

ENSAIO SOBRE ONTOLOGIA APLICADA NA RECUPERAÇÃO DA INFORMAÇÃO PARA A CIÊNCIA DA INFORMAÇÃO

Resumo: Apesar de não existir consenso sobre questões centrais no campo da Ciência da Informação, pelo menos três delas merecem destaque: a recuperação da informação, a relevância da informação recuperada e a interação entre o humano e a tecnologia durante estes processos. O processo de recuperação da informação é validado pela adequação das respostas obtidas face às necessidades dos usuários, seja ele realizado por métodos manuais ou automáticos. Não há como desassociar deste contexto a forma como o domínio do conhecimento está representado e organizado: quanto mais fidedigna a representação em relação à realidade, maior é a capacidade de se obter resultados consistentes. Ao contrário do que pregam as abordagens puramente tecnológicas de orientação estatística, advoga-se aqui que o trabalho intelectual de classificar e organizar ainda é o ponto fundamental na Recuperação da Informação. Neste ensaio, de caráter informativo e admitidamente incompleto, enfatizam-se aspectos inerentes às ontologias que permeiam questões centrais da área e que fundamentam novas possibilidades de pesquisa. Sugere-se que a criação da disciplina da Ontologia Aplicada no currículo da Ciência da Informação pode trazer benefícios e gerar oportunidades para pesquisadores, estudantes e profissionais.

Palavras-chave: Ontologia aplicada; representação do conhecimento; recuperação da informação; ciência da informação.

Jeanne Louize Emygdio
Doutoranda em Gestão e Organização
do Conhecimento pela Universidade
Federal de Minas Gerais (UFMG).
jeanne.emygdio@gmail.com

Maurício Barcellos Almeida
Doutor em Ciência da Informação
pela Universidade Federal de Minas
Gerais (UFMG). Bolsista de
Produtividade do Conselho Nacional
de Pesquisa (CNPQ). Professor
Associado da UFMG.
mba@fastmail.org

Lívia Marangon Duffles Teixeira
Doutora em Gestão e Organização do
Conhecimento pela Universidade
Federal de Minas Gerais (UFMG).
liviamarangon@gmail.com

ESSAY ABOUT APPLIED ONTOLOGY IN INFORMATION RETRIEVAL FOR INFORMATION SCIENCE

Abstract: Despite the lack of consensus on central issues in Information Science, at least three of them deserve to be highlighted: information retrieval, the relevance of retrieved information and the interaction between humans and technology during these processes. The information retrieval process is validated according to the suitability of the replies obtained regarding the user's needs, whether performed by manual or automatic methods. One can not detach this context from the way in which the knowledge domain is represented and organized: the more trustworthy the representation is in relation to reality, the greater the capacity for consistent results. In contrast to which purely technological approaches of statistical orientation recommend, we advocate here that the intellectual work of classifying and organizing is still a fundamental aspect for Information Retrieval. In this essay, which is informative and admittedly incomplete, we emphasize inherent aspects of ontologies that permeate central issues within Information Science and that underly new research possibilities. We suggest that the creation of an Applied Ontology discipline within the Information Science curriculum can bring benefits and generate opportunities for researchers, students, and professionals as well.

Keywords: Applied ontology; Knowledge Representation; Information Retrieval; Information Science.

1 INTRODUÇÃO

Os estudos sobre Recuperação da Informação (RI) têm sido frutíferos em campos científicos que buscam prover soluções de valor efetivo para a sociedade, como a Ciência da

Informação. Em função da complexidade dos problemas enfrentados hoje pela ciência – vide o problema da pandemia de 2020 – observa-se a conexão, nem sempre simples e direta, com teorias e práticas provenientes de outras áreas, como a Inteligência Artificial, a Linguística, a Matemática e a Estatística, para citar algumas.

Tendo em vista o caráter universal de uso da web e a relevância em se prover sentido para os dados que trafegam nas vias digitais, o interesse em estudos sobre ontologias tem sido crescente. Hoje, tais estudos representam um campo profícuo para as pesquisas em Ciência da Informação. O propósito prático destas pesquisas é o de fornecer soluções para a interoperabilidade semântica durante a interação entre humanos, máquinas e híbridos, aspecto fortemente estimulado pelo advento da Web Semântica (BERNERS-LEE; HENDLER, J.; LASSILA, 2001). As ontologias são recursos centrais em uma arquitetura semântica para web.

De forma análoga ao que ocorre com a RI, campos distintos da ciência se interessam pelo tema a começar pela Filosofia, a área de origem; a Ciência da Informação, ainda que de forma implícita; a Ciência da Computação, em particular, a Inteligência Artificial; e a Linguística, ainda que se use ali uma concepção completamente diferente do que é a ontologia. Desta forma, aportes teóricos, abordagens, métodos e técnicas de aplicação caracterizam um acervo multiforme, que faz uso de estratégias combinadas quando o assunto são as ontologias.

A Ciência da Informação reconhece a relevância da pesquisa sobre ontologia e a utilidade dos desenvolvimentos em seu entorno (ALMEIDA, 2013). É possível identificar a presença de princípios ontológicos na literatura da área, muito embora o termo “ontologia” nem sempre seja explicitamente mencionado (VICKERY, 1997). O entendimento comum na área relaciona as ontologias a artefatos que possuem dois propósitos distintos: i) prover compreensão do domínio e permitir classificação de termos, quando vistos como uma teoria informal e; ii) possibilitar a criação de vocabulários controlados e a recuperação de informação, quando vistos como um sistema conceitual informal (ALMEIDA, 2020).

Observa-se assim uma lacuna no que diz respeito ao reconhecimento da ontologia como disciplina específica, capaz de prover teorias formais para orientar a construção de artefatos de representação para fins computacionais. Por teoria formal, compreende-se a adoção de teorias filosóficas – a que se costuma denominar “ontologia como disciplina” (ALMEIDA, 2020) – associadas a teorias lógicas que confirmam: i) precisão semântica às representações e; ii) rigor necessário para sustentar a integração de artefatos de representação em um espectro de usos diverso. Por artefatos formais compreende-se a tradução das representações em linguagens

artificiais, aqui denominadas “ontologias como artefatos” (ALMEIDA, 2020). Um rótulo capaz de abarcar tais aspectos é “Ontologia aplicada” (MUNN; SMITH, 2008).

