electrical & Electronics Engineers (IEEE). Disponível em: <http://dx.doi.org/10.1109/tkde.2007.22>. Acesso em: 10 jun. 2017.
- CESARINO, M. A. da N. Sistemas de Recuperação da Informação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 157-168, set. 1985. Disponível em: <http://www.brapci.inf.br/index.php/article/view/0000009051/>. Acesso em: 17 out. 2019.
- CHANDRASEKARAN, B.; JOSEPHSON, J. R.; BENJAMINS, V. R. What Are Ontologies, and Why Do We Need Them? In: **IEEE Intelligent Systems**, v. 14, n. 1, p. 20-6, jan. 1999.
- CONEGLIAN, C. S; FUSCO, E; SANTARÉM, J. E. Arquitetura Semântica de Recuperação da Informação. **Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015**.

Disponível em: <http://dcevents.dublincore.org/IntConf/dc-2015/paper/download/390/400>. Acesso em: 17 out. 2019.

CONSTÂNCIO, A. S. **Ontologia para um motor de busca semântica para recuperação jurisprudencial no Brasil**. Dissertação (Mestrado). Universidade Federal do Paraná, Curitiba, 2017. Disponível em: <https://acervodigital.ufpr.br/bitstream/handle/1884/46276/R%20-%20D%20-%20ALEX%20SEBASTIAO%20CONSTANCIO.pdf?sequence=1&isAllowed=y>. Acesso em: 18 out. 2018.

DAS. **Formação Cientista de Dados**. Curso de Machine Learning ofertado por Data Science Academy. E-book. 2017.

FACHIN, G. R. B. Recuperação inteligente da informação e ontologias: um levantamento na área da Ciência da Informação. **Biblos**: revista do Instituto de Ciências Humanas e da Informação, Rio Grande, v. 23, n. 1, p. 259-283, 2009. Disponível em: http://www.brapci.inf.br/repositorio/2010/05/pdf_363eb96084_0010275.pdf. Acesso em: 17 out. 2019.

FERNEDA, E. **Ontologia como recurso de padronização terminológica em um Sistema de Recuperação de Informação**. Relatório de Pesquisa (Pós-doutorado). UFPB. Paraíba, 2013. Disponível em: <https://www.marilia.unesp.br/Home/Instituicao/Docentes/EdbertoFerna/pos-doutorado.pdf>. Acesso em: 17 out. 2019.

FERNEDA, E; NETO, J. Ontologia como recurso de padronização terminológica no processo de recuperação de informação. **Informação em Pauta**, Fortaleza, v. 1, n. 1, p. 30-45, jan./jun. 2016. Disponível em: <http://oaji.net/articles/2017/6265-1530036892.pdf>. Acesso em: 17 out. 2019.

GRUBER, T. R. **A translation approach to portable ontologies**. Knowledge Acquisition, v. 5, n. 2, p. 199-220, 1999. Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/12822/pdf_3. Acesso em: 09 jul. 2017.

GUARINO, N. Formal Ontology in Information Systems. **Proceedings of FOIS'98**, Trento, p. 6-8, jun. 1998. Disponível em: <https://klevas.mif.vu.lt/~donatas/Vadovavimas/Temos/OntologiskaiTeisingasKonceptinisModeliavimas/papildoma/Guarino98-Formal%20Ontology%20and%20Information%20Systems.pdf>. Acesso em: 3 jul. 2017.

KOWALSKI, G. J. **Information Retrieval Systems: Theory and Implementation**: Kluwer Academic Publishers. (Multimedia Information Retrieval: Content-Based Information Retrieval from Large Text and Audio Databases), 1997.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. Trad. Antônio Agenor Briquet de Lemos. Brasília: Briquet de Lemos/Livros, 1993. 347p.

