

Indexação automática e ontologias: identificação dos contributos convergentes na ciência da informação

Maria da Graça de Melo Simões

Doutora em Ciência da Informação pela Universidad de Salamanca (USAL) - Salamanca - Espanha.

Professora da Universidade de Coimbra (UC) - Coimbra - Portugal.

<http://lattes.cnpq.br/1226531375978630>

E-mail: gsimoesuc@gmail.com

Luís Miguel Oliveira Machado

Doutorando pela Universidade de Coimbra (UC), Coimbra, Portugal. Pesquisador pela Universidade de Coimbra (UC), Coimbra, Portugal.

<http://orcid.org/0000-0003-3403-5618>

E-mail: luismachado@fe.uc.pt

Renato Rocha Souza

Pós-doutorado pela University of South Wales (Southwales), Gales, UK. Pós-Doutorado pela Columbia University (Columbia) - New York - Estados Unidos. Doutor em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG) - Brasil. Professor e Pesquisador da Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil. Professor e Pesquisador da Escola de Matemática Aplicada (EMAp) da Fundação Getúlio Vargas (FGV) - Rio de Janeiro, RJ - Brasil.

<http://lattes.cnpq.br/4726949697973381>

E-mail: renato.souza@fgv.br

António Tavares Lopes

Doutorando em Informação e Comunicação em Plataformas Digitais pela Universidade do Porto (U.Porto) - Porto, Portugal. Assistente Convidado em Ciência da Informação na Faculdade de Letras da Universidade de Coimbra - Coimbra, Portugal.

<http://orcid.org/0000-0002-8876-5387>

E-mail: atlopes@fl.uc.pt

Submetido em: 10/07/2017. Aprovado em: 05/09/2017. Publicado em: 28/12/2017.

RESUMO

A indexação, humana e automática, foi desde sempre uma preocupação da área da ciência da informação (CI). A falta de consistência na indexação humana e de semântica na automática são apontadas como as grandes desvantagens. Uma possibilidade de dotar sistemas computacionais com maior capacidade de inferência reside na utilização de ontologias. Desta proposição objetiva-se identificar e analisar os estudos na área da CI que abordem as contribuições das antologias na indexação automática. Pretende-se (i) identificar os trabalhos científicos, nas bases Library & Information Science Source e Library, Information Science & Technology Abstracts que abordem este assunto, a sua distribuição temporal e geográfica; (ii) identificar e descrever a centralidade da abordagem temática aos dois conceitos (indexação automática e ontologias), e a abordagem metodológica dos respetivos artigos; (iii) identificar as contribuições presentes nos artigos que constituem o corpus relativamente às potencialidades da utilização conjunta dos dois conceitos. Partiu-se de um estudo exploratório assente numa revisão sistemática da literatura. Os resultados apontam como contribuições das ontologias na indexação automática: (i) desambiguar termos homógrafos e polissémicos; (ii) maior capacidade de integração de relações semânticas de forma automatizada; (iii) uma navegação e expansão de consultas através de relações semânticas; (iv) uma recuperação mais precisa e exaustiva da informação. Conclui-se que o desenvolvimento de sistemas que utilizam o potencial das ontologias na indexação automática procura contornar a sua falta de capacidade semântica. Apesar dos resultados promissores nesta matéria, infere-se que é ainda prematuro e inadequado falar de uma indexação semântica efetiva.

Palavras-chave: Indexação automática. Ontologias. Recuperação da informação. Indexação semântica.

Automatic indexing and ontologies: Identification of convergent contributions in Information Science

ABSTRACT

Automatic and traditional indexing has always been a concern of the area of Information Science (IS). The lack of consistency in human indexing, and the lack of semantics in the automatic are regarded as major disadvantages. One possibility of providing computational systems with greater inference power is the use of ontologies. This proposal aims to identify and analyse studies in the area of IS that address the contributions of ontologies in automatic indexing. It is intended to (i) identify the scientific works, in the scientific journal databases of Library & Information Science, and Information Science & Technology Abstracts that approach this subject, its temporal and geographical distribution; (ii) identify and describe the centrality of the subject approach to the two concepts (automatic indexing and ontologies), and the methodological approach of the respective articles; (iii) identify the contributions present in the articles that make up the corpus regarding the potentialities of the joint use of the two concepts. It was based on an exploratory study based on a systematic review of the literature. The results show the contributions of the ontologies in the automatic indexing, such as: (i) disambiguation of homographs and ambiguous terms; (ii) greater ability to integrate semantic relationships in an automated way; (iii) a navigation and expansion of queries through semantic relationships; (iv) a more accurate and exhaustive retrieval of information. We conclude that the development of systems that use ontologies in its full potential in automatic indexing seek to overcome their lack of semantic capacity. Despite the promising results in this regard, it may be still a little early to speak of effective semantic indexing.

Keywords: Automatic indexing. Ontologies. Information retrieval. Semantic indexing.

Indización automática y ontologías: Identificación de las contribuciones convergentes en la Ciencia de la Información

RESUMEN

La indización, humana y automática, ha constituido desde siempre una preocupación para el área de la Ciencia de la Información (CI). La falta de consistencia en la indización humana y de semántica en la automática, son apuntadas como las principales desventajas. Una posibilidad que permita dotar a los sistemas computacionales de una mayor capacidad de inferencia automática reside en la utilización de ontologías. Esta propuesta tiene como objetivo identificar y analizar los estudios del área de la CI que abordan las contribuciones de las ontologías en la indización automática. Se pretende (i) identificar los trabajos científicos en las bases de datos Library & Information Science Source y Library, Information Science & Technology Abstracts que abordan este tema, y su distribución temporal y geográfica; (ii) identificar y describir la centralidad en la aproximación temática a ambos conceptos (indización automática y ontologías), y el tratamiento metodológico de los artículos en cuestión; (iii) identificar las contribuciones de los artículos que constituyen el corpus existente sobre las potencialidades de utilización conjunta de los dos conceptos. Se ha partido de un estudio exploratorio asentado en una revisión sistemática de la literatura. Los resultados apuntan como contribuciones de las ontologías a la indización automática: (i) desambiguar términos homógrafos y polisémicos, (ii) mayor capacidad de integración de las relaciones semánticas de forma automatizada; (iii) una navegación y expansión de consultas a través de relaciones semánticas; (iv) una recuperación más precisa y exhaustiva de la información. Se concluye que el desarrollo de sistemas que utilizan el potencial de las ontologías en la indización automática contribuye a compensar su falta de capacidad semántica. Pese a los resultados prometedores en esta materia se infiere todavía prematuro e inapropiado hablar de una indización semántica eficiente.

Palabras clave: indización automática. Ontologías. Recuperación de la información. Indización semántica.

INTRODUÇÃO

Os sistemas de informação e de comunicação cobrem virtualmente todas as atividades sociais de tal modo que não mais podemos conceber a sociedade sem suas dependências das tecnologias de informação. Acompanhando o desenvolvimento dessas tecnologias, os repositórios de informações que são produzidos durante o desempenho das inúmeras atividades humanas vêm migrando para o ambiente on-line, de forma que, parafraseando (SHERA; CLEVELAND, 1977), “os registros da aventura intelectual humana” estejam cada vez mais em formatos digitais, acessíveis através de redes e sistemas de computadores. Nas palavras de Foskett (1997, p. 3), “as necessidades humanas de informação estão crescendo, na medida em que crescem as dependências de informação da sociedade, para sobreviver e florescer”. As redes e tecnologias digitais promovem o que vem sendo chamado de “tempos exponenciais”, sendo que os números que refletem a produção atual de registros de informação manifestam essa curva de crescimento. Dados estimados sobre a produção mundial de conteúdo digital (EMC, 2017) apontam para a cifra de 44 Zettabytes gerados no ano de 2020, ou seja, mais de 5 Terabytes para cada ser humano vivo. Comparados com os cerca de 50 milhões de minutos da vida de uma pessoa longeva, já temos hoje ordens de magnitude de conteúdo disponível em cada área do conhecimento a mais do que um ser humano poderia consumir, se decidisse dedicar toda a sua vida para tal, o que aumenta a dependência humana em relação aos sistemas de indexação e recuperação de informação (RI).