No contexto até aqui descrito, o objetivo do presente artigo é descrever argumentos que justifiquem uma disciplina “Ontologia aplicada” na Ciência da Informação, visando ampliar o campo teórico e prático da área. De forma genérica, acredita-se que tal empreendimento beneficia a área através do reconhecimento de sua *expertise* como provedora de soluções aos problemas de informação encontrados na sociedade.

O restante do presente artigo encontra-se assim estruturado. Enquanto a Seção 2 traz os aspectos históricos, teóricos e práticos relacionados à RI, a Seção 3 descreve aspectos dos estudos relacionados à ontologia. A Seção 4 desenvolve uma reflexão sobre as contribuições da ontologia aplicada para a RI e a Seção 5 oferece considerações finais.

2 RECUPERAÇÃO DA INFORMAÇÃO: ASPECTOS FUNDAMENTAIS

Cabe destacar que o argumento principal defendido aqui sugere que a RI, em determinados contextos – em particular, os de conhecimento altamente especializado – depende diretamente das formas de representar e organizar a informação e o conhecimento. Nesse sentido, a ontologia se revela um pressuposto importante para a RI, de fato, mais uma alternativa entre as diversas formas de representação e organização possíveis. Na presente seção, discorre-se brevemente sobre a RI, um tipo de pesquisa fundamental na CI, abordando suas origens e seus desafios.

2.1 ORIGENS E CONTEXTO DA RECUPERAÇÃO DA INFORMAÇÃO

Historicamente, atribuem-se à Calvin Mooers (1919-1994) as primeiras menções ao termo “recuperação da informação”, nos anos de 1950, em pesquisas relacionadas à recuperação digital de informações (MOOERS, 1956). O trabalho de Mooers relatou os desafios impostos à RI pela linguagem natural e sua complexa semântica, em um contexto de crescimento gradual do volume de publicações científicas organizadas em coleções (FELIPE, 2020).

Mooers (1951, p. 2) define a RI como “a descoberta e a localização de informações armazenadas para que possam ser utilizadas”, enfatizando que as estratégias para RI se distinguem sob dois propósitos elementares: i) atender necessidades de usuários individuais ou

instituições, enfatizando a organização de itens de informação de alta utilidade; ii) prover a organização completa ou, o mais próximo possível disso, de toda a informação e conhecimento existente no mundo.

O desconhecimento de tais estratégias, continua Mooers, pode resultar no desperdício de recursos onerosos como tempo, recursos de processamento computacional e espaço de armazenamento físico das coleções. Nesse contexto, a RI abrange um escopo de práticas de armazenamento, organização e acesso a recursos de informação que podem ser norteadas tanto por aspectos específicos de domínio quanto por aspectos mais abrangentes, sobrepondo-se às especificidades mencionadas.

Já as definições recentes para RI se fiam na descoberta de itens de informação que satisfaçam a necessidade humana frente a um amplo acervo, de natureza não estruturada. Está presente dentre os problemas que a ciência contemporânea enfrenta, mesmo antes do surgimento das inovações eletrônicas e computacionais (FELIPE, 2020; MANNING; RAGHAVAN, P.; SCHUTZE, 2008).

Dentre os itens sobre os quais aplicam-se as técnicas de RI tem-se: documentos, páginas web, catálogos online, itens de informação estruturados e não estruturados, objetos multimídia, cartas, documentos, jornais, artigos, livros, prontuários médicos, mapas, dentre outros (BAEZA-YATES; RIBEIRO-NETO, 2011; SALTON; MCGILL, 1983). De fato, a pesquisa em RI vem se desenvolvendo desde a década de 1940, e alguns marcos históricos desta evolução são dignos de nota (Quadro 1).

Quadro 1: Síntese histórica da pesquisa em RI

Eventos representativos	Referência
Concepção de modelos de recuperação baseados em artefatos terminológicos	Bernier e Crane (1948)
Primeiro uso do termo	Mooers (1950)
Primeiros modelos para armazenamento e recuperação da informação, baseado em sistemas de indexação para aprimorar aspectos de armazenamento e processamento, onerosos para a época.	Bush (1945)
Primeiros índices <i>Key Word in Context</i> (KWIC), estruturas de alinhamento ordenado de termos chave para permitir consultas a partir deste índice	Luhn (1953); Mooers (1956)

Primeira abordagem estatística para indexação	Luhn (1957)
Primeiro uso do termo Thesaurus como um artefato terminológico específico para a abordagem estatística de busca de informação literária	
Início da pesquisa sistemática em RI – <i>International Conference on Scientific Information</i>	1958
Início dos modelos de linguagem específica para indexação independente de linguagens terminológicas.	Mooers (1960)
Popularização do termo na Inglaterra	Robert A. Fairthorne em 1961
Início da pesquisa em RI por linguagem natural (<i>free-text searching</i>)	1964
Primeiro sistema de RI com indexação automática – SMART	Gerard A. Salton em 1965
Criação do formato MARC pela OCLC ¹ fomenta a integração e a computação em larga escala no âmbito da Biblioteconomia e Ciência da Informação	Final da década de 1960
Pesquisas sobre recuperação da informação em ambientes bilíngues, usando tesouros e mapeamento de termos em ambos os idiomas de uma mesma área	Salton (1968)
Impulso à RI automática, com o conceito de relevância na Ciência da Informação	Saracevic (1970)
Esforços para aprimorar a construção de tesouros	Cleverdon (1970)
Primeiros sistemas de RI interativos – DIALOG e MEDLINE	Década de 1970
Estímulo à adoção de métodos estatísticos na tradução automática de textos, em função da complexidade da linguística computacional.	Década de 1970
Surgimento dos “ <i>online public access catalogues</i> ” (OPACs)	
Surgimento dos princípios de classificação por probabilidades	Robertson (1977)
Primeira <i>Text REtrieval Conference</i> (TREC ²) encoraja pesquisas em RI sobre largas coleções textuais	1992
Adoção de técnicas provenientes das ciências comportamentais e cognitivas	Década de 1990

¹ *Online Computer Library Center*. Disponível em: <https://www.oclc.org/>. Acesso em 11/10/2021.