LIMA, G. A. de; MACULAN, B. C. M. dos S. Estudo comparativo das estruturas semânticas em diferentes sistemas de organização do conhecimento. **Ciência da Informação**, Brasília, v. 46, n. 1, p. 60-72, jan./abr. 2017. Disponível em: <http://revista.ibict.br/ciinf/article/view/4014/3453>. Acesso em 13 out. 2019.

MORSHED, A.; SINGH, R. Evaluation and Ranking of Ontology Construction Tools. **Unitn-eprints Research**. Mar. 2005. Disponível em: <http://eprints.biblio.unitn.it/archive/00000747/01/013.pdf>. Acesso em: 10 jun. 2017.

OSHAIBA, M. M. Z.; HOUBY, E. M. F. E.; SALAH, A. Semantic Annotation for Biological Information Retrieval System. **Advances in Bioinformatics**, 2015. Disponível em: <https://www.hindawi.com/journals/abi/2015/597170/>. Acesso em: 10 ago. 2018.

PEREIRA, J. W.; GONÇALVES, M. R. B.; SANTOS, M. T. P. Pré-Processamento para Recuperação de Informação em Textos Históricos do Século XIX. **Symposium on Knowledge Discovery, Mining and Learning - KDMiLe**, 2013. Disponível em: <https://homepages.dcc.ufmg.br/~gfrancis/kdmile/papers/paper%2010%20-%20Juliana%20e%20Marilde.pdf>. Acesso em: 18 out. 2018.

SAIAS, J. M. G. **Uma Metodologia para a construção automática de Ontologias e a sua aplicação em Sistemas de Recuperação de Informação**. Dissertação (Mestrado). Universidade de Évora. 2003. Disponível em: <http://www.di.uevora.pt/~jsaias/data/dissertacao-mei-js.pdf>. Acesso em: 16 out. 2019.

SALTON, G. **Automatic Information Organization and Retrieval**. New York: McGraw-Hill. 1968.

SARACEVIC, T. Information Science. **Journal of the American Society for Information Science**, v. 50, n. 12, p.1051-1063, 1999. Disponível em: <http://tefkos.comminfo.rutgers.edu/SaracevicInformationScienceELIS2009.pdf>. Acesso em: 10 jun. 2017.

SILVA, E. M. **Arquitetura para Recuperação de Informação em Documentos Anotados usando semântica**. Dissertação (Mestrado). UFSC. Santa Catarina, 2008.

SILVA, R. E.; SANTOS, P. L. V. A.C.; e FERNEDA, E. Modelos de Recuperação de Informação e Web Semântica: a questão da relevância. **Informação & Informação**, v. 18, n. 3, 2013. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2013v18n3p27>. Acesso em: 10 jun. 2017.

SMITH, B. Ontology. In: L. Floridi (Ed.). **Blackwell Guide to the Philosophy of Computing and Information**. Oxford: Blackwell, 2003. Ontology, p.155-166.

SOWA, J. F. Architectures for Intelligent Systems. **Special Issue on Artificial Intelligence of the IBM Systems Journal**, v. 41, n. 3, pp. 331-349, 2002. Disponível em: <http://www.jfsowa.com/pubs/arch.htm>. Acesso em: 5 jun. 2017.

SOUSA, E. E. **Uso de ontologia para recuperação da informação disponibilizada em vídeos por meio de indexação multimodal**. Brasília, 2011. 88 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Ciência da Informação, Universidade de Brasília. Disponível em: <http://repositorio.unb.br/handle/10482/9923>. Acesso em: 17 out. 2019.

SOUZA, R.R. 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/scielo.php?pid=S1413-99362006000200002&script=sci_abstract&tlng=pt. Acesso em: 17 out. 2019.

VALBUENA, S. J.; LONDOÑO, J.M. (2014). **Búsqueda de documentos basada en el uso de índices ontológicos creados con MapReduce**. Ciencia e Ingeniería Neogranadina, v. 24, n. 2, p. 57-75, 2014. Disponível em: <http://www.scielo.org.co/pdf/cein/v24n2/v24n2a04.pdf>. Acesso em: 10. ago. 2018.