A indexação, operação que consiste em descrever e identificar o conteúdo de um documento através de termos foi desde sempre uma preocupação da área da ciência da informação (CI). Seja na sua vertente manual, em que todo o processo de indexação é controlado por humanos, seja, um pouco mais recentemente, em modo automático, onde são artefatos computacionais programados os responsáveis por todas as etapas e decisões operacionais.

Existem ainda sistemas chamados de semiautomáticos, nos quais humanos e máquinas são responsáveis por diferentes etapas do processo cabendo, geralmente, aos primeiros a decisão final na atribuição dos termos. Tanto a abordagem manual quanto a automática apresentam prós e contras, por vezes em contraponto, como questões relacionadas com consistência e semântica, sendo a primeira uma desvantagem da indexação manual e a segunda da automática. Entretanto, o fenômeno dos *Big Data*, os grandes acervos digitais e a Web, em geral, tem ensejado maior desenvolvimento da indexação automática. Esta, que teve sua gênese com a disponibilidade de textos eletrônicos na década de 1950 (BAXENDALE, 1958; LUHN, 1957; MARON, 1961) é parte de, virtualmente, todos os processos de gestão de informação digital na Web, além de processos razoavelmente novos, como a análise de sentimentos (HU; LI, 2011) e artefatos baseados em inteligência artificial e machine learning (SEBASTIANI, 2002). Uma revisão histórica do assunto pode ser encontrada em Stevens (1965), Sparck Jones (1974) e Lancaster (2003).

Embora tenhamos evoluído sobremaneira nas duas últimas décadas, os sistemas de RI (SRIs) atuais ainda deixam a desejar no tocante às buscas baseadas em temas ou conceitos, com demanda por altas taxas de revocação e precisão. Diferenças nos vocabulários e terminologias específicas de nichos e domínios; documentos em linguagens múltiplas; fenômenos como a polissemia e sinonímia dificultam a indexação, e conseqüentemente, a posterior recuperação dos documentos e outros itens de informação.

Uma possibilidade de dotar sistemas computacionais com maior capacidade de inferência reside na utilização de sistemas de organização do conhecimento (SOCs/KOS), tais como linguagens de indexação - tesouros, listas de cabeçalhos de assunto ou esquemas de classificação bibliográfica - ou esquemas formais, como as ontologias. Quando os termos de indexação são controlados pelos SOCs, temos diversos benefícios, em comparação com a indexação por termos livres que ocorre nos

motores de busca publicamente disponíveis, como Google, Bing; entre outros. No caso das ontologias, podemos contar com a riqueza de um conjunto de relações semanticamente ricas, com a facilidade de acoplamento destas ontologias a sistemas de recuperação de informações e a naturalidade da representação destes artefatos nas linguagens da Web Semântica.

Guarino (1998), partindo de uma definição de Gruber (1992, p. 199), descreve uma ontologia como sendo um artefato de engenharia, constituído por um vocabulário intencional utilizado para descrever certa realidade, em conjunto com pressupostos explícitos em uma teoria lógica representando conceitos e relações de maneira clara e não ambígua. Apesar da falta de consenso entre alguns autores em relação à classificação de tipos de ontologias, Guarino (1998) classifica ontologias em quatro categorias com base em seu conteúdo: de alto nível, de domínio, de tarefas e de aplicação. As ontologias de alto nível descrevem conceitos bastante gerais, tais como, espaço, tempo, matéria, objeto, etc., que são independentes de um domínio particular. As ontologias de domínio descrevem um vocabulário relacionado a um domínio genérico, tal como medicina. As ontologias de tarefas também expressam um vocabulário inerente a um domínio genérico ou uma tarefa genérica, como vendas. Ambos os tipos são especializações de termos advindos da ontologia de alto nível. Percebe-se que a definição de ontologia de domínio e de tarefa se sobrepõe, estabelecendo certa confusão em relação ao termo a ser utilizado. E, finalmente, as ontologias de aplicação descrevem conceitos dependentes de domínio e de tarefa particulares, que geralmente são especializações dos termos advindos das ontologias de domínio e de tarefa (SILVA; SOUZA, 2013). As ontologias de domínio são o tipo mais comumente desenvolvido, sendo que diversos trabalhos podem ser encontrados na literatura, como na área de química (GÓMEZ-PÉREZ; FERNÁNDEZ; VICENTE, 1996), modelagem de negócio (GRÖNINGER; FOX, 1995), bioinformática (STEVENS et al., 2004), modelagem de processos de *software* (FALBO, 1998), entre outros.

Ontologias são, atualmente, objetos de estudo tanto da ciência da computação (CC) como da ciência da informação (CI). Desta proposição surge a motivação de investigar o modo como a área da CI tem abordado a potencialidade da relação das ontologias com a indexação automática. Assim, propõe-se identificar e analisar os estudos na área da CI que abordem as contribuições das ontologias para a indexação automática. Especificamente pretende-se: (i) identificar os trabalhos científicos, com revisão por pares, nas bases específicas da área Library & Information Science Source (LISS) e Library, Information Science & Technology Abstracts (LISA) que abordem este assunto assim como a sua distribuição temporal e geográfica; (ii) identificar e descrever a centralidade da abordagem temática aos dois conceitos operacionais (indexação automática e ontologias), assim como a abordagem metodológica aplicada nos respetivos artigos; (iii) identificar as contribuições presentes nos artigos que constituem o *corpus* de estudo relativamente às potencialidades da utilização conjunta dos dois conceitos.

Para dar cumprimento aos objetivos enunciados, partiu-se de um estudo exploratório assente numa revisão sistemática da literatura baseada num *corpus* de estudo de 25 artigos.

METODOLOGIA

Para a seleção e análise do *corpus* de estudo recorreu-se a um conjunto de técnicas e procedimentos metodológicos inseridos no espectro da análise qualitativa nomeadamente a análise de conteúdo e a análise de palavras-chave em contexto (BARDIN, 2011; BERNARD; RYAN, 2010; COUTINHO, 2013; GIL, 2008). Importa referir que, indo ao encontro do expresso por Coutinho (2013, p. 35–36), adotou-se uma “estratégia flexível” adaptada não só ao problema a analisar como também ao desenvolvimento da investigação.

RECOLHA E SELEÇÃO DO CORPUS DE ESTUDO

O *corpus* de estudo foi extraído da pesquisa efetuada em 14 de junho de 2017 no portal EBSCO *Discovery Service*, do qual foram selecionadas duas bases: LISS e LISTA. Utilizou-se uma expressão de pesquisa abrangente (*index* AND ontolog**), restringindo-a aos campos *título* e *assuntos*. Tais opções decorreram de testes prévios efetuados que tiveram como objetivo, no caso da expressão usada, a obtenção de maior revocação e, relativamente aos campos de pesquisa, a intenção de refinamento em termos de precisão, para reduzir, deste modo, o nível de revocação obtido pela expressão de pesquisa. Uma vez que a verificação da distribuição temporal dos artigos faz parte dos objetivos, não se restringiu a pesquisa a um período específico.

Da aplicação desses critérios metodológicos foram recuperados 65 trabalhos (LISS – 34 e LISTA – 31). Deste número foram excluídos os registos duplicados, restando 39. Da pré-análise efetuada aos 39 trabalhos recuperados, verificou-se que um destes era a apresentação de um número da publicação periódica *Scire*, pelo que foi descartado, restando assim 38 artigos.