² Maiores informações em: <https://trec.nist.gov/>. Acesso em: 11/10/2021.

Surgimento dos primeiros sites de busca	1994
Surgimento dos conceitos de revocação e precisão, desenvolvimento de mecanismos de avaliação de sistemas de recuperação da informação	LESK (1996)
Criação do navegador <i>Mosaic</i> disseminou o uso e o conhecimento da web no mundo.	NCSA (1993)
Criação do padrão de metadados Dublin Core fomenta interoperabilidade entre sistemas.	Década de 1990
A IA se envolve com a sumarização automática, técnica que permite a extração de conteúdo para transmitir a ideia principal do texto de forma automatizada.	

Fonte: adaptado de Felipe (2020) e Gonzalez e Lima (2003).

Os propósitos da RI se distinguem basicamente quanto a dois aspectos principais: i) à recuperação de documentos a partir de sistemas de indexação, busca e classificação, nomeados na CI como Sistemas de Organização do Conhecimento (SOCs) (SOERGEL, 1999; SOUZA; TUDHOPE; ALMEIDA, 2012; VICKERY, 1986) e; ii) à recuperação de dados a partir de sistemas de gerenciamento de bancos de dados (CHEN, 1976; CODD, 1970; JACOBSON et al., 1992; JARDINE, 1976), recursos de persistência em sistemas de informação no escopo da Ciência da Computação.

2.2 ABORDAGENS, MÉTODOS E TÉCNICAS ADOTADOS NAS PESQUISAS EM RI

A RI moderna tem uma longa história no século XX. De acordo com Baeza-Yates e Ribeiro-Neto (2011), três gerações de sistemas de RI podem ser identificadas na literatura científica:

- a) *Primeira geração*: caracterizada pela adoção de catálogos de cartões que armazenavam, basicamente, o nome do autor e título do documento;
- b) *Segunda geração*: caracterizada pela ampliação das funcionalidades de busca, possibilitando a pesquisa por assunto, palavras-chave e consultas complexas;
- c) *Terceira geração*: caracterizada por práticas de RI a partir de interfaces gráficas e formulários eletrônicos, com características de hipertexto e arquiteturas de sistemas.

Ao longo das gerações, percebe-se igualmente, uma evolução das estratégias concebidas para RI, organizadas por Gonzalez e Lima (2003) nas seguintes categorias:

- *Soluções universais*: ênfase em uma RI capaz de tratar coleções universais de documentos e possibilitar recuperação de informação de qualquer natureza, oriunda de qualquer fonte;
- *Busca de textos com foco em sintaxe*: utilização de bases de dados de referências bibliográficas para consultas com operadores lógicos, índices invertidos, indexação automática, recuperação *full-text* (década de 1980), redução de palavras ao radical, estatísticas de co-ocorrência de termos, técnicas de probabilidade e busca por proximidade de palavras;
- *Busca de conceitos*: captura de informações dos textos para a construção de índices de forma semiautomática para etiquetagem de termos, normalização semântica e expansão de consultas baseadas em conceitos.

Dentre os desafios em geral distintos na literatura sobre RI, destacam-se (BAEZA-YATES; RIBEIRO-NETO, 2011):

- A indexação de grandes volumes de documentos;
- As lacunas de interoperabilidade técnica e sintática entre as coleções a serem tratadas;
- O tratamento da linguagem natural para a recuperação semântica de informações;
- A relevância dos recursos de informação recuperados;
- A RI sobre Big Data;
- A complexidade das demandas de informação apresentadas pelos usuários;
- A capacidade de especificação das consultas sobre coleções;
- O domínio da sintaxe da linguagem de consulta para formulação adequada à demanda;
- A compreensão sobre a forma de se utilizar uma informação para detectar sua relevância;
- A qualidade da interface de busca.

3 ONTOLOGIA APLICADA: ASPECTOS FUNDAMENTAIS

Cabe novamente reiterar o argumento principal, já mencionado, que sugere o impacto das formas de representar e organizar a informação e o conhecimento na respectiva RI. Nesse sentido, a ontologia é importante ao transportar para os sistemas automatizados do século XXI as técnicas de representação e organização do conhecimento consagradas pela CI ao longo de décadas. Para continuar a compor o argumento, a presente seção discorre brevemente sobre a ontologia, em particular, a ontologia aplicada.

3.1 ORIGENS E CONTEXTO DA ONTOLOGIA APLICADA

Os estudos sobre “ontologia” originaram-se na Filosofia, a partir do trabalho do filósofo grego Aristóteles (384 – 322), o qual era voltado à compreensão da existência em seus aspectos primordiais, estruturais e relacionais. Aristóteles acreditava que a realidade, em sua completude, poderia ser representada em um único sistema de categorias.

Os filósofos posteriores, a partir de críticas ao sistema aristotélico, conceberam novos sistemas de categorias. Na época da Revolução Científica, talvez entre 1550 e 1770, os trabalhos realizados pelo filósofo francês *René Descartes* (1596 – 1650) e pelo filósofo alemão *Immanuel Kant* (1724 – 1804) resultaram no deslocamento do interesse central da metafísica para a epistemologia (SMITH; KLAGGES, 2008).

No século XX, após um longo período de esquecimento, surgiu um novo interesse nas teorias de Aristóteles a partir dos anos de 1970, do ponto de vista teórico. Cabe citar a necessidade de noções aristotélicas no âmbito de novas teorias lógicas da época – em particular, a Lógica Modal – desenvolvida por *Saul Kripke* (1940 –). Kripke reviveu e estendeu as noções aristotélicas necessárias à sua própria teoria e acabou provendo respostas e argumentos aos neo-aristotélicos, naquele momento, atacados por exemplo, pelo filósofo norte-americano W. V. Quine e seus colegas.

