

ISSN 0100-1965 e-ISSN 1518-8353

CIÊNCIA DA INFORMAÇÃO

v.46 n.1 jan./abr. de 2017

Special thematic issue
Número especial temático

Ontologies in information science: Brazilian state of the art

**Ontologias na ciência da informação:
estado da arte no Brazil**

**Ontologías en la ciencia da la información:
estado del arte en Brazil**

Ciência da Informação
v. 46 n.1 jan./abr. 2017

ISSN 0100-1965 eISSN 1518-8353

Special thematic issue

Número especial temático

**Ontologies in information science:
Brazilian state of the art**

**Ontologias na ciência da informação:
estado da arte no Brazil**

**Ontologías en la ciencia da la información:
estado del arte en Brazil**

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

Diretoria

Cecília Leite Oliveira

Seção de Editoração

Ramón Martins Sodoma da Fonseca

Indexação

Ciência da Informação tem seus artigos indexados ou resumidos.

Bases Internacionais: Paschal Thema: Science de L'Information, Documentation Library and Information Science Abstracts PAIS Foreign Language Index Information Science Abstracts Library and Literature Páginas de Contenido: Ciencias de la Información EDUCACION: Noticias de Educación, Ciencia y Cultura Iberoamericanas Referativnyi Zhurnal: Informatika. ISTA Information Science & Technology Abstracts. LISTA Library, Information Science & Technology Abstracts. SciELO Scientific Electronic Library On-line. Latindex – Sistema Regional de Información em Línea para Revistas Científicas de América Latina el Caribe, España y Portugal, México. INFOBILA: Información Bibliotecológica Latinoamericana.

Indexação em Bases de Dados Nacionais

Portal de Periódicos: LivRe – Portal de Periódicos de Livre Acesso

Comissão Nacional de Energia Nuclear (Cnen). Portal Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes).

Portal de Associações Nacionais: Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (Ancib).

Bases de Dados Nacionais: Base de Dados Referenciais de Artigos de Periódicos de Ciência da Informação da Universidade Federal do Paraná (Brapci). Escola de Ciência da Informação da Universidade Federal de Minas Gerais (Peri).

Instruções para aquisição

Abrir o *site* do Ibict (www.ibict.br).

Selecionar a opção “Guia de Recolhimento da União” da lista “Outros links”.

Preencher os campos “Código”, “Gestão” e “Recolhimento Código” como indicado abaixo:

Código: 240121;

Gestão: 00001;

Recolhimento Código: 288187

Selecionar opção “Emitir GRU simples”, imprimindo-a em seguida.

Pagar GRU simples em uma agência do Banco do Brasil.

Enviar, via fax, o comprovante de pagamento (contendo nome, endereço e telefone do usuário) para o Ibict/NCP, Fax: (61) 3217-6494.

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

Núcleo de Comercialização de Produtos (NCP)

SAUS, Quadra 5, Lote 6, Bloco H, sobreloja

70070-912 – Brasília-DF

Tels.: 55 (61) 3217-6161

3217-6162

nucom@ibict.br

Editada em novembro de 2017.

Última edição em janeiro de 2018.

Publicada em dezembro de 2017.

Ciência da Informação
v. 46 n.1 jan./abr. 2017

ISSN 0100-1965 eISSN 1518-8353

Special thematic issue

Número especial temático

**Ontologies in information science:
Brazilian state of the art**

**Ontologias na ciência da informação:
estado da arte no Brazil**

**Ontologías en la ciencia da la información:
estado del arte en Brazil**



© 2017 Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

Os autores são responsáveis pela apresentação dos fatos contidos e opiniões expressas nesta obra.

Equipe técnica

Editora Científica

Cecília Leite Oliveira

Coordenador Editorial do número

Maurício B. Almeida

Editor Executivo

Ramón Martins Sodoma da Fonseca

Projeto Gráfico

Anderson Moraes

Leila Mendonça Raulino

Capa

Rodrigo Azevedo

Editoras assistentes

Gislaine Russo de Moraes Brito

Revisão gramatical e visual

Margaret de Palermo Silva

Tradução

COED/Ibict

Normalização de referências

Priscilla Mara Bermudes (Normalização Bibliográfica)

Comitê Editorial (agosto de 2016 a agosto de 2018)

Lena Vânia Ribeiro Pinheiro

Instituto Brasileiro de Informação em Ciência e Tecnologia,
Coordenação de Ensino e Pesquisa – Rio Janeiro, RJ – Brasil.

Luc Quoniam

Centro Franco-Brasileiro de Documentação Técnica e Científica
(CenDoTeC) – São Paulo, SP – Brasil.

Mariângela Spotti Lopes Fujita

Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade
de Filosofia e Ciências de Marília, Departamento de Ciência da
Informação – Marília, SP – Brasil.

Guilherme Ataíde Dias

Universidade Federal para Paraíba, Departamento de Ciência da
Informação – Paraíba, PB – Brasil.

Aurora Cuevas Cerveró

Universidad Complutense de Madrid, Departamento de
Biblioteconomía y Documentación – Espanha.

Lillian Maria Araújo Rezende Alvares

Instituto Brasileiro de Informação em Ciência e Tecnologia,
Coordenação-Geral de Pesquisa e Manutenção de Produtos
Consolidados – Brasília, DF – Brasil.

Marcus Granato

Museu de Astronomia e Ciências Afins – Rio de Janeiro, RJ – Brasil.

Gaëlle Bequet

Diretora

Centor Internacional do ISSN – Cieps – França

NOTAS DO EDITOR

Agradecimento especial a todos os avaliadores, em especial aos profs. Lillian Alvares e Maurício B. Almeida, na idealização e produção deste número. Para baixar o PDF de cada artigo da revista *Ciência da Informação* a partir do seu smartphone ou tablet, escaneie o QR Code publicado em cada artigo da versão impressa.

Mais informações pelo telefone: (61) 3217-6145

Ciência da Informação/Instituto Brasileiro de Informação em Ciência e Tecnologia

– Vol. 1, n. 1 (1972) – Brasília: Ibict, 1972 –

Quadrimestral

Até o v. 20, 1991, publicada semestralmente. De 1972 a 1975 editada pelo Instituto Brasileiro de Bibliografia e Documentação (IBBD).

ISSN impresso 0100-1965. eISSN 1518-8353.

1. Ciência da Informação – Periódicos I. Brasil, Instituto Brasileiro de Informação em Ciência e Tecnologia.

CDU 02 (05)

CDD 020.5

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

Setor de Autarquias Sul (SAUS)
Quadra 05, Lote 06, Bloco H – 5º Andar
Cep: 70070-912 – Brasília, DF
Telefones: 55 (61) 3217-6360
55 (61) 3217-6350
www.ibict.br

Rua Lauro Muller, 455 - 4º Andar - Botafogo
Cep: 22290-160 – Rio de Janeiro, RJ
Telefones: 55 (21) 2275-0321
Fax: 55 (21) 2275-3590
<http://www.ibict.br/capacitacao-e-ensino/pos-graduacao-em-ciencia-da-informacao>
<http://www.ppgci.ufrj.br>

AVALIADORES DESTE NÚMERO

Benildes Coura Moreira dos Santos Maculan

Doutora em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG) - Brasil. Professora da Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/5336218259257800>

Fabício Martins Mendonça

Doutor em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil, com período sanduíche na Universidade do Porto (U.Porto) - Portugal. Professor da Universidade Federal de Juiz de Fora (UFJF) - Juiz de Fora, MG - Brasil.

<http://lattes.cnpq.br/7587726616949092>

Maria Luiza de Almeida Campos

Pós-Doutorado pela Universidade de São Paulo (USP) - São Paulo, SP - Brasil.