Numa primeira leitura abrangente entendeu-se que os 38 artigos não responderiam ao primeiro objetivo (identificar os estudos nas bases LISS e LISTA que abordassem o assunto em questão), devido ao facto de a expressão de pesquisa ter sido deliberadamente formulada para se atingir maior revocação, o que concorreria, inevitavelmente, para menor grau de precisão; e ainda pela circunstância do conteúdo do campo assuntos, da responsabilidade das respetivas bases pesquisadas, ter mostrado, em vários casos, características de exaustividade, aumentando também, a revocação em detrimento da precisão. Assim, houve a necessidade de se aferir o nível de abordagem temática relativamente aos conceitos operacionais considerados (indexação automática e ontologias). Neste sentido foram estabelecidas quatro variáveis ordinais, refletindo a centralidade da abordagem expressa numa escala de intensidade negativa (ver tabela 1), (BARDIN, 2011, p. 141; GIL, 2008, p. 84).

Tabela 1 – Variáveis usadas na categoria de análise *abordagem temática*

Abordagem do artigo aos conceitos indexação automática e ontologias	Nível de aproximação	Valor ¹
O conceito é parte central do estudo.	Central	0
O conceito é abordado devido à sua relação intrínseca com o objeto de estudo.	Inerente	-0,5
O conceito é abordado devido a uma relação secundária com o objeto de estudo.	Periférico	-1,5
O conceito não é abordado, mas infere-se uma ligação temática.	Inferido	-3

Os conceitos foram analisados de forma independente, uma vez que alguns trabalhos apresentam diferentes abordagens temáticas relativamente a cada um deles. Como critérios de seleção dos trabalhos, para que eles fossem considerados relevantes, considerou-se que não poderiam abordar qualquer um dos conceitos de forma *inferida* nem os dois de forma *periférica*. Atendendo aos valores de cada menção, a soma deles não poderia ser igual ou inferior a -3; todas as restantes somas foram consideradas para efeito de seleção.

A classificação foi efetuada através da análise dos resumos, de onde se extraiu o foco do trabalho. A menção atribuída foi posteriormente validada pela análise do contexto, no qual as expressões “ontolog” e “index” surgem em cada artigo, assim como pela sua frequência e coocorrência (BARDIN, 2011, p. 138–144; COUTINHO, 2013, p. 219–220).

Após esse procedimento chegou-se ao *corpus* de estudo, que contabilizou 25 artigos (cf. tabela 2, ver apêndice A para consulta da lista de referências bibliográficas completa).

¹ O aumento progressivo dos intervalos, além de espelhar o afastamento temático de cada tipo de abordagem, permitiu estabelecer regras simples de exclusão dos artigos (no caso, valor da soma ser menor ou igual a -3).

Tabela 2 – Corpus do estudo

Ref.	Autoria	Ano	Título
01	Gilchrist, A.	2003	Thesauri, taxonomies and ontologies - an etymological note
02	Kabel, S.; De Hoog, R.; Wielinga, B.J.; Anjewierden, A.	2004	The added value of task and ontology-based markup for information retrieval
03	Bayer, O.; Höhfeld, S.; Josbächer, F.; Kimm, N.; Kradepohl, I.; Kwiatkowski, M.; Puschmann, C.; Sabbagh, M.; Werner, N.; Vollmer, U.	2005	Evaluation of an Ontology-based Knowledge-Management-System. A Case Study of Convera RetrievalWare 8.0
04	Ciravegna, F.; Petrelli, D.	2006	Annotating document content: a knowledge-management perspective
05	Köhler, J.; Philippi, S.; Specht, M.; Rwegg, A.	2006	Ontology based text indexing and querying for the semantic web
06	Tsinarakis, C.; Polydoros, P.; Christodoulakis, S.	2007	Interoperability Support between MPEG-7/21 and OWL in DS-MIRF
07	Hernandez, N.; Mothe, J.; Chrisment, C.; Egret, D.	2007	Modeling context through domain ontologies
08	Pirró, G.; Talia, D.	2008	LOM: a linguistic ontology matcher based on information retrieval
09	Allampalli-Nagaraj, G.; Bichindaritz, I.	2009	Automatic semantic indexing of medical images using a web ontology language for case-based image retrieval
10	Moura, M.A.	2009	Informação, ferramentas ontológicas e redes sociais Ad Hoc: A interoperabilidade na construção de tesauros e ontologias
11	Good, B.M.; Tennis, J.T.	2009	Term based comparison metrics for controlled and uncontrolled indexing languages
12	Solskinnsbakk, G.; Gulla, J.A.	2010	Combining ontological profiles with context in information retrieval
13	Bouramoul, A.	2011	The Semantic Dimension in Information Retrieval, from Document Indexing to Query Reformulation
14	Kara, S.; Alan, Ö.; Sabuncu, O.; Akpınar, S.; Cicekli, N.K.; Alpaslan, F.N.	2012	An ontology-based retrieval system using semantic indexing
15	De Maio, C.; Fenza, G.; Loia, V.; Senatore, S.	2012	Hierarchical web resources retrieval by exploiting Fuzzy Formal Concept Analysis
16	Chiaravalloti, M.T.; Pasceri, E.; Taverniti, M.	2012	URT "Indexing and Classification Systems" Projects and Biomedical Knowledge Standards
17	Willis, C.; Losee, R.M.	2013	A Random Walk on an Ontology: Using Thesaurus Structure for Automatic Subject Indexing
18	Qiu, J.; Lou, W.	2014	Constructing an information science resource ontology based on the Chinese Social Science Citation Index
19	Thenmalar, S.; Geetha, T.V.	2014	Enhanced ontology-based indexing and searching
20	Bendib, I.; Laouar, M.R.; Hacken, R.; Miles, M.	2014	Semantic ontologies for multimedia indexing (SOMI): Application in the e-library domain
21	Gödert, W.	2014	Ein Ontologie-basiertes Modell für Indexierung und Retrieval
22	Pang, C.; Hendriksen, D.; Dijkstra, M.; Van Der Velde, K.J.; Kuiper, J.; Hillege, H.L.; Swert, M.A.	2015	BiobankConnect: software to rapidly connect data elements for pooled analysis across biobanks using ontological and lexical indexing
23	Du Preez, M.	2015	Taxonomies, folksonomies, ontologies: what are they and how do they support information retrieval?
24	Vlachidis, A.H.; Tudhope, D.	2016	A Knowledge-Based Approach to Information Extraction for Semantic Interoperability in the Archaeology Domain
25	Gödert, W.	2016	An Ontology-Based Model for Indexing and Retrieval

PROCEDIMENTOS METODOLÓGICOS

A seleção do *corpus* de estudo concorreu para o cumprimento parcial do primeiro e segundo objetivos, na medida em que foi determinado tanto o número de artigos que focam o objeto de estudo quanto o nível de centralidade da abordagem temática aos conceitos operacionais considerados.

Para dar pleno cumprimento ao primeiro objetivo, no que respeita à determinação geográfica e cronológica dos estudos foram considerados os países associados às afiliações dos autores e o ano da publicação dos artigos.

Para completar o cumprimento do segundo objetivo, procedeu-se à determinação da abordagem metodológica através da análise das introduções e resumos dos artigos, privilegiando-se a informação relativa à descrição dos procedimentos metodológicos. Adotou-se, nesta categoria de análise, três variáveis nominais (ver tabela 3), (BERNARD; RYAN, 2010, p. 151; GIL, 2008, p. 82).