Além disso, contribuíram para o renascimento da teoria aristotélica, a partir dos anos 1980, os desafios lançados contra a teoria da causalidade que questionavam as teorias devidas ainda a David Hume (1711 – 1776). Ainda, os debates na Filosofia da Ciência geraram ideias, com destaque para as teorias de Nancy Cartwright (1957 –) cujos resultados retratavam cenário tipicamente descrito na teoria de Aristóteles. Todas essas razões, dentre outras, levam filósofos contemporâneos a afirmar que categorias aristotélicas não são apenas produtos da imaginação

daquele filósofo, mas sim reais manifestações da forma pela qual os seres humanos acessam o mundo (ALMEIDA, 2020).

Do ponto de vista prático, a ontologia aristotélica se adequa muito bem a uma forma de estruturar o mundo passível de ser incorporada a computadores. Filósofos contemporâneos com Jonathan Lowe, Barry Smith, Ingvar Johansson, Ludger Jassen, dentre muitos outros,³ em um esforço colaborativo com grupos e entidades científicas ao redor do mundo, delinearam a “Ontologia Aplicada”: uma conjunção entre Filosofia e tecnologia da informação que objetiva prover melhorias aos sistemas e possibilitar novos patamares de qualidade às instituições.

No início do século XXI, novos estudos em metafísica analítica e biologia, em particular, em bioética, fizeram ressurgir certos tipos de conhecimentos dados como obsoletos. E novamente, dentre eles, destacaram-se os postulados de Aristóteles (SMITH; KLAGGES, 2008). Naquele mesmo período ocorreu a criação da primeira disciplina de “Ontologia Aplicada” na Universidade de *Buffalo*⁴ como um braço da filosofia aplicada. O objetivo da disciplina em suas primeiras edições foi registrar *insights* e métodos filosóficos da ontologia pura, para alcançar uma compreensão filosófica dos aspectos da realidade, objetos de pesquisa científica aplicada (SMITH; KLAGGES, 2008).

No aspecto prático, os artefatos ontológicos produzidos a partir da ontologia aplicada são utilizados para atenuar as inconsistências semânticas de um domínio de conhecimento e possibilitar raciocínio automático sobre o conhecimento representado (ALMEIDA, 2013, 2020; MUNN; SMITH, 2008). Os princípios metafísicos norteiam a construção de metodologias para representação do conhecimento em ontologias (SMITH; CEUSTERS, 2010).

Todo esse empreendimento se fundamenta atualmente sob a influência de três sistemas clássicos de categorias: o sistema realista de Aristóteles, o sistema cognitivo de *Kant* e o sistema de ontologias formais do de Edmund Husserl (1859 – 1938). O resultado é um conjunto de teorias bem fundamentadas, que combina consistência teórica e utilidade prática nos modernos sistemas de informação.

3.2 ASPECTOS TEÓRICOS, METODOLÓGICOS E TECNOLÓGICOS

³ Ver *Ontology Research Group*, disponível na Internet em <http://org.buffalo.edu/>. Acesso em: 13/10/2020.

⁴ *State University of New York at Buffalo, New York, United States*. Mais informações na internet em: <https://www.buffalo.edu/>. Acesso em: 10/10/2021.

A literatura científica revela trabalhos seminais que envolvem ontologias, compondo um acervo de visões distintas, mesmo se considerado apenas o viés filosófico.

Sob o enfoque teórico, identificam-se como seminais, de ponto de vista multidisciplinar: na Filosofia, Smith (2003) e Smith e Welty (2001); na Inteligência Artificial, Alexander e outros (1986), Brachman e Levesque (2004); na Ciência da Computação, Gruber (1994), Uschold; Gruninger (1996) e Guarino, 1998; em Sistemas de Informação, Wand e Weber (1990) e Guizzardi (2005), e na Ciência da Informação, Vickery (1986), Soergel (1999), e Almeida (2013, 2020, 2021).

Sob o enfoque metodológico destacam-se a *Methontology* (GÓMEZ-PÉREZ; FERNÁNDEZ; VICENTE, 1996), a TOVE (GRÜNINGER; FOX, 1995), a *NeOn Methodology* (SUÁREZ-FIGUEROA; GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ, 2008); e a *OntoForInfoScience* (MENDONÇA, 2015). Sob o enfoque tecnológico, a construção de ontologias como artefatos hoje demanda a linguagem OWL, recomendada pela W3C em função de sua capacidade de escrita, publicação e compartilhamento na web. A OWL deriva-se da DAML+OIL⁵ e estende o vocabulário do RDF e RDFS, mantendo os fundamentos da lógica descritiva (ARP; SMITH; SPEAR, 2015). Em sua versão atual compreende três sublinguagens distintas quanto à capacidade de representação (expressividade):

- OWL 2 DL: provê semântica com características computacionais desejáveis para sistemas de raciocínio;
- OWL 2 FULL: estende a semântica presente em RDFs, possibilita visualização de grafos RDF;
- OWL 2 EL: aplicada a sistemas de alta complexidade, possibilitando a definição de classes complexas de domínio neutro e axiomas sobre as mesmas.

Além das especificidades das sublinguagens é possível optar entre quatro sintaxes distintas, RDF/XML, OWL/XML, Sintaxe funcional, Sintaxe Manchester, Sintaxe Turtle (W3C, 2012).

A construção de ontologias é uma tarefa complexa, de dedicação intelectual e resultados nem sempre de curto prazo. São diversos os desafios que se apresentam até mesmo antes da etapa de representação, dos quais destacam-se:

⁵ Maiores informações em: <https://www.w3.org/TR/daml+oil-reference>. Acesso em: 25/03/2019.