Pós-Doutorado pela Fundação Instituto Oswaldo Cruz (Fiocruz) - Rio de Janeiro, RJ - Brasil. Doutora em Ciência da Informação pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict/UFRJ) - Rio de Janeiro, RJ - Brasil. Professora da Universidade Federal Fluminense (UFF) - Niterói, RJ - Brasil.

<http://lattes.cnpq.br/9545682339961651>

Renata Maria Abrantes Baracho Porto

Doutora em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG) - Brasil, com período sanduíche na The Pennsylvania State University, Pensilvânia - Estados Unidos. Professora da Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/4218954956709188>

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>

Beatriz Valadares Cendón

Doutora em Information Science pela University of Texas at Austin (UT Austin), Estados Unidos. Professora da Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/5217795669575718>

Dagobert Soergel

Doutor em Political Science pela Universität Freiburg (UNI FREIBURG), Alemanha. Professor da University of Maryland (UMD), Estados Unidos.

<http://lattes.cnpq.br/2442859142109715>

Frederico Fonseca

Ph.D. em Spatial Information Science and Engineering pela University of Maine (UMaine), Estados Unidos. Professor da Pennsylvania State University (Penn State), Estados Unidos.

<https://ist.psu.edu/directory/faculty/fuf1>

Maurício B. Almeida

Pós-Doutorado pela State University of New York at Buffalo (SUNY) - Estados Unidos. Doutor em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG) - Brasil. Professor da Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/5218069708058487>

Ricardo Rodrigues Barbosa

Pós-Doutorado pela University of Toronto (UTORONTO), Canadá. Doutor em Administração de Empresas pela Columbia University (COLUMBIA), Estados Unidos. Professor da Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/6523065261260011>

Ciência da Informação

Volume 46 - número 1 - jan./abr. 2017

Sumário

Table of Contents

Presentation	9
<i>Apresentação</i>	
<i>Presentación</i>	
Dagobert Soergel	
<hr/>	
Editorial	18
Mauricio B. Almeida	
<hr/>	
Part I Theoretical and methodological contributions of research on ontologies	22
<i>Parte I Contribuições teóricas e metodológicas da pesquisa em ontologias</i>	
<i>Parte I Contribuciones teóricas y metodológicas de la investigación en ontologías</i>	
Ontologias em ciência da informação: um estudo bibliométrico no Brasil	23
<i>Ontologies in information science: a bibliometric study in Brazil</i>	
<i>Ontología en ciencia de la información: un estudio bibliométrico en Brasil</i>	
Frederico de Carvalho Figueiredo	
Fernanda Gomes Almeida	
Metrics and methods for comparative ontology evaluation	34
<i>Métricas e métodos para a avaliação comparativa de ontologias</i>	
<i>Métricas y métodos para la evaluación comparativa de ontologías</i>	
Amanda Hicks	
Construindo ontologias com a metodologia ontoforinfoscience: uma abordagem detalhada das atividades do desenvolvimento ontológico	43
<i>Bulding ontologies with ontoforinfoscience methodology: a detailed approach to the activiyies of ontology development</i>	
<i>Construyendo ontologías con la metodología ontoforinfoscience: un enfoque detallado de las actividades del desarrollo ontológico</i>	
Antônio Lucas Soares	
Estudo comparativo das estruturas semânticas em diferentes sistemas de organização do conhecimento	60
<i>Comparative study of semantic structures in different systems of knowledge organization</i>	
<i>Estudio comparativo de estructuras semánticas en diferentes sistemas de organización del conocimiento</i>	
Gercina Ângela de Lima	
Benildes Coura Moreira dos Santos Maculan	

Guidelines for writing definitions in ontologies	73
<i>Diretrizes para criação de definições em ontologias</i>	
<i>Directrices para creación de definiciones en ontologías</i>	
Selja Seppälä	
Alan Ruttenberg	
Barry Smith	
Elaboração de modelos de domínio em ontologias: a abordagem onomasiológica e a função da definição	89
<i>Domain models development in ontologies: the onomasiological approach and the function of the definition</i>	
<i>Elaboración de modelos de dominio en ontologías: el enfoque onomasiológico y la función de la definición</i>	
Maria Luiza de Almeida Campos	
Desambiguação de relações em tesouros e o seu reuso em ontologias	102
<i>An approach to deal with ambiguity of relationships in thesauri for reuse in ontologies</i>	
<i>Desambiguación de relaciones en tesoros y su reutilización en ontologías</i>	
Benildes Coura Moreira dos Santos Maculan	
Elisângela Cristina Aganette	
<hr/>	
Part II Applications and uses of ontologies in information science	121
<i>Parte I Contribuições teóricas e metodológicas da pesquisa em ontologias</i>	
<i>Parte I Contribuciones teóricas y metodológicas de la investigación en ontologías</i>	
Construção de ontologia na prática: um estudo de caso aplicado ao domínio obstétrico	122
<i>Ontology building in practice: a case study applied to the obstetric domain</i>	
<i>Construcción de la ontología en la práctica: un estudio de caso aplicado al dominio obstétrico</i>	
Fernanda Farinelli	
Peter L. Elkin	
Semantic middleware for industrial sensors	141
<i>Middleware semântico para sensores industriais</i>	
<i>Middleware semánticos para sensores industriales</i>	
Fernando Silva Parreiras	
Vitor Afonso Pinto	
Marco Antônio Calijorne Soares	
Daniel Henrique Mourão Falci	
Indexação automática e ontologias: identificação dos contributos convergentes na ciência da informação	152
<i>Automatic indexing and ontologies: Identification of convergent contributions in Information Science</i>	
<i>Indización automática y ontologías: Identificación de las contribuciones convergentes en la Ciencia de la Información</i>	
Maria da Graça de Melo Simões	
Luís Miguel Oliveira Machado	
Renato Rocha Souza	
António Tavares Lopes	

Aplicação e uso de ontologias musicais	169
<i>Application and use of music ontologies</i>	
<i>Aplicación y uso de ontologías musicales</i>	
Lígia Maria Arruda Café Camila Monteiro de Barros	
Ontologias como suporte à modelagem da informação na arquitetura, engenharia e construção	183
<i>Ontologies as support for information modeling in architecture, engineering and construction</i>	
<i>Ontologías como soporte para modelar la información en arquitectura, ingeniería y construcción</i>	
Renata M. A. Baracho Livia Marangon Duffles Teixeira Mário Lucio Pereira Junior	
Ontologia aplicada a um modelo de gestão organizacional: contribuições da ciência da informação	196
<i>Ontology applied to an organizational management model: contributions of information science</i>	
<i>Ontología aplicada a un modelo de gestión organizacional: contribuciones de la ciencia de la información</i>	
Cláudio Gottschalg Duque Geraldino Gonçalves Bastos	
Ontological approach to the normative dimension of organizations: an application of Documents Acts Ontology	213
<i>Abordagem ontológica à dimensão normativa das corporações: uma aplicação da ontologia Atos dos Documentos</i>	
<i>Enfoque ontológico a la dimensión normativa de las corporaciones: una aplicación de la ontología de Actos de Documentos</i>	
Maurício B. Almeida Mathias Brochhausen Fabiana Bigão Silva	

Presentation

Dagobert Soergel
PhD, University of Maryland

What follows are some remarks on the topics covered in the papers in this issue, culminating in a call for stronger collaboration among the communities dealing with issues of ontologies and other knowledge organization systems and an effort towards unification of the thinking and the results achieved by these communities.

I begin with recalling some distinctions to place the papers in this issue in context. These distinctions are pragmatic and useful, but they can be — and have been — debated endlessly, sometimes on logical grounds. Most of these debates I find decidedly not useful.