Tabela 3 – Variáveis usadas na categoria de análise abordagem metodológica

Abordagem metodológica do artigo (descrição)	Classes (tipo de abordagem)
Os autores analisam e discutem a relevância da utilização convergente de ontologias e da indexação automática para o avanço, consolidação, ou afirmação da CI enquanto área do conhecimento e do saber-fazer.	Análise epistemológica
Os autores analisam e discutem um sistema de indexação automática (ou modelo, enquanto representação conceptual de um sistema), seja real ou apresentado como proposta, com integração de ontologias.	Análise de sistema
Os autores apresentam os resultados operacionais de um sistema, em atuação, sobre um estudo de caso bem definido, com medidas de impacto da aplicação de ontologias nos processos de indexação automática.	Estudo experimental

Ainda que não seja incomum, relativamente à abordagem metodológica, que os trabalhos analisados pudessem, em teoria, ser incluídos em mais de uma classe estabelecida, tal situação não foi observada pelo facto de a classificação adotada ter por principal propósito valorizar a contribuição mais significativa de cada um dos textos identificados na recolha. Nesta perspectiva, a opção por três grandes abordagens metodológicas confere às classes um carácter de exclusão mútua e contribui para a homogeneidade, pertinência, fidelidade e produtividade de acordo com as qualidades expressas por Bardin (2011, p. 149–150) para a categorização.

Para dar cumprimento ao terceiro objetivo, identificar as contribuições relativamente às potencialidades da utilização conjunta dos dois processos, foram analisadas as secções dos artigos relativas às conclusões ou considerações finais. Nos casos em que as conclusões não apresentavam informações robustas para o propósito pretendido, recorreu-se aos resumos dos respectivos artigos, a fim de identificar as contribuições relativamente às potencialidades da utilização das ontologias para o processo de indexação automática. Ao contrário das duas etapas anteriores (análise das abordagens temática e metodológica), das quais se partiu de classes determinadas *a priori*, se bem que posteriormente ajustadas de acordo com os preceitos da análise categorial (BARDIN, 2011, p. 148–149; GIL, 2008, p. 176–177), nesta fase, o carácter exploratório da análise determinou um diferente procedimento, sendo o sistema classificativo resultante da mesma (BARDIN, 2011, p. 149; COUTINHO, 2013, p. 217). Neste sentido foram consideradas quatro classes: precisão e exaustividade; desambiguação; indexação semântica e navegação orientada (ver tabela 4).

Tabela 4 – Variáveis usadas na categoria de análise contribuições

Potencialidades da utilização conjunta dos dois processos	Classes (contribuições)
Desambiguação de termos homógrafos e polissêmicos.	Desambiguação
Maior capacidade de integração de relações semânticas automatizadas.	Indexação semântica
Disponibilidade de pesquisar via relações semânticas existentes nas ontologias.	Navegação orientada
Recuperação mais precisa e exaustiva da informação.	Precisão e exaustividade

Importa referir que nesta categoria de análise cada artigo poderá estar representado em mais de uma classe, uma vez que as unidades de referência são as contribuições neles identificadas usando um recorte semântico na sua limitação (BARDIN, 2011, p. 135).

RESULTADOS E DISCUSSÃO

Os 25 trabalhos identificados nas bases LISS e LISTA, que abordam a indexação automática e as ontologias (ver tabela 2), encontram-se distribuídos de uma forma assimétrica entre os anos 2003 e 2016 (cf. figura 1). Acentua-se o aumento do número de artigos a partir de 2011, sendo que 25% estão concentrados entre 2014 e 2016, e os restantes 75% encontram-se dispersos nos anos anteriores, como mostra o diagrama da figura 2.

Atendendo à afiliação dos seus autores, esses artigos apresentam uma distribuição geográfica por 16 países (cf. figura 3). A distribuição foi feita com base em dois referentes, artigo e autor. Na contabilização por artigo, o total é de 27, uma vez que se verifica em dois artigos a combinação de dois países (Canadá/EUA e Argélia/EUA) tendo os mesmos sido considerados para a soma dos respetivos países. No que se refere à contabilização dos autores a soma total é de 72. A única repetição ocorre nos autores alemães, em que um dos 14 está associado a dois artigos, ambos sem coautoria.

Após a análise às abordagens temática e metodológica dos artigos (ver apêndice B), verificou-se, em termos gerais, a seguinte situação: por um lado, maior aproximação temática ao conceito *ontologias* do que ao conceito *indexação automática*; por outro lado, no que concerne à metodologia, comparando as três abordagens, observou-se uma percentagem significativamente maior (72%) no que respeita aos *estudos experimentais* (cf. figura 4).

Em relação à abordagem temática, comparando a quantidade de artigos identificados com os níveis de aproximação *central*, *inerente* ou *periférica* (cf. figura 5), em cada conceito (*ontologias* e *indexação automática*), constata-se que o primeiro referido apresenta 60% das atribuições do nível *central*, nível de maior aproximação temática. Ao conceito *indexação automática* foi atribuído 82% das menções correspondentes ao nível de maior afastamento (aproximação *periférica*).

Comparando o resultado das duas categorias analisadas (ver figura 6), destaca-se um conjunto de observações:

- i. a abordagem metodológica *análise epistemológica* não apresenta artigos cujo *grau de aproximação temática* seja zero, (cf. coluna *Grau (soma a+b)* da figura 6);
- ii. a abordagem metodológica *estudo experimental* é a única das três abordagens a apresentar em mais de metade dos artigos aí incluídos (61%) um *grau de aproximação temática* igual ou maior que -0,5 (que representam os valores de maior aproximação temática);
- iii. apenas a abordagem metodológica *análise de sistema* apresenta artigos cuja abordagem temática ao conceito *ontologias* apresenta o nível mais afastado de aproximação (nível *periférico*).

Figura 1 – Distribuição temporal dos artigos constituintes do corpus de análise

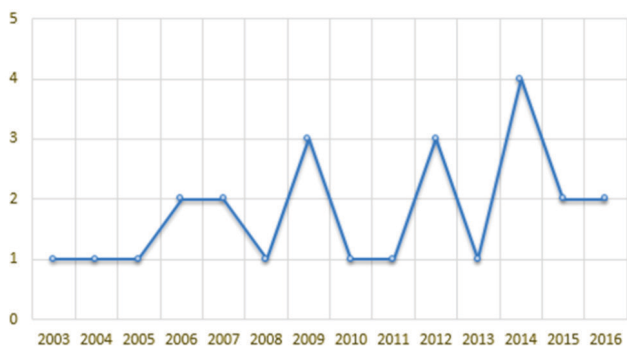


Figura 2 – Diagrama de extremos e quartis da distribuição temporal

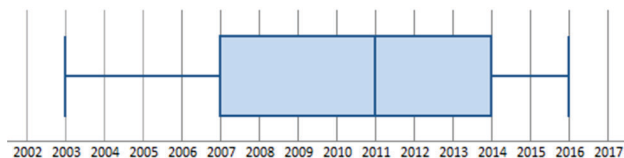


Figura 3 – Distribuição do corpus de análise atendendo ao país associado à afiliação dos autores

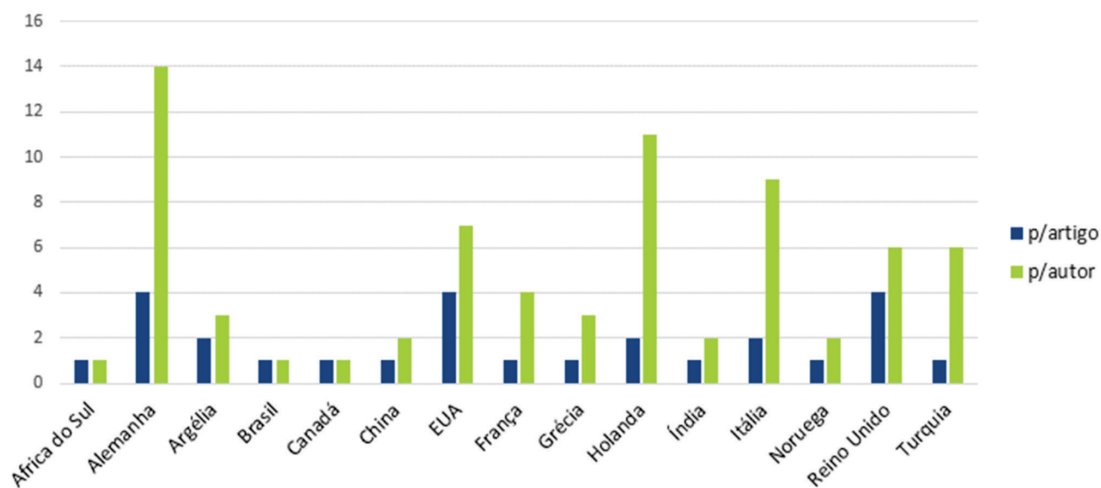


Figura 4 – Distribuição percentual do tipo de abordagem identificada nos 25 artigos