- A profusão terminológica e conceitual presente em cada domínio de especialidade e que demanda compreensão e desambiguação (jargões) (SCHULZ et al., 2017);
- A profusão de padrões terminológicos em uso concomitante em um mesmo domínio, o que resulta em sobreposições sintáticas e semânticas e impacta nos sistemas de informação (SCHULZ et al., 2017);
- A profusão de sistemas computacionais em uso em um domínio e as heterogeneidades sintáticas, estruturais e semânticas que os permeiam, e os impactos no planejamento da arquitetura tecnológica;
- A capacitação de arquitetos da informação em ontologia (ALMEIDA; FELIPE; BARCELOS, 2019);
- A representação de documentos médicos (ALMEIDA; FARINELLI, 2017);
- A identificação e distinção entre aspectos ontológicos e epistemológicos em padrões terminológicos (TEIXEIRA, 2019);
- A carência de maior detalhamento das fases de tratamento informacional em metodologias para construção de ontologias (MENDONÇA, 2015);
- A articulação de equipe multidisciplinar para envolvimento em diferentes etapas e frentes (ALMEIDA, 2020).

Todos esses desafios carecem de aporte teórico e técnico da Ciência da Informação e áreas afins, para que se possa abordar problemas complexos que afligem áreas vitais da sociedade em todo o mundo. Um exemplo relevante é a conhecida falta de interoperabilidade nos registros e sistemas médicos, que pode ser mitigada a partir de ontologias, mas não com ferramentas e técnicas que os profissionais da informação usavam há 40 ou 50 anos (EMYGDIO, 2021; SCHULZ et al., 2017).

4 CONTRIBUIÇÕES DA ONTOLOGIA APLICADA PARA A RI

Existem aspectos cruciais da representação do conhecimento em ontologias formais que são fortes subsídios para aprimorar a qualidade das informações a serem recuperadas em ambientes diversos. Tais aspectos repousam sobre a teorias no entorno da ontologia aplicada tendo por base os postulados aristotélicos revistos no realismo ontológico (SMITH; CEUSTERS, 2010) e os aspectos da lógica descritiva (BAADER et al., 1991).

A partir da retomada aos princípios filosóficos e do estabelecimento da ontologia aplicada, os estudos sobre os sistemas clássicos de categorias foram revisitados. Já se havia evidenciado a impossibilidade da representação completa da realidade em um único sistema de categorias, de forma que a pesquisa buscou articular um conjunto de princípios de organização do conhecimento em domínios específicos. Pelo menos oito princípios são adotados com este propósito (ALMEIDA, 2020; SMITH et al., 2007). Uma breve explicação sobre eles se faz necessária para esclarecer a abrangência da disciplina em questão.

- realismo;
- perspectivismo;
- falibilismo;
- adequatismo;
- reutilização;
- utilidade;
- atualização
- facilidade.

O princípio do realismo advoga que o objetivo de uma ontologia é o de descrever a realidade, garantindo que cada termo a ser representado possua um referente na realidade e ainda, que as entidades da realidade sejam consideradas em sua natureza própria, abstendo-se de aspectos subjetivos. Esse princípio, apesar de polêmico, evita a profusão desenfreada de novos termos, sobreposições e confusão terminológica.

O princípio do perspectivismo ontológico advoga que uma multiplicidade de ontologias – organizadas em uma arquitetura multinível de sistemas parciais de categorias – é necessária para abranger os vários aspectos da realidade, representados em diversas áreas da pesquisa científica. Este princípio permite a compreensão e a representação da realidade sob ângulos distintos de observação e em níveis igualmente distintos de granularidade, prevenindo a desarticulação quanto à essência da realidade e a representação. Cada sistema parcial de categorias organiza o domínio em classes e tipos, de forma análoga à divisão taxonômica de Carl Linnaeus (1707 – 1778), estabelecida para organizar o domínio de organismos em várias categorias de alto nível (reino, filo, classe, espécie etc.). Essa iniciativa é hoje padronizada em

trabalhos como a International Code of Zoological Nomenclature⁶ e a International Code of Nomenclature of Bacteria.⁷

O princípio do falibilismo implica na necessidade de se estabelecer estratégias de versionamento dos trabalhos de aquisição do conhecimento, garantindo sua aderência às novas descobertas científicas. Este princípio garante uma redução das proliferações de modelos conceituais sobrepostos, uma vez que, restringe a demanda por novas ontologias em função de atualizações, optando por versioná-las, o que pode ocorrer de forma integral ou particular.

O princípio do adequatismo advoga que artefatos ontológicos devem ser desenvolvidos pautados pela interoperabilidade com outros com as quais mantém fronteiras. Tal aspecto também concorre para redução do número de artefatos desenvolvidos, uma vez que demanda apreensão não apenas dos aspectos elementares do domínio, mas dos artefatos preexistentes com os quais deverá se articular de forma consistente. Uma ontologia de alto nível, por exemplo, é um artefato preferencial em atividades de mapeamento, impactando diretamente na economia de recursos de processamento computacional e tempo. Portanto, quanto maior a qualidade na sua construção, maior será a sua relevância para alcance de interoperabilidade semântica.

O princípio da reutilização advoga que ontologias existentes são consideradas referência e devem ser prioritariamente reutilizadas. Desta forma, ampliam-se as possibilidades de avanços em teorias de ontologias uma vez que ontologias de referência, em arquiteturas concebidas a partir desta abordagem, encerram em si mesmas teorias ontológicas, explicitamente declaradas e com sentido formalmente restringido por proposições lógicas. Sua reutilização possibilita releitura de seus pressupostos, revisão à luz dos avanços científicos e atualização, caso necessário. O princípio da utilidade advoga que os princípios realistas devem ser preservados, mantendo-se o foco sobre a característica de alto nível que se propõe para o artefato, não devendo-se abster deles em função de demandas imediatistas.

O princípio da atualização remete à revisão da ontologia em função de avanços científicos e o princípio da facilidade orienta à construção de um artefato a partir dos casos mais simples para entendimento e posterior avanço para aspectos mais complexos do domínio. Este

⁶ Disponível em: <https://code.iczn.org/>. Acesso em: 13/10/2021.

⁷ Disponível em: <http://www.the-icsp.org/bacterial-code>. Acesso em: 13/10/2021.

princípio contribui para a avaliação constante do artefato em construção a partir do processamento de inferências sobre as representações realizadas.