In dealing with ontologies and their use in representing data there are three levels:

The papers in this issue focus on Entity Value Ontologies. One gets the impression that for some authors what makes a system worthy of the name “ontology” depends not so much on the content or on the level (in the sense presented in the table), but rather on the degree of formality, such as being expressed in RDF, OWL, or some more expressive logic format. This stance should be revisited

what is important the content a system has to offer and how precisely that content is expressed (Things can be expressed very precisely in natural language).

Three levels in representing data		
Data model ontologies	Define entity types and relationship types. Entity types are classes, but not all classes are entity types. For example, taxon is an entity type and also a class the members of the class, or the entity values of the entity type, are taxa, such as <i>Rosa multiflora</i> <i>Rosa multiflora</i> is, in turn, a class whose members are individual plant specimens. Examples: Basic Format Ontology (BFO, including BFO relationships), Unified Foundational Ontology (UFO), FOAF, Dublin Core	Tbox Terminological axioms
Entity value ontologies	These systems deal with entity values – biological taxa, body parts and body systems (anatomy plus), genes, biological functions, diseases and disorders, places, historical periods, cultures, artistic styles, cultural objects, concepts – and terms used to name them. Such systems are also known as classification schemes, taxonomies, subject heading lists, or thesauri, collectively known as knowledge organization systems. Systems that use the term “ontology” in their name usually have a more elaborate structure with differentiated relationships. Examples: The Universal Decimal Classification (UDC), the Gene Ontology, (GO) SnoMed, the International Classification of Diseases (ICD), the Medical Subject Headings (MeSH), AGROVOC, the Thesaurus of Psychological Index Terms, the Getty vocabularies, including the Art and Architecture Thesaurus (AAT).	
Data sets, data files, databases	Data sets are here construed as collections of assertions about reality (a term used here naively without taking any philosophical stance) or representing thought or imagination and dreams. Assertions are formed from relationship types by filling in entity values for the relationship arguments. In their simplest form assertions take the form of triples, but representing the complexity of reality and thought requires multi-argument relationships. The examples in Hick’s paper on ontology evaluation (Paper I-2) are mostly assertion, indicating that she considers data sets as ontologies, at least if they are presented in a formal way.	Abox Assertional axioms
Boundaries between levels are porous	The boundaries between these levels are by no means strict. For example, the Unified Medical Language System (UMLS) Semantic Types list animal (a class) and even mammal and human as an entity type, while to me animal is a value of the entity type organism, albeit high up in the hierarchy of these entity values. Many thesauri, such as MeSH or the Alcohol and Other Drug Thesaurus include relationships that are assertions about reality. For example, in the MeSH hierarchy there is a heading anti-infective agents listing a chemical substance under this heading is an assertion about reality. MeSH includes such relationships because they are useful for searching. In a broad view, all three levels could be combined into one big data base without worrying which level an element belongs to. On the other hand, distinction between the levels is useful. Systems must decide pragmatically what information to include at what level	

The papers discuss design and structure, development (closely linked to design), evaluation, and application of ontologies, broadly construed.

Lima and Maculan (*Comparative study of semantic structures in different knowledge organization systems*) and Maculan and Aganette (*Ontologia aplicada a um modelo de gestão organizacional: contribuições da Ciência da Informação*) discuss the relational structure of ontologies and other knowledge organization systems. Seppälä, Ruttenberg and Smith (*Guidelines for Writing Definitions in Ontologies*), as well as Maria Campos (*Elaboration of domain models in ontologies: the onomasiological approach and the function of the definition*), discuss a specific but important aspect of design, namely the bringing together concepts and terms and writing of good definitions.

Mendonça (*Bulding ontologies through the OntoForInfoScience Methodology: a detailed approach to perform activities of the ontology development*), and Soares, Farinelli and Elkin (*Ontology building in practice: a case study applied to the obstetric domain*), deal with the process of developing ontologies.

Evaluation of ontologies is the topic of Hicks (*Metrics and methods for comparative ontology evaluation*). This paper makes excellent points, but some of the metrics are more suitable at the assertion level than at the data model and the entity value levels, where the focus of evaluation should be. The literatures on the evaluation of data models and on the evaluation of thesauri, classifications, and similar knowledge organization systems have useful insights to offer.

The remaining papers, except *Ontologies in the context of Information Science: a bibliometric study in Brazil*, deal with various applications of ontologies, by function – harmonizing sensor data (*Semantic Middleware for Industrial Sensors*), indexing (*Automatic indexing and ontologies: identification of convergent contributions in Information Science*) and by domain, for example,

music (*Application and use of music ontologies*)

architecture, engineering, and civil construction (*Ontologies as support for information modeling in architecture, engineering and construction*), and, on a deep philosophical level, organizations and corporations (*An ontological approach to the normative dimension of organizations: an application of Documents Acts Ontology*).

Figueiredo and Almeida report on a bibliographic analysis of the literature on ontologies in Brazil. This is a good starting point, but any such analysis is limited by the source(s) covered and the terms used for searching. The unification of thinking and results from several fields called for in the next paragraph could be supported by extending this work, using a much broader group of sources and a broader approach to the selection of search terms.

It is welcome that this issue combines contributions from *information science* (a label commonly used for the community best represented by organizations such as the Association for Information Science and Technology (ASIST), the International Association for Knowledge Organization (ISKO), and, to some extent, library organization) and *ontology* (a label commonly used for the community best represented by the International Association for Ontology and its Applications (IAOA) and the Ontology Summit). These communities deal with much the same issues – from a different perspective perhaps, drawing on different scientific traditions, using different terminology and different methods – but the same issues nonetheless. As this issue demonstrates, there is growing recognition of this commonality of interests and that joining forces would be better for making progress than working in parallel and separate universes.

Much closer collaboration between OntoBras, ISKO Brasil, and the Brasil chapter of ASIST in organizing conferences and in other applications and closer collaboration between university

faculty from both communities (often across the boundary information science – computer science – would be beneficial for both sides. There are two more communities that should be involved: data modeling and linguistics.

Linguists have developed their own upper level ontologies (not too dissimilar from BFO, for example) for their purpose of explaining language structure and use. Unifying all these contributions may require for data modeling, ontologies, and other knowledge organization systems and effort comparable in scope and insight to what the group known under the name Bourbaki did for mathematics

Apresentação

Dagobert Soergel

PhD, Universidade de Maryland

A seguir estão algumas observações sobre os tópicos abordados nos trabalhos desta questão, culminando em um apelo a uma colaboração mais forte entre as comunidades que lidam com questões de ontologias e outros sistemas de organização do conhecimento e um esforço para a unificação do pensamento e os resultados alcançados por estas comunidades.

Começo por recordar algumas distinções para colocar os documentos nesta edição em contexto. Essas distinções são pragmáticas e úteis, mas podem ser — e têm sido — debatidas sem parar, às vezes por motivos lógicos. A maioria desses debates é decididamente inútil.

Ao lidar com ontologias e seu uso na representação de dados, existem três níveis:

Os trabalhos nesta edição se concentram em Ontologias de Valor de Entidade. Pode-se ter a impressão de que, para alguns autores, o que faz um sistema digno do nome "ontologia" não depende tanto do conteúdo quanto do nível (no sentido apresentado na tabela), mas sim do grau de formalidade, como sendo expresso em RDF, OWL, ou algum formato de lógica mais expressiva. Esta posição deve ser revisada

O que é importante o conteúdo que um sistema tem para oferecer e com que precisão esse conteúdo é expresso (as coisas podem ser expressas de forma muito precisa em linguagem natural).