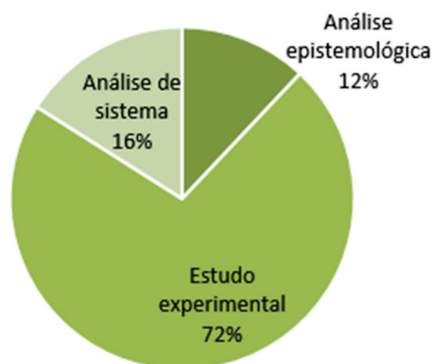


Figura 5 – Distribuição percentual, por nível de aproximação, da abordagem temática aos dois conceitos operacionais (indexação automática e ontologias)

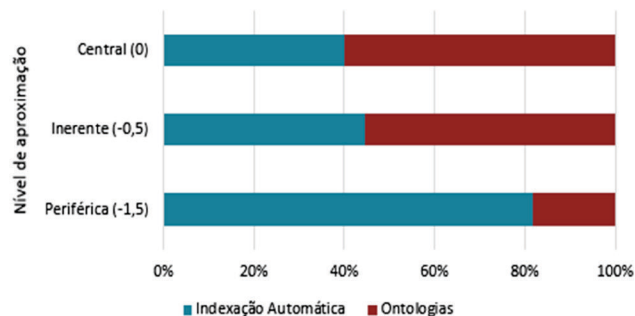
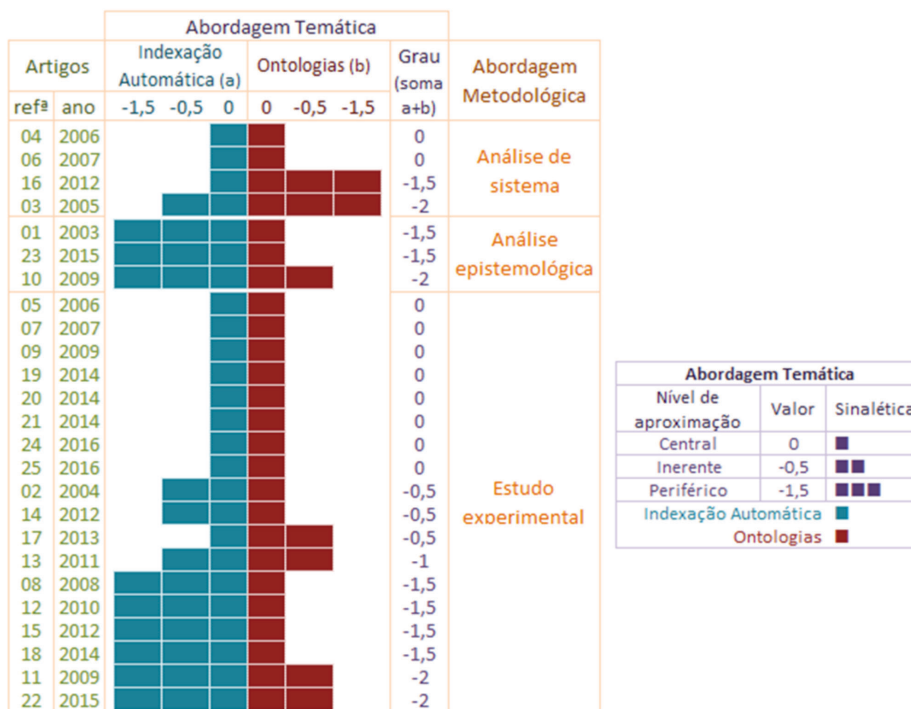


Figura 6 – Comparação entre a abordagem temática e a abordagem metodológica identificadas nos 25 dos artigos



No que concerne à identificação e contabilização das contribuições relativamente às potencialidades da utilização conjunta dos dois conceitos operacionais, a tabela 5 apresenta os respetivos resultados, de onde se destaca o seguinte:

- i. à exceção da contribuição navegação orientada, identificada em apenas dois artigos (8%), as restantes apresentam uma frequência muito aproximada (em cinco (20%) e seis artigos (24%));
- ii. em 48% dos artigos não foram identificadas contribuições, sendo que, destes, apenas 8% (dois de 12) estão associados a maior grau de aproximação temática (-0,5 ou zero);
- iii. não foram encontradas quaisquer contribuições em artigos associados ao menor grau de aproximação temática (-2) e apenas um com o grau imediatamente anterior (-1,5), sendo essa contribuição relativamente à indexação semântica;
- iv. as contribuições identificadas concentram-se em artigos associados a maior grau de aproximação temática (-0,5 ou zero), concretamente em 84% (16 de 19) das contribuições identificadas.

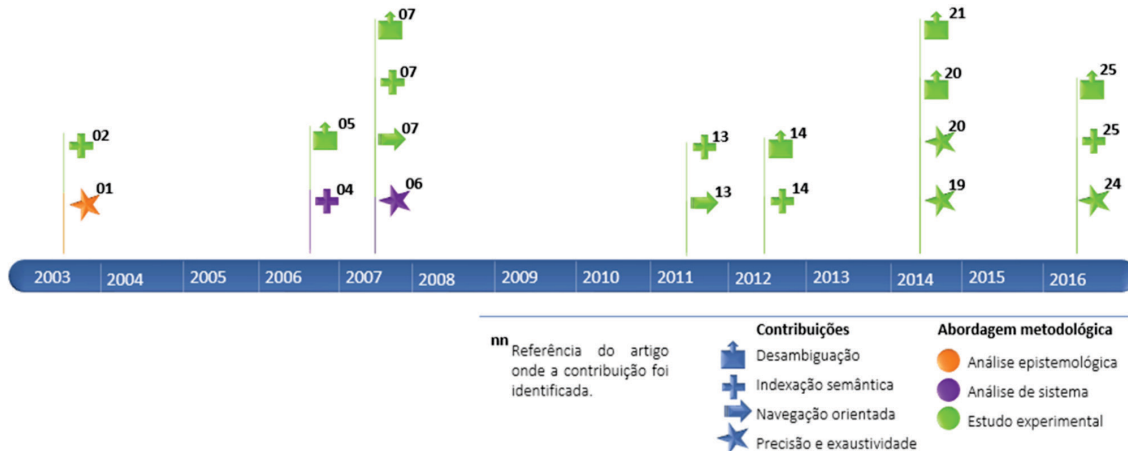
Por fim, a figura 7 apresenta a distribuição temporal das contribuições identificadas, de acordo com o ano de publicação dos artigos de onde foram coligidas. No referido esquema é possível constatar uma dispersão no tempo das várias contribuições, não se verificando, assim, concentração de qualquer tipologia específica.