O conjunto de princípios apresentado até aqui resume os postulados aristotélicos e os fundamentos da Basic Formal Ontology⁸ (BFO) uma ontologia de alto nível considerada um *hub* que sustenta atualmente, mais de 300 ontologias subjacentes, no consórcio de ontologias biomédicas abertas denominado Open Biomedical Ontologies (OBO) Foundry.⁹ A BFO possui ampla aceitação no domínio médico e biomédico, a partir da qual foram construídas outras ontologias de referência para a representação do conhecimento, tais como:

- *Gene Ontology* (ASHBURNER et al., 2000);
- *Cell Type Ontology* (BARD; RHEE; ASHBURNER, 2005);
- *Foundational Model of Anatomy* (ROSSE; MEJINO, 2003);
- *Ontology for General Medical Science* (SCHEUERMAN; CEUSTERS; SMITH, 2009);
- *Obstetric and Neonatal Ontology* (FARINELLI et al., 2016).

A maturidade deste projeto e da própria abordagem do realismo ontológico permite inferir que sua adoção tende a impactar positivamente nos aspectos de RI em contextos organizacionais. Isso pode ocorrer de forma direta e indireta, uma vez que ontologias promovem:

- Melhoria na qualidade das representações especialmente pela adoção da lógica descritiva cuja sintaxe e recursos favorecem a declaração explícita do conhecimento e da validação contínua do artefato em construção por inferência;
- Estímulos a melhorias da capacidade intelectual para a elaboração de expressões de consulta (*queries*) inteligentes que resultem em assertividade nos resultados;
- Transparência quanto às entidades ontológicas e epistemológicas do domínio, favorecendo a construção de mapeamentos entre as fontes de informação, e completude de consultas;
- Melhoria no alcance de interoperabilidade semântica no domínio;
- Transparência nos processos organizacionais existentes, favorecendo a identificação de demandas de reengenharia para eliminação de tarefas redundantes;

⁸ Disponível em: <https://basic-formal-ontology.org/>. Acesso em: 13/10/2021.

⁹ Disponível em: <http://www.obofoundry.org/>. Acesso em 03 abr. 2020.

- Melhorias na compreensão em perspectiva sobre a realidade organizacional, aumentando a percepção de oportunidades de inovação em produtos, processos e serviços;
- Oferta de insumos para avaliação dos aspectos de qualidade dos dados organizacionais;
- Capacitação de ontologistas em lógica descritiva e ontologia, ampliando naturalmente seu campo de atuação no mercado.

Os benefícios elencados compreendem um levantamento preliminar, mas possibilitam apreender a necessidade em se estabelecer a ontologia aplicada como uma disciplina obrigatória no currículo da Ciência da Informação.

5 CONSIDERAÇÕES FINAIS

O presente artigo trouxe reflexões para sugerir a criação de uma disciplina de Ontologia aplicada no currículo da Ciência da Informação. Para tanto, a Seção 2 apresentou aportes teóricos sobre RI, além dos principais desafios elencados na literatura. A Seção 3 descreveu aportes relacionados à Ontologia Aplicada, suas origens, aspectos teóricos metodológicos e tecnológicos relacionados, incluindo também os desafios à construção de artefatos ontológicos de qualidade. A Seção 4 sintetizou contribuições da Ontologia Aplicada para a RI, a partir da qualidade das representações via a adoção da abordagem do realismo ontológico e da lógica descritiva.

Conclui-se que a compreensão e a adoção dos princípios da Basic Formal Ontology (BFO), a qual é a representação computacional de princípios discutidos até aqui, tende a ampliar a abrangência das pesquisas da Ciência da Informação, especificamente, no campo da Recuperação da Informação. Atualmente, no Brasil, as atividades relacionadas à BFO e desenvolvimentos em seu entorno restringem-se:

- Às pesquisas realizadas pelo grupo Representação do Conhecimento, Ontologias e Linguagem (RECOL),¹⁰ vinculado ao diretório de Pesquisa do CNPQ;

¹⁰ Maiores informações em: <http://recol.eci.ufmg.br/>. Acesso em: 13 out. 2021.

- Iniciativas isoladas como o Centro Nacional de Pesquisa em Ontologias (NCOR-BR)¹¹, a associação sem fins lucrativos – capítulo do NCOR¹² – onde promovem-se discussões, dissemina-se conhecimento e boas práticas para artefatos ontológicos de qualidade.

Todas essas possibilidades encontram respaldo ainda no fato de que a BFO é a base para elaboração da Norma ISO/IEC PRF 21838-2.2 *Information technology — Top-level ontologies (TLO) — Part 2: Basic Formal Ontology (BFO)*.¹³ Trata-se da norma internacional, mas a sua tradução está em andamento para adoção no Brasil.

Para uma mudança evolutiva na RI no contexto da Ciência da Informação, não se pode prever a realização de pesquisas de desenvolvimento de algoritmo e avanços nas arquiteturas de *hardware*. Estes são trabalhos típicos da Ciência da Computação. Na CI, pode-se, por outro lado, evoluir em aspectos que lhe são caros, como a modelagem, o estudo de usuário, a organização e a representação do conhecimento, as técnicas de categorização e classificação, para citar alguns.

Nesse contexto, o presente ensaio preconiza que a adoção da Ontologia Aplicada como disciplina no corpo de conhecimento da CI, vai impactar positivamente na RI praticada pela área, abrir novos horizontes para pesquisadores e estudantes, assim como aumentar o reconhecimento da área no contexto científico. Para que isso ocorra, a pesquisa em CI ligada a suas origens na Biblioteconomia, precisa evoluir alguns pressupostos que são essenciais para lidar com o ambiente de informação do século XXI. Um desses pressupostos é que a informação foi há muito “desterritorializada”, um termo usado por *Pierre Levi* (1956 –), o chamado “filósofo da cibercultura”, para explicar com o advento da Web impactou em técnicas que se limitam a pensar documentos e espaço físico para eles, em um mundo cada vez mais digital.