Três níveis de representação de dados

Três níveis de representação de dados		
Ontologias de modelo de dados	Defina tipos de entidades e tipos de relacionamento. Os tipos de entidades são classes, mas nem todas as classes são tipos de entidade. Por exemplo, o taxon é um tipo de entidade e também uma classe os membros da classe, ou os valores da entidade do tipo de entidade, são taxa, como Rosa multiflora Rosa multiflora é, por sua vez, uma classe cujos membros são espécimes de plantas individuais. Exemplos: Ontologia de Formato Básico (BFO, incluindo relacionamentos BFO), Ontologia Fundamental Unificada (UFO), FOAF, Dublin Core	Tbox Axiomas terminológicos
Ontologias de valor de entidade	Esses sistemas lidam com valores de entidades - taxa biológica, partes do corpo e sistemas corporais (anatomia mais), genes, funções biológicas, doenças e distúrbios, lugares, períodos históricos, culturas, estilos artísticos, objetos culturais, conceitos - e termos usados para nomeá-los. Tais sistemas também são conhecidos como esquemas de classificação, taxonomias, listas de títulos de assunto ou tesouros, conhecidos coletivamente como sistemas de organização do conhecimento. Os sistemas que utilizam o termo "ontologia" em seu nome geralmente possuem uma estrutura mais elaborada com relacionamentos diferenciados. Exemplos: A Classificação Decimal Universal (UDC), a Ontologia de Gênero, (GO) SnoMed, a Classificação Internacional de Doenças (ICD), as Seções de Assunto Médico (MeSH), AGROVOC, o Tesouro de Termos de Índice Psicológica, os vocabulários de Getty, incluindo o Thesaurus de Arte e Arquitetura (AAT).	
Conjuntos de dados, conjuntos de arquivos, bases de dados	Os conjuntos de dados são aqui interpretados como coleções de afirmações sobre a realidade (um termo usado aqui ingenuamente sem tomar qualquer posição filosófica) ou representando pensamento ou imaginação e sonhos. As afirmações são formadas a partir de tipos de relacionamento preenchendo valores de entidade para os argumentos de relacionamento. Na sua forma mais simples, as afirmações assumem a forma de triplos, mas representar a complexidade da realidade e do pensamento requer relacionamentos multi-argumentos. Os exemplos no artigo de Hick sobre a avaliação da ontologia (Paper I-2) são principalmente afirmativos, indicando que ela considera os conjuntos de dados como ontologias, pelo menos, são apresentados de forma formal.	Abox Axiomas assertivos
Limites entre níveis são porosos	Os limites entre esses níveis não são estritos. Por exemplo, os tipos semânticos do Sistema de linguagem médica unificada (UMLS) listam animais (uma classe) e até mesmo mamíferos e humanos como um tipo de entidade, enquanto que eu animal é um valor do organismo de tipo de entidade, embora alto na hierarquia destes valores da entidade. Muitos tesouros, como MeSH ou o Thaurus de álcool e outros medicamentos, incluem relacionamentos que são asserções sobre a realidade. Por exemplo, na hierarquia MeSH existe um título de agentes anti-infecciosos Listar uma substância química sob este título é uma afirmação sobre a realidade. A MeSH inclui tais relacionamentos porque são úteis para pesquisa. Em uma visão ampla, todos os três níveis poderiam ser combinados em uma grande base de dados sem se preocupar com qual nível pertence um elemento. Por outro lado, a distinção entre os níveis é útil. Os sistemas devem decidir pragmaticamente quais informações incluir em que nível	

Os trabalhos discutem *design* e estrutura, desenvolvimento (estritamente ligado ao design), avaliação e aplicação de ontologias, amplamente interpretadas.

Lima e Maculan (*Comparative study of semantic structures in different knowledge organization systems*) e Maculan e Aganette (*Ontologia aplicada a um modelo de gestão organizacional: contribuições da Ciência da Informação*) discutem a estrutura relacional das ontologias e outros sistemas de organização do conhecimento. Seppälä, Ruttenberg e Smith (*Guidelines for Writing Definitions in Ontologies*), bem como Maria Campos (*Elaboration of domain models in ontologies: the onomasiological approach and the function of the definition*) discutem um aspecto específico, mas importante, do *design*, a saber, reunir conceitos e termos e escrever boas definições.

Mendonça (*Bulding ontologies through the OntoForInfoScience Methodology: a detailed approach to perform activities of the ontology development*) e Soares, Farinelli e Elkin (*Ontology building in practice: a case study applied to the obstetric domain*) lidam com o processo de desenvolvimento de ontologias.

A avaliação de ontologias é o tópico de Hicks (*Metrics and methods for comparative ontology evaluation*). Este artigo faz pontos excelentes, mas algumas métricas são mais adequadas ao nível de afirmação do que no modelo de dados e os níveis de valor da entidade, onde o foco da avaliação deve ser. As bibliografias sobre a avaliação de modelos de dados e a avaliação de tesouros, classificações e sistemas semelhantes de organização de conhecimento têm informações úteis a serem oferecidas.

Os documentos restantes, exceto *Ontologies in the context of Information Science: a bibliometric study in Brazil*, lidam com várias aplicações de ontologias, por função - harmonizando dados de sensores (*Semantic Middleware for Industrial Sensors*), indexação (*Automatic indexing and ontologies: identification of convergent contributions in Information Science*) e por domínio, por exemplo, música (*Application and use of music ontologies*)

arquitetura, engenharia e construção civil (*Ontologies as support for information modeling in architecture, engineering and construction*), e, em nível filosófico profundo, organizações e corporações (*An ontological approach to the normative dimension of organizations: an application of Documents Acts Ontology*).

Figueiredo e Almeida relatam uma análise bibliográfica da literatura sobre ontologias no Brasil. Este é um bom ponto de partida, mas qualquer análise é limitada pela(s) fonte(s) coberta(s) e os termos usados para pesquisa. A unificação do pensamento e os resultados de vários campos exigidos no próximo parágrafo podem ser apoiados ao ampliar esse trabalho, usando um grupo muito mais amplo de fontes e uma abordagem mais ampla para a seleção de termos de pesquisa.

É bem-vinda que esta edição combine contribuições da *ciência da informação* (um rótulo comumente usado para a comunidade melhor representada por organizações como a Associação de Ciência da Informação e Tecnologia (ASIST), a Associação Internacional para a Organização do Conhecimento (ISKO) e, para alguns extensão, organização da biblioteca) e ontologia (um rótulo comumente usado para a comunidade melhor representada pela Associação Internacional para Ontologia e suas Aplicações (IAOA) e a Cúpula de Ontologia). Essas comunidades lidam com os mesmos problemas — de uma perspectiva diferente talvez, com base em diferentes tradições científicas, usando diferentes terminologias e diferentes métodos —, mas os mesmos problemas, no entanto. Como esta edição demonstra, há um crescente reconhecimento dessa comunidade de interesses e que unir forças seria melhor para fazer progressos do que trabalhar em universos paralelos e separados.

Uma colaboração muito mais estreita entre OntoBras, ISKO Brasil e o ASIST Brasil na organização de conferências e em outras aplicações e uma colaboração mais estreita entre professores universitários de ambas as comunidades (muitas vezes na ciência da informação de fronteira —

ciência da computação — seria benéfico para ambos os lados. São duas comunidades mais que devem estar envolvidas: modelagem de dados e linguística.

Os linguistas desenvolveram suas próprias ontologias de nível superior (não muito diferentes da BFO, por exemplo) para o propósito de explicar a estrutura e uso da linguagem. Unificar todas essas contribuições pode exigir modelagem de dados, ontologias e outros sistemas e esforços de organização do conhecimento comparáveis em alcance e visão ao que o grupo conhecido como Bourbaki fez para a matemática

Presentación

Dagobert Soergel

PhD, Universidad de Maryland

Lo que sigue son algunos comentarios sobre los temas tratados en este documento, que culminan en un llamado a una colaboración más fuerte entre las comunidades que se ocupan de cuestiones de ontologías y otros sistemas de organización del conocimiento y un esfuerzo hacia la unificación del pensamiento y los resultados logrados por estas comunidades.