Por meio da figura 7 é possível observar-se que o número mais elevado de contribuições se verifica nos artigos cuja abordagem metodológica se baseia no *estudo experimental*. Embora em termos absolutos exista uma diferença acentuada entre o número total de artigos com a referida abordagem (18), relativamente ao número de artigos das outras duas abordagens (três para a *análise epistemológica* e quatro para a *análise de sistema*), no que respeita a termos percentuais, considerando o número de artigos com contribuições identificadas, observa-se maior equilíbrio, particularmente nas abordagens iminentemente práticas: *estudo experimental* – 56% (10); *análise de sistema* – 50% (dois). A *análise epistemológica* apresentou o valor de 33% (um artigo).

Tabela 5 – Contribuições identificadas nos 25 artigos e respetiva contabilização

Contribuições (classes)	Quantidade de artigos					Soma	Porcentagem de cada classe relativamente aos 25 artigos
	por <i>Graus de Aproximação Temática</i>						
	0	-0,5	-1	-1,5	-2		
Desambiguação	5	1	0	0	0	6	24%
Indexação semântica	3	1	1	1	0	6	24%
Navegação orientada	1	0	1	0	0	2	8%
Precisão e exaustividade	4	1	0	0	0	5	20%
Sem contribuições identificadas	1	1	0	6	4	12	48%

Figura 7 – Distribuição temporal dos contributos de acordo com o ano dos respectivos artigos onde foram identificados



CONCLUSÕES

Deste estudo exploratório ficou evidente, tanto através da literatura consultada como do trabalho empírico, a importância das ontologias, nomeadamente as de domínio, na indexação automática. Este interesse manifesta-se quer na representação da informação mais consistente, quer, em especial, na recuperação da informação a qual atinge níveis de elevada precisão. Conclui-se que a questão das potencialidades da relação das ontologias com a indexação automática tem vindo, na área da CI, a ser trabalhada de maneira continuada tal como mostra o resultado do estudo empírico, com especial ênfase nos últimos quatro anos que antecedem este estudo.

Infer-se deste resultado um interesse crescente no estudo desta matéria, o qual se apresenta transversal em termos geográficos, embora o grupo de países anglo-saxónicos mostre relativo destaque, indo ao encontro do tradicional interesse do estudo sobre estas matérias, quando comparado com outros países francófonos e/ou de outras culturas idiomáticas.

Os trabalhos analisados, apesar de centrados na CI, mostraram grande proximidade e intercâmbio com a área da ciência da computação, em particular no que se refere à recuperação de informação em sistemas digitais apoiados pela “inteligência artificial”.

Os resultados do estudo empírico fortalecem esta inferência, na medida em que os trabalhos associados a maior centralidade temática, relativa aos dois conceitos (indexação automática e ontologias), apresentam uma abordagem metodológica essencialmente prática baseada em metodologias experimentais e não em abordagens teórico-metodológicas (análise epistemológica).

Com base nesta ideia, infere-se que a natureza operacional dos dois conceitos concorre para tal situação, no sentido em que, por um lado, a indexação é simultaneamente processo e resultado; por outro lado, as ontologias, no contexto estudado, apresentam uma função prática e objetiva de, enquanto artefato computacional, representar e modelar uma área específica do conhecimento humano através de relações semânticas. Neste contexto, a contribuição das ontologias para a indexação automática consubstancia-se em: i. potenciar a desambiguação de termos homógrafos e polissémicos; ii. imprimir maior capacidade de integração de relações semânticas de forma automatizada; iii. possibilitar a navegação e expansão de consultas através de relações semânticas; iv. concorrer para uma recuperação mais precisa e exaustiva da informação.

Infere-se, também, que o esforço aplicado no desenvolvimento de sistemas que façam uso do evidente potencial da utilização conjunta dos dois conceitos estudados tem por base a tentativa de contornar a falha inerente à indexação automática – a menor capacidade semântica. O conceito de “indexação semântica”, presente em vários artigos analisados, aplicado a processos automatizados de indexação, é um passo nesse sentido. Sistemas que fazem uso de ontologias e técnicas de processamento de linguagem natural (PLN) foram já desenvolvidos e encontram-se em fase experimental, apresentando resultados bastante promissores nesta matéria; infere-se que é ainda prematuro e inadequado falar de uma indexação semântica efetiva.

REFERÊNCIAS

- BARDIN, L. *Análise de conteúdo*. Tradução Luís Antero Reto Augusto Pinheiro. São Paulo: Almedina, 2011.
- BAXENDALE, P. B. Machine-made index for technical literature: an experiment. *IBM Journal of Research and Development*, v. 2, n. 4, p. 354-361, 1958.
- BERNARD, H. R.; RYAN, G. W. *Analysing qualitative data: systematic approaches*. Los Angeles: SAGE, 2010.
- COUTINHO, C. P. *Metodologia de investigação em ciências sociais e humanas: teoria e prática*. 2. ed. Coimbra: Almedina, 2013.
- EMC. *The digital universe of opportunities*. Disponível em: <<https://www.emc.com/infographics/digital-universe-2014.htm>>. Acesso em: 08 ago. 2017.
- FALBO, R. de A. *Integração de conhecimento em um ambiente de desenvolvimento de software*. Rio de Janeiro: Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Universidade Federal do Rio de Janeiro, 1998.
- FOSKETT, D. J. *Thesaurus: readings in information retrieval*. 1997.
- GIL, A. C. *Métodos e técnicas de pesquisa social*. 6. ed. São Paulo: Atlas S.A., 2008.
- GÓMEZ-PÉREZ, A.; FERNÁNDEZ, M.; VICENTE, A. de. *Towards a method to conceptualize domain ontologies*. In: WORKSHOP: ONTOLOGICAL ENGINEERING, 1996. *Electronic proceedings...* 1996. Disponível em: <<http://oa.upm.es/7228/>>. Acesso em: 06 nov. 2017.
- GRUBER, T. R. What is an ontology?. *International Journal Human-Computer Studies*, v. 43, p. 907-928, 1992.
- GRUNINGER, M.; FOX, M. S. Methodology for the design and evaluation of ontologies. In: INTERNATIONAL JOINT CONFERENCE ON AI, 1995, Montreal. *Electronic proceedings...* Montreal: [s.n.], 1995. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.44.8723>>. Acesso em: 06 nov. 2007.
- GUARINO, N. Formal ontology and information systems. In: INTERNATIONAL CONFERENCE (FOIS'98), 1., 1998. *Proceedings...* Amsterdam: IOS Press, 1998. P. 3-15.
- HU, Y.; LI, W. Document sentiment classification by exploring description model of topical terms. *Computer Speech & Language*, v. 25, n. 2, p. 386-403, 2011.
- LANCASTER, F. W. Do indexing and abstracting have a future?. *Anales de Documentación*, 2003.
- LUHN, H. P. A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research and Development*, v. 1, n. 4, p. 309-317, 1957.
- MARON, M. E. Automatic indexing: an experimental inquiry. *Journal of the ACM (JACM)*, v. 8, n. 3, p. 404-417, 1961.

SEBASTIANI, F. Machine learning in automated text categorization. *ACM Computing Surveys (CSUR)*, v. 34, n. 1, p. 1-47, 2002.

SHERA, J. H.; CLEVELAND, D. B. History and foundations of information-science. *Annual Review of Information Science and Technology*, v. 12, p. 249-275, 1977.

SILVA, D. L. DA; SOUZA, R. R. Representação de documentos multimídia: dos metadados às anotações semânticas. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB) 14., 2013, Florianópolis. *Anais eletrônicos...* 2013. Disponível em: <<http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/view/430/427>>

SPARCK JONES, K. *Automatic indexing 1974: a state of the art review*. [S.l.]: Computer Laboratory, University of Cambridge, 1974.

STEVENS, E. M.; URBAN, G. H. *Automatic indexing using cited titles*. In: STATISTICAL ASSOCIATION METHODS FOR MECHANIZED DOCUMENTATION SYMPOSIUM, 1965. *Proceedings...* 1965.

STEVENS, R. et al. Ontologies in bioinformatics. In: STAAB, S.; STUDER, R. (Ed.). *Handbook on ontologies*. Berlin: Springer-Verlag, 2004. P. 635-658.