REFERÊNCIAS

ALEXANDER, J. H. *et al.* Knowledge Level Engineering Ontological Analysis. In: NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE (AAAI-86). 1986, [Philadelphia]. **Proceedings** [...]. [Philadelphia: s. n.], 1986. p. 6. Disponível em: <https://www.aaai.org/Papers/AAAI/1986/AAAI86-159.pdf>

¹¹ Maiores informações em: <https://ncor-brasil.org/>. Acesso em: 13 out. 2021.

¹² Maiores informações em: <https://ubwp.buffalo.edu/ncor/>. Acesso em: 13 out. 2021.

¹³ Maiores informações em: <https://www.iso.org/standard/74572.html>. Acesso em: 27 maio 2021.

ALMEIDA, M. B. Revisiting ontologies: A necessary clarification. **Journal of the American Society for Information Science and Technology**, v. 64, n. 8, p. 1682-1693, ago. 2013.

ALMEIDA, M. B. **Ontologia em Ciência da Informação: teoria e método**. Curitiba: CRV, 2020. v. 1.

ALMEIDA, M. B. **Ontologia em Ciência da Informação: tecnologia e aplicações**. Curitiba: CRV, 2021. v. 2.

ALMEIDA, M. B.; FARINELLI, F. Ontologies for the representation of electronic medical records: The obstetric and neonatal ontology. **Journal of the Association for Information Science and Technology**, v. 68, n. 11, p. 2529-2542, 2017.

ALMEIDA, M. B.; FELIPE, E. R.; BARCELOS, R. Toward a document-centered ontological theory for information architecture in corporations. **Journal of the Association for Information Science and Technology**, v. 71, n. 11, p. 1308-1326, Nov. 2020.

ARP, R.; SMITH, B.; SPEAR, A. D. **Building Ontologies with Basic Formal Ontology**. Cambridge, Massachusetts: The MIT Press, 2015.

ASHBURNER, M. *et al.* Gene Ontology: tool for the unification of biology. **Nature Genetics**, v. 25, p. 25-29, May 2000.

BAADER, F. *et al.* **Terminological Knowledge Representation: A Proposal for a Terminological Logic**. [S. l.: s. n.], 1991.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern information retrieval: the concepts and technology behind search**. 2. ed. New York: Addison Wesley, 2011.

BARD J.; RHEE SY.; ASHBURNER M. An ontology for cell types. **Genome Biology**, London, v. 6, n. 2, 2005.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, v. 1, p. 34-43, May 2001.

BERNIER, C. L.; CRANE, E. J. Indexing abstracts. **Industrial and Engineering Chemistry**. v. 40, n. 4, p. 725-30, 1948.

BRACHMAN, R. J.; LEVESQUE, H. J.; PAGNUCCO, M. **Knowledge Representation and Reasoning**. 1. ed. [s. l.]: Morgan Kaufmann, 2004.

BUSH, V. As We May Think. **Life Magazine**, p. 112-124, Oct. 1945.

CHEN, P. P.-S. The Entity-Relationship Model-Toward a Unified View of Data. **ACM Transactions on Database Systems**, v. 1, n. 1, p. 9-36, 1976.

CLEVERDON, C. Evaluation Tests of Information Retrieval Systems. **Journal of Documentation**, v. 26, n. 1, p. 55-67, jan. 1970.

CODD, E. F. A Relational Model of Data for Large Shared Data Banks. **Commun. ACM**, v. 13, n. 6, p. 377-387, jun. 1970.

DAHLBERG, I. A referent-oriented analytical concept theory of interconcept. **International Classification**, v. 5, n. 3, p. 142-150, 1978.

DATE, C. J. **Introdução a sistemas de bancos de dados**. Rio de Janeiro: Campos, 1984.

EMYGDIO, J. L. **Interoperabilidade semântica orientada por ontologia para a Ciência da Informação**: a metodologia Onto4All-Interoperability como resultado de estudo de caso no domínio de energia. 2021. Tese (Doutorado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2021.

FARINELLI, F. *et al.* IT403: OntONeo: The Obstetric and Neonatal Ontology. *In*: INTERNATIONAL CONFERENCE ON BIOMEDICAL ONTOLOGY AND BIOCREATIVE, 7., 2016, Corvallis, Oregon. **Proceedings** [...]. [S. l.]: CEUR-ws.org, 2016.

FELIPE, E. R. **A expansão de queries sobre terminologias biomédicas**: uma comparação de artefatos de representação do conhecimento para Recuperação de Informações. 2020. Tese (Doutorado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2020.

GÓMEZ-PÉREZ, A.; FERNÁNDEZ, M.; VICENTE, A. de. Towards a Method to Conceptualize Domain Ontologies. *In*: 12TH EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE. 1996, Budapest. **Proceedings** [...]. Budapest, Rumania: Facultad de Informática (UPM), 1996.

GONZALEZ, M.; LIMA, V. L. S. de. **Recuperação de informação e processamento da linguagem natural**. [Marília: Unesp], 2003. Disponível em: <https://www.marilia.unesp.br/Home/Instituicao/Docentes/EdbertoFerneda/mri-06---gonzales-e-lima-2003.pdf>. Acesso em: 11 out. 2021.

GRUBER, T. R. Towards principles for the design of ontologies used for knowledge sharing. *In*: GUARINO, N.; POLI, R. (ed.). **Formal Ontology in Conceptual Analysis and Knowledge Representation**. [London: England Academic Press], 1994.

GRÜNINGER, M.; FOX, M. S. Methodology for the Design and Evaluation of Ontologies. Workshop on Basic Ontological Issues in Knowledge Sharing. *In*: THE 1995 INTERNATIONAL JOINT CONFERENCE ON AI. 1995, Montreal. **Anais** [...]. Montreal: [AAAI; CSCSI], 1995.

GUARINO, N. Formal Ontology and Information Systems. *In*: FORMAL ONTOLOGY IN INFORMATION SYSTEMS. 1998. Trento, Italy. **Proceedings** [...]. Amsterdam: IOS Press, 1998.

GUIZZARDI, G. **Ontological Foundations For Structural Conceptual Models**. Enschede: Centre for Telematics and Information Technology, University of Twente, 2005.

JACOBSON, I. *et al.* **Object-oriented software engineering: a use case driven approach.** New York: Addison-Wesley, 1992.