Comienzo por recordar algunas distinciones para contextualizar los documentos sobre este tema. Estas distinciones son pragmáticas y útiles, pero pueden ser, y han sido, debatidas interminablemente, a veces por motivos lógicos. La mayoría de estos debates los considero decididamente no útiles.

Al tratar con las ontologías y su uso en la representación de datos, existen tres niveles:

Los documentos en este número se centran en Entity Value Ontologies. Uno tiene la impresión de que para algunos autores lo que hace que un sistema digno del nombre "ontología" dependa no tanto del contenido o del nivel (en el sentido presentado en la tabla), sino más bien del grado de formalidad, como expresándose en RDF, OWL o algún formato de lógica más expresiva. Esta postura debe ser revisada

lo importante es el contenido que un sistema tiene para ofrecer y la precisión con que se expresa ese contenido (las cosas se pueden expresar con mucha precisión en el lenguaje natural).

Tres niveles en representación de datos		
Ontologías de modelo de datos	Defina tipos de entidades y tipos de relación. Los tipos de entidades son clases, pero no todas las clases son tipos de entidad. Por ejemplo, el taxón es un tipo de entidad y también una clase los miembros de la clase, o los valores de la entidad del tipo de entidad, son tasa, como Rosa multiflora La rosa multiflora es, a su vez, una clase cuyos miembros son especímenes de plantas individuales. Ejemplos: Ontología de Formato Básico (BFO, incluyendo relaciones BFO), Ontología Fundamental Unificada (UFO), FOAF, Dublin Core	Tbox Axiomas terminológicos
Ontologías de valor de entidad	Estos sistemas se ocupan de valores de entidades - taxa biológica, partes del cuerpo y sistemas corporales (anatomía más), genes, funciones biológicas, enfermedades y trastornos, lugares, períodos históricos, culturas, estilos artísticos, objetos culturales, conceptos - y términos usados para nombrarlos. Estos sistemas también se conocen como esquemas de clasificación, taxonomías, listas de títulos de asunto o tesauros, conocidos colectivamente como sistemas de organización del conocimiento. Los sistemas que utilizan el término "ontología" en su nombre generalmente poseen una estructura más elaborada con relaciones diferenciadas. Los siguientes son los siguientes: La Clasificación Decimal Universal (UDC), la Ontología de Género, (GO) SnoMed, la Clasificación Internacional de Enfermedades (ICD), las Secciones de Asunto Médico (MeSH), AGROVOC, el Tesoro de Términos de Índice Psicológico, los vocabularios de Getty, incluyendo el Thesaurus de Arte y Arquitectura (AAT).	
Conjuntos de datos, conjuntos de archivos, bases de datos	Los conjuntos de datos se interpretan aquí como colecciones de afirmaciones sobre la realidad (un término usado aquí ingenuamente sin tomar ninguna posición filosófica) o representando pensamiento o imaginación y sueños. Las afirmaciones se forman a partir de tipos de relaciones que rellenan valores de entidad para los argumentos de relación. En su forma más simple, las afirmaciones asumen la forma de triples, pero representar la complejidad de la realidad y del pensamiento requiere relaciones multi-argumentos. Los ejemplos en el artículo de Hick sobre la evaluación de la ontología (Paper I-2) son principalmente afirmativos, indicando que considera que los conjuntos de datos como ontologías, al menos, se presentan de forma formal.	Abox Axiomas asertivos
Límites entre niveles son porosos	Los límites entre estos niveles no son estrictos. Por ejemplo, los tipos semánticos del sistema de lenguaje médico unificado (UMLS) enumeran animales (una clase) e incluso mamíferos y humanos como un tipo de entidad, mientras que el animal es un valor del organismo de tipo de entidad, aunque alto en jerarquía de estos valores de la entidad. Muchos tesauros, como MeSH o el Thaurus de alcohol y otros medicamentos, incluyen relaciones que son aserciones sobre la realidad. Por ejemplo, en la jerarquía MeSH existe un título de agentes antiinfecciosos Listar una sustancia química bajo este título es una afirmación sobre la realidad. MeSH incluye tales relaciones porque son útiles para la investigación. En una visión amplia, los tres niveles podrían combinarse en una gran base de datos sin preocuparse de qué nivel pertenece un elemento. Por otro lado, la distinción entre los niveles es útil. Los sistemas deben decidir pragmáticamente qué información incluir en qué nivel	

Los trabajos discuten *design* e estrutura, desenvolvimento (estritamente ligado ao design), avaliação e aplicação de ontologias, amplamente interpretadas.

Lima y Maculan (*Comparative study of semantic structures in different knowledge organization systems*) y Maculan y Aganette (*Ontologia aplicada a um modelo de gestão organizacional: contribuições da Ciência da Informação*) discuten la estructura relacional de ontologias y otros sistemas de organización del conocimiento. Seppälä, Ruttenberg y Smith (*Guidelines for Writing Definitions in Ontologies*), así como Maria Campos (*Elaboration of domain models in ontologies: the onomasiological approach and the function of the definition*) discuten aspecto específico, pero importante, del *design*, a saber, reunir conceptos y términos y escribir buenas definiciones.

Mendonça (*Building ontologies through the OntoForInfoScience Methodology: a detailed approach to perform activities of the ontology development*) y Soares, Farinelli y Elkin (*Ontology building in practice: a case study applied to the obstetric domain*) se ocupan con el proceso de desarrollo de ontologias.

La evaluación de ontologias es el asunto de Hicks (*Metrics and methods for comparative ontology evaluation*). Este artículo levanta excelentes puntos, pero algunas métricas son más adecuadas al nivel de afirmación que en el modelo de datos y los niveles de valor de entidad, donde debe ser el foco de la evaluación. Las bibliografías sobre evaluación de modelos de datos y evaluación de tesauros, clasificaciones y sistemas semejantes de organización del conocimiento tienen informaciones útiles a ofrecer.

Los documentos restantes, excepto *Ontologies in the context of Information Science: a bibliometric study in Brazil*, tratan de varias aplicaciones de ontologias, por función — armonizando datos de sensores (*Semantic Middleware for Industrial Sensors*), indexación (*Automatic indexing and ontologies: identification of convergent contributions in Information Science*) y por dominio, por ejemplo, música (*Application and use of music ontologies*)

arquitectura, ingeniería y construcción civil (*Ontologies as support for information modeling in architecture, engineering and construction*), y, en nivel filosófico profundo, organizações e corporações (*An ontological approach to the normative dimension of organizations: an application of Documents Acts Ontology*).

Figueiredo y Almeida relatan análisis bibliográfico de la literatura sobre ontologias en Brasil. Este es un buen punto de partida, pero cualquier análisis es limitado por la(s) fuente(s) cubierta(s) y los términos utilizados para investigación. La unificación del pensamiento y los resultados de varios campos exigidos en el siguiente párrafo pueden ser apoyados al ampliar el trabajo, usando un grupo mucho más amplio de fuentes y un enfoque más amplio para la selección de términos de investigación.

Es bienvenida que esta edición combine contribuciones de la *ciencia de la información* (rótulo comúnmente usado para la comunidad mejor representada por organizaciones como la Asociación de Ciencia de la Información y Tecnología (ASIST), la Asociación Internacional para la Organización del Conocimiento (ISKO) y, para algunos extensión, organización de la biblioteca) y ontología (rótulo comúnmente usado para la comunidad mejor representada por la Asociación Internacional para Ontología y sus Aplicaciones (IAOA) y la Cúpula de Ontología). Esas comunidades se ocupan con los mismos problemas — de una perspectiva diferente talvez, con base en diferentes tradiciones científicas, usando diferentes terminologias y diferentes métodos —, pero los mismos problemas, entretanto. Como esta edición demuestra, hay creciente reconocimiento de esa comunidad de intereses y que unir fuerzas seria mejor para hacer progresos que trabajar en universos paralelos y separados.