APÊNDICE A

REFERÊNCIAS DO CORPUS DE ESTUDO

- ALLAMPALLI-NAGARAJ, G.; BICHINDARITZ, I. Automatic semantic indexing of medical images using a web ontology language for case-based image retrieval. *Engineering Applications of Artificial Intelligence*, v. 22, n. 1, p. 18-25, Feb. 2009. DOI: 10.1016/j.engappai.2008.04.018.
- BAYER, O. et al. Evaluation of an ontology-based knowledge-management-system: a case study of convera retrievalware 8.0. *Information Services & Use*, July 2005. Disponível em: <<http://content.iospress.com/articles/information-services-and-use/isu471>>. Acesso em: 06 nov. 2017.
- BENDIB, I. et al. Semantic Ontologies for Multimedia Indexing (SOMI): application in the e-library domain. *Library Hi Tech*, v. 32, n. 2, p. 206-218, Apr. 2014. DOI: 10.1108/LHT-08-2013-0108.
- BOURAMOUL, A. The semantic dimension in information retrieval, from document indexing to query reformulation. *Knowledge Organization*, v. 38, n. 5, p. 425-437, Nov. 2011.
- CHIARAVALLLOTI, M. T.; PASCERI, E.; TAVERNITI, M. URT "Indexing and Classification Systems" Projects and Biomedical Knowledge Standards. *Knowledge Organization*, Jan. 2012.
- CIRAVEGNA, F.; PETRELLI, D. Annotating document content: a knowledge-management perspective. *Indexer*, Apr. 2006.
- DE MAIO, C. et al. Hierarchical web resources retrieval by exploiting fuzzy formal concept analysis. *Information Processing & Management*, v. 48, n. 3, p. 399-418, May 2012. DOI: 10.1016/j.ipm.2011.04.003.
- DU PREEZ, M. Taxonomies, folksonomies, ontologies: what are they and how do they support information retrieval?. *Indexer*, Mar. 2015.
- GILCHRIST, A. Thesauri, taxonomies and ontologies: an etymological note. *Journal of Documentation*, v. 59, n. 1, p. 7-18, Feb. 2003. DOI: 10.1108/00220410310457984.
- GÖDERT, W. Ein ontologie-basiertes modell für indexierung und retrieval. *Information -- Wissenschaft und Praxis*, Apr. 2014. Disponível em: <<http://10.0.5.235/iwp-2014-0017>>. Acesso em: 06 nov. 2017.
- _____. An ontology-based model for indexing and retrieval. *Journal of the Association for Information Science and Technology*, v. 67, n. 3, p. 594-609, Mar. 2016. DOI: 10.1002/asi.23420.
- GOOD, B. M.; TENNIS, J. T. Term based comparison metrics for controlled and uncontrolled indexing languages. *Information Research*, v. 14, n. 1, p. 10, Mar. 2009. Disponível em: <<http://informationr.net/ir/14-1/paper395.html>>. Acesso em: 06 nov. 2017.
- HERNANDEZ, N. et al. Modeling context through domain ontologies. *Information Retrieval*, v. 10, n. 2, p. 143-172, Feb. 2007. DOI: 10.1007/s10791-006-9018-0.
- KABEL, S. et al. The added value of task and ontology-based markup for information retrieval. *Journal of the American Society for Information Science & Technology*, Feb. 2004. DOI: 10.1002/asi.10379. Acesso em: 06 nov. 2017.
- KARA, S. et al. An ontology-based retrieval system using semantic indexing. *Information Systems*, v. 37, n. 4, p. 294-305, June 2012. DOI: 10.1016/j.is.2011.09.004.
- KÖHLER, J. et al. Ontology based text indexing and querying for the semantic web. *Knowledge-Based Systems*, v. 19, n. 8, p. 744-754, Dic. 2006. DOI: 10.1016/j.knsys.2006.04.015.
- MOURA, M. A. Informação, ferramentas ontológicas e redes sociais Ad Hoc: a interoperabilidade na construção de tesouros e ontologias. *Informação & Sociedade*, mar. 2009. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/viewFile/2396/2688>>.
- PANG, C. et al. BiobankConnect: software to rapidly connect data elements for pooled analysis across biobanks using ontological and lexical indexing. *Journal of the American Medical Informatics Association*, v. 22, n. 1, p. 65-75, 1 Jan. 2015. DOI: 10.1136/amiajnl-2013-002577.
- PIRRO, G.; TALIA, D. LOM: a linguistic ontology matcher based on information retrieval. *Journal of Information Science*, v. 34, n. 6, p. 845-860, 8 Dic. 2008. DOI: 10.1177/0165551508091014.
- QIU, J.; LOU, W. Constructing an information science resource ontology based on the Chinese Social Science Citation Index. *Aslib Journal of Information Management*, v. 66, n. 2, p. 202-218, Mar. 2014. DOI: 10.1108/AJIM-10-2013-0114.
- SOLSKINNSBAKK, G.; GULLA, J. A. Combining ontological profiles with context in information retrieval. *Data & Knowledge Engineering*, v. 69, n. 3, p. 251-260, Mar. 2010. DOI: 10.1016/j.datak.2009.10.006.
- THENMALAR, S.; GEETHA, T. V. Enhanced ontology-based indexing and searching. *Aslib Journal of Information Management*, v. 66, n. 6, p. 678-696, Nov. 2014. DOI: 10.1108/AJIM-08-2014-0098.
- TSINARAKI, C.; POLYDOROS, P.; CHRISTODOULAKIS, S. Interoperability support between MPEG-7/21 and OWL in DS-MIRF. *IEEE Transactions on Knowledge and Data Engineering*, v. 19, n. 2, p. 219-232, Feb. 2007. DOI: 10.1109/TKDE.2007.33.
- VLACHIDIS, A.; TUDHOPE, D. A knowledge-based approach to information extraction for semantic interoperability in the archaeology domain. *Journal of the Association for Information Science and Technology*, v. 67, n. 5, p. 1138-1152, May 2016. DOI: 10.1002/asi.23485.
- WILLIS, C.; LOSEE, R. M. A random walk on an ontology: using thesaurus structure for automatic subject indexing. *Journal of the American Society for Information Science and Technology*, v. 64, n. 7, p. 1330-1344, July 2013. DOI: 10.1002/asi.22853.