JARDINE, D. A. The ANSI/SPARC DBMS model. *In: SHARE WORKING CONFERENCE ON DATABASE MANAGEMENT SYSTEMS*, 2., 1976, Montreal. **Proceedings [...]**. Amsterdam: North Holland, 1976. p. 1-225.

LESK, M. **The seven ages of information retrieval.** Ottawa: IFLA, 1996. v. 5. Disponível em: <https://archive.ifla.org/VI/5/op/udtop5/udt-op5.pdf>. Acesso em: 11 Oct. 2021.

LUHN, H. P. A new method of recording and searching information. **American Documentation**, v. 4, n. 1, p. 14-16, jan. 1953.

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, Oct. 1957.

MANNING, C. D.; RAGHAVAN, P.; SCHUTZE, H. **Introduction to Information Retrieval.** United States of America: Cambridge University Press, 2008.

MENDONÇA, F. M. **Ontoforinfoscience: metodologia para construção de ontologias pelos cientistas da informação - Uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano (HEMONTA).** 2015. Tese (Doutorado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

MOOERS, C. N. **The Theory of Digital Handling of Non-numerical Information And Its Implications to Machine Economics.** Boston: Zator, 1950.

MOOERS, C. N. **Making information retrieval pay.** Boston: Zator, 1951.

MOOERS, C. N. Zatorcoding and Developments in Information Retrieval. **Aslib Proceedings**, v. 8, n. 1, p. 3-22, jan. 1956.

MOOERS, C. N. The next twenty years in information retrieval; some goals and predictions. **American Documentation**, v. 11, n. 3, p. 229-236, Aug. 1960.

MUNN, K.; SMITH, B. **Applied Ontology: An Introduction.** [s. l.] Frankfurt: Ontos, 2008.

NATIONAL CENTER FOR SUPERCOMPUTING APPLICATIONS. **NCSA Mosaic.** (1993) Seção **Enabling discovery.** Disponível em: <http://www.ncsa.edu/Projects/mosaic.html>. Acesso em: 11 Oct. 2021.

ROBERTSON, S. The Probability Ranking Principle in IR. **Journal of Documentation**, v. 33, p. 294-304, dez. 1977.

ROSSE, C.; MEJINO, J. L. V. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. **Journal of Biomedical Informatics, Unified Medical Language System**, v. 36, n. 6, 2003.

SARACEVIC, T. **On the concept of relevance in information science**. [Cleveland]: Case Western Reserve University, 1970.

SALTON, Gerald. **Automatic information organization and retrieval**. New York: McGraw-Hill Computer Science Series, 1968.

SALTON, G. MCGILL, M. J. **Introduction to modern information retrieval**. New York: McGraw-Hill Computer Science Series, 1983.

SCHEUERMANN, R. H.; CEUSTERS, W.; SMITH, B. Toward an ontological treatment of disease and diagnosis. **Summit on Translat Bioinforma**, v. 2009, (20090301), p. 116-120, Mar. 2009.

SCHULZ, S. *et al.* Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration. **Studies in health technology and informatics**, v. 245, p. 940-944, 2017.

SMITH, B. Ontology. *In: The Blackwell Guide to the Philosophy of Computing and Information*. Oxford: Luciano Floridi, 2003.

SMITH, B.; CEUSTERS, W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. **Applied ontology**, v. 5, n. 3/4, p. 139-188, Nov. 2010.

SMITH, B.; KLAGGES, B. Philosophy and Biomedical Information Systems *In: MUNN, K.; SMITH, B. Applied Ontology: An Introduction*. Frankfurt: Ontos, 2008. Chapter 1.

SMITH, B.; WELTY, C. FOIS introduction - Ontology: Towards a New Synthesis. *In: THE INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS*, 2001, New York. **Anais [...]**. New York: ACM Press, 2001. Disponível em: <http://portal.acm.org/citation.cfm?doid=505168.505201>

SMITH, B. *et al.* The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. **Nature Biotechnology**, v. 25, n. 11, p. 1251-1255, Nov. 2007.

SOERGEL, D. The Rise of Ontologies or the Reinvention of **Classification**. **Journal of the American Society for Information Science**, v. 50, n. 12, p. 1119-1120, Oct. 1999.

SOUZA, R. R.; TUDHOPE, D.; ALMEIDA, M. B. Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems. **Knowledge Organization**, v. 39, n. 3, p. 179-192, 2012.

SUÁREZ-FIGUEROA, M. C.; GÓMEZ-PÉREZ, A.; FERNÁNDEZ-LÓPEZ, M. The NeOn Methodology for Ontology Engineering. *In: SUÁREZ-FIGUEROA, M. C. et al. (ed.). Ontology Engineering in a Networked World*. Berlin: Springer Berlin Heidelberg, 2012. p. 9-34

TEIXEIRA, L. M. D. **Princípios ontológicos aplicados à classificação internacional de doenças**: alternativas para a busca por interoperabilidade semântica entre sistemas de

prontuários eletrônicos de pacientes. 2019. Tese (Doutorado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2019.

USCHOLD, M.; GRUNINGER, M. Ontologies: principles, methods and applications. **Knowledge Engineering Review**, v. 11, n. 2, 1996.

WAND, Y.; WEBER, R. An ontological model of an information system. **IEEE Transactions on Software Engineering**, v. 16, n. 11, 1990.

WUESTER, E. L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses. *In*: RONDEAU, G.; FELBER, F. (org.). **Textes choisis de terminologie**. I. Fondements théoriques de la terminologie. Québec: GIRSTERM, 1981. p. 57-114.

W3C. (2012). OWL-2 Web Ontology Language document Overview – W3C Recommendation. Disponível em: <http://www.w3.org/TR/2012/REC-owl2-overview-20121211/>. Acesso em: 10 ago. 2021.

VICKERY, B. C. Knowledge representation: a brief review. **Journal of Documentation**, v. 42, n. 2, p. 145-159, 1986.

VICKERY, B. C. Ontologies. **Journal of Information Science**, v. 23, n. 4, p. 277-286, 1997.