Una colaboración mucho más estrecha entre OntoBras, ISKO Brasil y el ASIST Brasil en la organización de conferencias y en otras aplicaciones una colaboración más estrecha entre profesores universitarios de ambas comunidades

(muchas veces en la ciencia de la información de frontera — ciência de la computación — sería benéfico para ambos lados. Son dos comunidades más que deben estar involucradas: modelaje de datos y lingüística.

Los lingüistas desarrollaron sus propias ontologías de nivel superior (no muy distintas de la BFO, por ejemplo) con el propósito de explicar la estructura y uso del lenguaje. Unificar todas esas contribuciones puede exigir modelaje de datos, ontologías y otros sistemas y esfuerzos de organización del conocimiento comparables en alcance y visión a lo que el grupo conocido como Bourbaki hizo para la matemática.

EDITORIAL

Computadores necessitam de modelos para manipular representações das coisas do mundo, ou seja, coisas com as quais as pessoas lidam diariamente, como árvores, bactérias, carros, instituições, livros, loterias, governos, outras pessoas, para citar alguns poucos. Esses modelos, chamados “ontologias”, fornecem esquemas que permitem aos computadores suportar inferências lógicas, isto é, raciocínio automático.

De fato, ontologias são representações formais e rigorosas das coisas do mundo e das relações entre elas. Essa é a abordagem da disciplina Ontologia Aplicada, na qual a ênfase recai na representação das coisas em si e não nas palavras utilizadas para descrever tais coisas.

A ênfase em representar as coisas e não a complexidade da linguagem humana faz com que as ontologias sejam um instrumento adotado em todo o mundo para a integração básica entre sistemas de informação, uma alternativa para solucionar o problema em geral conhecido como “falta de interoperabilidade”. Nos dias de hoje, a Ontologia Aplicada é um campo de pesquisa frutífero que estuda as entidades da realidade e as teorias que suportam a construção de representações compartilhadas de domínios do conhecimento.

Existe grande interesse na aplicação de ontologias para resolver problemas de modelagem e classificação em uma variedade de áreas, tais como ciência da computação, ciência da informação, filosofia, inteligência artificial, linguística, gestão do conhecimento e muitas outras. Este número especial da revista *Ciência da Informação* configura-se em uma oportunidade para que os pesquisadores em ontologias apresentem e discutam teorias, metodologias, linguagens e experiências relativas ao desenvolvimento e aplicação de ontologias. Especialistas do campo da ciência da informação foram convidados a submeter contribuições de pesquisa teóricas e práticas.

Assim, este número especial contém artigos nacionais e internacionais, representativos da pesquisa no Brasil e no mundo, organizados em duas seções: i) contribuições teóricas e metodológicas da pesquisa em ontologias

ii) aplicações e usos da ontologia em ciência da informação. Este número contém então 14 artigos completos, selecionados por um comitê de avaliação também composto por membros nacionais e internacionais. Agradecemos ao IBICT e à equipe da revista científica *Ciência da Informação* pela oportunidade, aos autores pelas suas contribuições, e aos revisores por seu trabalho.

Mauricio B. Almeida

Editor Científico do número especial
Ontologias em Ciência da Informação

Doutor em ciência da informação

Professor Associado, Programa de Pós-Graduação em
Gestão & Organização do Conhecimento
Universidade Federal de Minas Gerais

EDITORIAL

Pesquisador bolsista CNPQ
Computers need models to deal with representations of things of the world, that is, things with which people live daily, such as trees, bacteria, cars, institutions, books, lotteries, governments, other people, their professions, to name but a few. These models, called “ontologies”, provide schemas to computers enabling them to support logical inferences, that is, automatic reasoning.

Ontologies are formal and rigorous representations of things of the world and relations between them. This is the approach of the Applied Ontology discipline in which the emphasis rests on the representation of things and not on the words used to describe the very same things.

The emphasis on representing things, rather than the complexity of human language, makes ontologies largely used around the world promoting integration among information systems, as an alternative to an issue in information systems commonly called “lack of interoperability”. Nowadays, Applied Ontology is a fruitful research field concerning the study of entities of reality and theories that support the building of shared representations of knowledge domains.

A growing interest in the application of ontologies to solve modeling and classification problems can be observed in several different areas such as Computer Science, Information Science, Philosophy, Artificial Intelligence, Linguistics, Knowledge Management and many others. This special number of Instituto Brasileiro de Ciência e Tecnologia (IBICT) is an opportunity to researchers on ontologies present and discuss theories, methodologies, languages, tools and experiences related to the development and application of ontologies. Specialist researchers and practitioners from the Information Science field were invited to submit theoretical and practical research contributions. The special number contains papers, representative of the research in Brazil and abroad, which are organized in two sections, namely: i) theoretical and methodological contributions of research on ontologies

ii) applications and uses of ontologies in Information Science.

This volume is thus constituted of 14 full papers, selected by a review committee, which is composed by national and international referees. We thank the IBICT and the team of *Ciência da Informação Journal* for the opportunity, the authors for their valuable contributions, and the referees for their work.

Mauricio B. Almeida, PhD.

Scientific Editor of the special issue on
Ontologies in Information Science

Associate Professor – Graduate Program of
Management & Knowledge Organization
Federal University of Minas Gerais

Researcher – National Council of

EDITORIAL

Scientific Development (CNPq) Computadoras necesitan modelos para manipular representaciones de las cosas del mundo, es decir, cosas con las que las personas tratan diariamente, como árboles, bacterias, coches, instituciones, libros, loterías, gobiernos, otras personas, por citar algunos pocos. Estos modelos, llamados "ontologías", proporcionan esquemas que permiten a los equipos soportar inferencias lógicas, es decir, razonamiento automático.

De hecho, las ontologías son representaciones formales y rigurosas de las cosas del mundo y de las relaciones entre ellas. Este es el enfoque de la disciplina Ontología Aplicada, en la cual el énfasis recae en la representación de las cosas en sí y no en las palabras utilizadas para describir tales cosas.

El énfasis en representar las cosas y no la complejidad del lenguaje humano hace que las ontologías sean un instrumento adoptado en todo el mundo para la integración básica entre sistemas de información, una alternativa para solucionar el problema en general conocido como "falta de interoperabilidad". En los días de hoy, la Ontología Aplicada es un campo de investigación fructífero que estudia las entidades de la realidad y las teorías que soportan la construcción de representaciones compartidas de dominios del conocimiento.

Hay gran interés en la aplicación de ontologías para resolver problemas de modelado y clasificación en una variedad de áreas, tales como ciencia de la computación, ciencia de la información, filosofía, inteligencia artificial, lingüística, gestión del conocimiento y muchas otras. Este número especial de la revista Ciencia de la Información se configura en una oportunidad para que los investigadores en ontologías presenten y discutan teorías, metodologías, lenguajes y experiencias relativas al desarrollo y aplicación de ontologías. Los expertos del campo de la ciencia de la información fueron invitados a someter contribuciones de investigación teóricas y prácticas.

Así, este número especial contiene artículos nacionales e internacionales, representativos de la investigación en Brasil y en el mundo, organizados en dos secciones: i) contribuciones teóricas y metodológicas de la investigación en ontologías

ii) aplicaciones y usos de la ontología en ciencia de la información. Este número contiene entonces 14 artículos completos, seleccionados por un comité de evaluación también compuesto por miembros nacionales e internacionales. Agradecemos al IBICT y al equipo de la revista científica de la Ciencia de la Información por la oportunidad, a los autores por sus contribuciones, ya los revisores por su trabajo.