APÊNCICE B

ANÁLISE DOS ARTIGOS QUANTO ÀS ABORDAGENS TEMÁTICA E METODOLÓGICA

Artigos	Foco determinado pela análise dos respetivos resumos	Abordagem temática		Abordagem metodológica
		Index.Aut.	Ontologias	
[01] <i>Thesauri, taxonomies ...</i> (2003)	Analisa, com base numa revisão da literatura, os termos: tesauros, taxonomias e ontologias, no contexto da sua utilização nas áreas da CI e da CC, em especial no campo da IA e WS, com vista à sua precisão terminológica.	Periférica	Central	Análise epistemológica
[02] <i>The added value of task ...</i> (2004)	Apresenta uma comparação de diferentes tipologias de indexação, desde palavras-chave extraídas do conteúdo dos documentos até à indexação estruturada assente em ontologias, com base num estudo experimental que simula a tarefa de criação de material didático recorrendo a uma base de dados específica.	Inerência	Central	Estudo experimental
[03] <i>Evaluation of an Ontology ...</i> (2004)	Analisa a avaliação das funções e componentes da aplicação <i>RetrievalWare</i> relativa à indexação e recuperação de informação, considerando os aspetos tecnológicos e modos operacionais, com o foco na criação de bibliotecas e tesauros.	Inerência	Periférica	Análise de sistema
[04] <i>Annotating document ...</i> (2006)	Analisa metodologias de anotação de documentos baseada em ontologia com o fim de tornar o seu conteúdo acessível à pesquisa automática e à indexação mesmo quando não existe indexador profissional.	Central	Central	Análise de sistema
[05] <i>Ontology based text ...</i> (2006)	Apresenta um sistema que recorre a ontologias e a técnicas de processamento de linguagem natural para indexação de textos, cujo objetivo é a desambiguação de conceitos assim como a recuperação de textos com palavras equivalentes através da sua ligação a conceitos existentes em ontologias.	Central	Central	Estudo experimental
[06] <i>Interoperability Support ...</i> (2007)	Analisa o software DS-MIRF para o desenvolvimento de um sistema de organização do conhecimento sobre informação multimídia em termos de: recuperação, filtragem, navegação, extração, segmentação e descrição. Apresenta, ainda uma metodologia para a avaliação e comparação de resultados no que se refere à recuperação de informação.	Central	Central	Análise de sistema
[07] <i>Modeling context through ...</i> (2007)	Apresenta uma proposta de modelo para a recuperação da informação que considere o contexto da procura em dois aspetos: os temas tratados nos documentos onde o utilizador procura e os dados especificados pelo mesmo na sua pesquisa, ambos modelados por meio de ontologias especificamente designadas para cada um dos aspetos considerados.	Central	Central	Estudo experimental
[08] <i>LOM: a linguistic ontology ...</i> (2008)	Apresenta um modelo construído a partir de técnicas linguísticas para modelagem de ontologias para implementar no sistema <i>Lucene</i> , com vista ao melhoramento da sua eficácia em termos de precisão e revocação.	Periférica	Central	Estudo experimental
[09] <i>Automatic semantic indexing ...</i> (2009)	Apresenta a implementação de um sistema para a indexação e recuperação de imagens médicas, a partir da extração automática de características imagéticas de baixo nível (cor, textura, forma, etc.) e da sua comparação com os termos e respetivos conceitos modelados em uma ontologia de domínio relativa à anatomia humana.	Central	Central	Estudo experimental

(Continua)

ANÁLISE DOS ARTIGOS QUANTO ÀS ABORDAGENS TEMÁTICA E METODOLÓGICA

Artigos	Foco determinado pela análise dos respetivos resumos	Abordagem temática		Abordagem metodológica
		Index.Aut.	Ontologias	
[10] <i>Informação, ferramentas ...</i> (2009)	Analisa, com base numa revisão de literatura, os novos instrumentos de representação da informação em ambiente digital, entre os quais o conceito de ferramentas ontológicas e a sua interoperabilidade com as redes sociais ad hoc, assim como a nova configuração da garantia (literária, estrutural e de uso) no processo de construção de linguagens de indexação, nesse mesmo ambiente.	Periférica	Inerência	Análise epistemológica
[11] <i>Term based comparison ...</i> (2009)	Apresenta uma coleção de métricas para descrever e comparar conjuntos de termos em linguagens de indexação controladas e não controladas, e a forma como as mesmas podem ser usadas para caracterizar as referidas linguagens, desde as <i>folksonomias</i> até às ontologias, passando pelos tesauros.	Periférica	Inerência	Estudo experimental
[12] <i>Combining ontological ...</i> (2010)	Apresenta um protótipo de sistema de pesquisa com uso a perfis ontológicos, enquanto extensão semântica de uma ontologia, para refinamento e expansão de pesquisas.	Periférica	Central	Estudo experimental
[13] <i>The Semantic Dimension ...</i> (2011)	Apresenta um sistema de representação semântica do conteúdo de documentos e reformulação automática de pesquisas usando uma ontologia de domínio no sentido de melhorar o desempenho na recuperação de informação.	Inerência	Inerência	Estudo experimental
[14] <i>An ontology-based retrieval system ...</i> (2012)	Apresenta um modelo assente numa ontologia para extração e recuperação de informação aplicado ao domínio específico do futebol, através de uma abordagem baseada na recuperação semântica de palavras-chave	Inerência	Central	Estudo experimental
[15] <i>Hierarchical web resources ...</i> (2012)	Apresenta uma metodologia para a recuperação de informação baseada em ontologia que suporta organização e visualização de dados, e providencia uma interface de navegação intuitiva, através da exploração da lógica difusa da teoria de análise formal de conceitos formais.	Periférica	Central	Estudo experimental
[[16] <i>URT Indexing and ...</i> (2012)	Analisa as atividades desenvolvidas pela URT-CNR, que trata da pesquisa em sistemas de gestão de documentos, indexação, classificação, organização de conhecimento e gestão de conteúdo, em particular o projeto Infraestrutura tecnológica de registo de saúde eletrónica com vista à definição de modelos ontológicos e vocabulários que apoiem o EHR na construção de uma linguagem atualizada da área médica.	Central	Periférica	Análise de sistema
[17] <i>A Random Walk on an ...</i> (2013)	Apresenta uma metodologia para análise e modelagem de processos de indexação baseada num algoritmo ponderado de passeio aleatório, testada no contexto da indexação automática efetuada em 4 coleções de documentos pré-indexados com 4 diferentes tesauros, com o fim de analisar a vantagem da estrutura dos tesauros no processo de indexação.	Central	Inerência	Estudo experimental
[18] <i>Constructing an information ...</i> (2014)	Apresenta a construção de uma ontologia sobre recursos bibliográficos em chinês relativa à Ciência da Informação, explorando um novo método para a construção semiautomática de ontologias	Periférica	Central	Estudo experimental

(Continua)

ANÁLISE DOS ARTIGOS QUANTO ÀS ABORDAGENS TEMÁTICA E METODOLÓGICA (Conclusão)

Artigos	Foco determinado pela análise dos respetivos resumos	Abordagem temática		Abordagem metodológica
		Index.Aut.	Ontologias	
[19] <i>Enhanced ontology-based indexing ...</i> (2014)	Apresenta um modelo baseado na incorporação de informação ontológica, tal como conceitos e relações semânticas, assente na ontologia proposta por Kohler <i>et al.</i> com o propósito de melhorar a indexação, em particular no que se refere à precisão da recuperação da informação.	Central	Central	Estudo experimental
[20] <i>Semantic ontologies ...</i> (2014)	Apresenta um modelo que facilita a automação do processo de indexação para registos multimídia, baseado em ontologias e técnicas de extração de relações "semânticas".	Central	Central	Estudo experimental
[21] <i>Ein Ontologie-basierter ...</i> (2014)	Apresenta um modelo assente numa ontologia que combina os métodos e a experiência do uso das linguagens de indexação tradicionais com as potencialidades da modelagem formal de representação do conhecimento.	Central	Central	Estudo experimental
[22] <i>Biobank Connect ...</i> (2015)	Apresenta um software - <i>BiobankConnect</i> , para combinar semi-automaticamente os dados de diferentes Biobancos, com o fim de acelerar a recolha e a harmonização das bases dos referidos bancos.	Periférica	Inerência	Estudo experimental
[23] <i>Taxonomies, folksonomies ...</i> (2015)	Analisa, com base numa revisão de literatura, vários tipos de vocabulários referindo a sua utilidade para os indexadores.	Periférica	Central	Análise epistemológica
[24] <i>A Knowledge-Based Approach ...</i> (2016)	Apresenta um método para a indexação semântica automática de artigos científicos de arqueologia, usando técnicas empíricas (baseadas em regras e recursos terminológicos) de extração de informação em combinação com sistemas de organização de conhecimento de domínio específico (ontologias de domínio).	Central	Central	Estudo experimental
[25] <i>An Ontology-Based Model ...</i> (2016)	Apresenta um modelo assente em ontologia para indexação e recuperação de informação combinando os métodos e experiência do uso das linguagens de indexação tradicionais com as potencialidades da modelagem formal de representação do conhecimento.	Central	Central	Estudo experimental