Mauricio B. Almeida

Editor Científico del número especial
Ontologías en Ciencia de la Información

PhD en ciencia de la información

Profesor Asociado, Programa de Postgrado en
Gestión y Organización del Conocimiento

Universidad Federal de Minas Gerais
Investigador bursátil CNPQ

Part I
Theoretical and methodological contributions
of research on ontologies

Parte I
Contribuições teóricas e metodológicas da pesquisa em ontologias

Parte I
Contribuciones teóricas y metodológicas de la investigación en ontologías

Ontologias em ciência da informação: um estudo bibliométrico no Brasil

Frederico de Carvalho Figueiredo

Doutor em Gestão Urbana pela Pontifícia Universidade Católica do Paraná (PUC/PR) - Curitiba, PR – Brasil. Professor do Centro Universitário UNA (UNA) - Belo Horizonte, MG - Brasil. Professor e pesquisador do Centro Universitário de Belo Horizonte (UniBH) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/8877626854487189>

E-mail: frederico_miarelli@hotmail.com

Fernanda Gomes Almeida

Doutoranda em Ciência da Informação pela Universidade Federal de Minas Gerais (UFMG) - Brasil.

Doutoranda em Gestão e Organização do Conhecimento Universidade Federal de Minas Gerais (UFMG)

- Brasil. Mestre em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG) - Brasil.

Bibliotecária pela Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/5601300780102290>

E-mail: nandaalmeida@ufmg.br

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

RESUMO

O estudo de ontologias como ferramenta de organização do conhecimento vem despertando interesse da academia. Cada vez mais publicações sobre o tema vem sendo produzidas, especialmente no âmbito da web semântica. O presente trabalho busca compreender o contexto de produção científica sobre ontologias no Brasil, analisando artigos científicos indexados pela Brapci – Base de Dados em Ciência da Informação – Acervo de Publicações Brasileiras em Ciência da Informação, mantido pela Universidade Federal do Paraná. A análise foi realizada por meio de um estudo bibliométrico que determinou os autores de maior relevância para a temática no Brasil, as principais relações entre os autores e instituições, bem como a evolução da produção sobre ontologia nas últimas décadas. Os resultados apontam a necessidade da criação de vínculos entre os diversos polos de estudo sobre ontologias no Brasil, com a finalidade de ampliar o relacionamento dos autores e instituições bem como a troca de experiências e conhecimento. Ainda demonstram o aumento de artigos na temática desde 2001 em razão das relações entre ontologias, web semântica e ciência da informação, bem como do crescente interesse da tecnologia da informação. Por fim, é possível verificar, pelas palavras-chave dos artigos, a interdisciplinaridade e a influência da filosofia e tecnologia da informação nos trabalhos ontológicos na ciência da informação.

Palavras-chave: Bibliometria. Ontologias. Web semântica. Redes de coautores. BRAPCI.

Ontologies in information science: a bibliometric study in Brazil

ABSTRACT

Studies on ontologies as tools for knowledge organization have been drawing the attention of the academy. More and more works on the subject are produced, especially in the scope of the semantic web. Thus, this paper aims to understand the context of the Brazilian scientific production on ontologies, analyzing the scientific works indexed by BRAPCI – Information Science database – Collection of Brazilian Publications on Information Science sponsored by the Federal University of Paraná. The analysis of the indexed papers was bibliometric, determining the most prominent authors in the thematic, the relations between authors and institutions, as well as the evolution of the academic production over the last decades. The results show that it would be important the collaboration among different ontologies research centers, thus broadening the relations among authors and institutions, and the exchange of perspectives and experiences. It also points the increase in the number of papers since 2001, possibly because of the relations between ontologies semantic web and Information Science. Finally, it is possible to verify, analyzing the keywords in the papers, the interdisciplinarity and the influence of the philosophical and technological views on the ontological works of the Information Science.

Keywords: *Bibliometry. Ontologies. Web semantic. Co-authors network. BRAPCI*

Ontología en ciencia de la información: un estudio bibliométrico en Brasil

RESUMEN

El estudio de ontologías como herramienta de organización del conocimiento ha despertado el interés de la academia internacional. Cada vez más trabajos sobre el tema son producidos, sobre todo en el contexto de la web semántica. Así, este trabajo intenta comprender el contexto de la producción científica acerca de ontologías en Brasil, mediante el análisis de artículos científicos indexados por la BRAPCI - Base de Datos en Ciencia de la Información - Acervo de Publicaciones Brasileñas en Ciencia de la Información, mantenido por la Universidad Federal de Paraná. El análisis fue realizado mediante un estudio bibliométrico que determinó cuáles son los autores de mayor relevancia para la temática en Brasil, las principales relaciones entre autores e instituciones y la evolución de la producción sobre ontología en las últimas décadas. Los resultados apuntan a la necesidad de establecer vínculos entre los diversos polos de estudio sobre ontologías en Brasil, con el fin de ampliar la relación entre autores e instituciones y el intercambio de experiencias y conocimientos. Demuestra aún el incremento del número de artículos cerca de la temática desde 2001, posiblemente a causa de las relaciones entre ontologías y web semántica y al creciente interés de la Tecnología de la Información. Al fim, se verifica a partir de las palabras-clave utilizadas en los artículos la interdisciplinaridad y la influencia de la filosofía y de la Tecnología de la Información en los trabajos ontológicos en la Ciencia de la Información.

Palabras clave: *Bibliometria. Ontologías. Web semantica. Red de co-autores. BRAPCI*

INTRODUÇÃO

Com a necessidade de interoperabilidade sistêmica e o surgimento da web semântica, o campo das ontologias vem ganhando importância como temática de pesquisa dentro da ciência da informação. Em razão de sua forte ligação com axiomas filosóficos de base, permite uma fiel representação de determinado domínio do conhecimento, da forma pela qual aquele ramo da ciência o define e demonstra sua existência.

O crescimento do estudo de ontologias no contexto da ciência da informação pode ser apresentado como resultado da importância que o tema vem exercendo no contexto de tal domínio nos últimos anos. Estudos anteriores (por exemplo, ZHU *et al*, 2015 e DING, 2010 – o último incluindo as ontologias dentro do contexto da web semântica) apontam o crescimento da produção científica sobre ontologias desde 1990. Zhu *et al* (2015) afirmam que até tal ano os conceitos relacionados à concepção filosófica da ontologia foram introduzidos na ciência da informação. Nesse estudo (ZHU *et al*, 2015), o Brasil ficou em décimo-quarto lugar na produção total de artigos sobre ontologia nas bases de dados Science Citation Index e Social Science Citation Index, acessados pela plataforma Web of Science.

Em face do contexto científico-geográfico apresentado, é importante questionar: quais as características da produção acadêmica brasileira sobre ontologias na ciência da informação? O presente trabalho busca compreender o contexto de produção científica sobre ontologias no Brasil. Especificamente, este artigo analisa os artigos científicos produzidos no Brasil e indexados pela Brapci – Base de Dados em Ciência da Informação – Acervo de Publicações Brasileiras em ciência da informação, mantido pela Universidade Federal do Paraná. Esta base de dados possui trabalhos científicos indexados na área de ciência da informação a partir de 1972.

A análise foi realizada por meio de um estudo bibliométrico, que apontou, além dos autores mais produtivos sobre o tema, as principais relações entre os autores e a evolução da produção sobre ontologia nas últimas décadas.

Após a introdução, a segunda parte apresenta conceitos básicos utilizados na presente pesquisa, em especial ontologia e bibliometria. Na terceira parte do trabalho são elencados os métodos utilizados para a realização da pesquisa. Na quarta parte, são analisados os dados obtidos e apresentada a discussão sobre os resultados. Por fim, são apresentadas considerações finais.

BACKGROUND

Esta seção visa apresentar conceitos necessários para a operacionalização da pesquisa. Assim, é discutida a definição – ou as definições – de ontologia na ciência da informação, com a finalidade de orientar os trabalhos de coleta de dados. As diferentes visões e concepções sobre ontologias foram utilizadas também para subsidiar a análise dos dados.

Além disso, é apresentado o conceito de bibliometria, bem como sua importância não só no contexto da ciência da informação, mas como ferramenta de análise da produção científica de um campo científico.

ONTOLOGIA

Em sua concepção original, ontologia pode ser definida como um ramo da filosofia que tem por objetivo estudar o ser. Assim, busca determinar as propriedades principais dos seres, assim como suas relações, para então classificá-los (SMITH, 2004; ALMEIDA, 2013; FONSECA, 2007).

Em trabalhos recentes, pode-se constatar a utilização do termo ontologia como referência a artefatos de engenharia de *software*, na área de tecnologia da informação. ao mesmo tempo, a ciência da informação também se apropria de tal conceito, tratando-o como um sistema de organização do conhecimento (ALMEIDA, 2013). Verifica-se, então, aplicações da teoria ontológico-filosófica que são adequadas aos diversos campos nos quais é inserida (GUARINO, 1998).

Guarino e Giaretta (1995) apontam a diversidade que o termo ontologia pode ostentar, trazendo sete diferentes possibilidades de utilização, que variam

desde a disciplina filosófica até vocabulário lógico e sistema conceitual. Visando a sistematização das apresentações desse conceito, os autores apresentam uma diferenciação essencial: Ontologia (com o maiúsculo), como a disciplina filosófica; e ontologias (com o minúsculo, plural) como objetos relacionados aos níveis semânticos ou sintáticos – desde sistemas até vocabulários.

Fonseca (2007), no entanto, afirma que podem ser localizados trabalhos sobre ontologias para sistemas de informação e ontologias de sistemas de informação. Assim, diferenciam os dois aspectos, sendo aquele para apoiar a criação de ferramentas de modelagem, enquanto este apoia a criação de sistemas de informação direcionados a ontologias.

A variedade de visões sobre ontologia tanto, demonstra a amplitude de domínios aos quais pode-se aplicá-la, quanto por outro, demonstra um caráter interdisciplinar do estudo sobre ontologia que deve ser levado em consideração quando da análise do tema. Tais visões foram reconhecidas em Zhu *et al* (2015) ao analisar a produção científica internacional sobre ontologias até 2012. Os pesquisadores, inclusive, apontam a mudança de rumo das pesquisas sobre ontologia do viés filosófico para as suas aplicações na ciência da informação.

Com o foco na utilização de ontologias na ciência da informação, Roa, Sadiq e Indulska (2014) demonstram a função de organização do conhecimento por elas realizada:

ontologias são usadas em diferentes campos da administração, ciência e engenharia para modelar fenômenos, situações ou domínios de interesse. São uma base para a transferência de conhecimento, desenvolvendo o entendimento de um domínio, ou estimulando o objeto que representam (ROA; SADIQ; INDULSKA, 2014, p.2, tradução nossa)

A utilização de ontologias como sistemas de organização do conhecimento possui diversas aplicações, como, por exemplo, maior facilidade da interação homem-máquina, interoperabilidade entre sistemas computacionais e a melhoria de especificações na recuperação de informação, bem como a confiabilidade e reusabilidade de sistemas (ROA; SADIQ; INDULSKA, 2014).

Para permitir a inter-relação, são criadas ontologias de alto nível. As ontologias de alto nível podem ser definidas como instrumentos para integrar conhecimento heterogêneo que podem ser extraídos de áreas diversas (MASCARDI; CORDÌ; ROSSO, 2007). Tal conhecimento se vincula em um nível superior, permitindo a relação de conteúdos diversos em nível ontológico. Algumas ontologias de alto nível vêm sendo desenvolvidas fornecendo-se referências para trabalhos ontológicos que visem, por exemplo, interoperabilidade sistêmica. Entre elas, pode-se mencionar a Basic Formal Ontology (BFO); Descriptive Ontology form Linguistic and Cognitive Engineering (DOLCE); Cyc; General Formal Ontology (GFO); entre outras (MASCARDI; CORDÌ; ROSSO, 2007).

Tal amplitude de finalidades, bem como os benefícios que proporciona, permitem explicar o aumento do número de pesquisas sobre ontologias realizadas ao redor do mundo (ZHU *et al*, 2015).

BIBLIOMETRIA

A bibliometria analisa informações quantitativas sobre os diversos formatos bibliográficos, bem como avalia dados quantitativos de textos, autores e as citações entre eles. A utilização de ferramentas bibliométricas permite ao pesquisador conhecer melhor certo campo científico. Com tal tratamento das informações sobre determinada área acadêmica, é possível definir, entre outras características, as temáticas mais estudadas e eventuais mudanças nos rumos das pesquisas. Braga (1973) apresenta a seguinte afirmação, definindo bibliometria:

Entre as leis que contribuíram para o desenvolvimento da Ciência da Informação estão as grupadas pelo termo Bibliometria e que indicam o tratamento quantitativo da informação; entre as pesquisas bibliométricas estão as relacionadas ao estudo de citações bibliográficas, que possuem padrões de comportamento, obedecendo a determinadas leis. (BRAGA, 1973, p.9).

Mueller (2008) afirma que as avaliações realizadas sobre a produção científica são realizadas com base nas publicações que são resultado das pesquisas, que, por sua vez, certificam e divulgam os resultados apresentados.

Para que tal avaliação seja realizada, são necessários indicadores e métricas para a análise da produção científica.

Sobre a utilização de aspectos quantitativos para a análise da produção científica, Price (1976) declara a relevância de tais pesquisas para o desenvolvimento científico:

Deixando de lado os julgamentos de valor, parece clara a importância de se dispor de uma distribuição que nos informe sobre o número de autores, trabalhos, países ou revistas que existem em cada categoria de produtividade, utilidade ou o que mais desejarmos saber. (PRICE, 1976, p. 39).

Uma evolução dos estudos bibliométricos trouxe uma visão que permite relacionar nos métodos quantitativos que lhe são peculiares, e utilizados desde os seus primórdios (cujos exemplos mais relevantes são os modelos matemáticos, epidemiológicos e entrópicos da informação) e ferramentas provenientes das ciências sociais (como análises históricas, de redes sociais e políticas) (CARDOSO, 1996; ARAÚJO, 2006).

METODOLOGIA

A metodologia utilizada neste trabalho é de caráter exploratório, visando a análise da situação atual do estudo sobre ontologias no escopo da ciência da informação.

Como fonte dos trabalhos consultados por esta pesquisa foi escolhida a Brapci – Base de Dados em Ciência da Informação – Acervo de Publicações Brasileiras em Ciência da Informação. Tal base é a maior do Brasil em conteúdo no campo da ciência da informação, e está em funcionamento desde 1972.

Esta base de dados engloba periódicos e eventos científicos, abrangendo, conforme informações de seu *website*¹, 14.221 autores, 17.275 trabalhos em revistas científicas e 2.243 trabalhos publicados em anais de eventos acadêmicos. Além disso, permite a busca pelos campos autor, título, palavra-chave e resumo.

Salienta-se que a base de dados possui 22.367 palavras-chave em português, 16.956 palavras-chave em inglês e números pouco expressivos de palavras-chave em outras línguas. A base de dados escolhida indexa trabalhos desde 1972 até o presente.

Como critério de inclusão de artigos foi definido a utilização da palavra ontologia ou derivadas dela no título, palavras-chave e resumo do trabalho. Assim, através da busca on-site, foram utilizados os termos de busca “ontologia”, “ontologias”, “ontology”, “ontologies”, “ontológico”, “ontológica”, e “ontologic”.

Como critério de exclusão foi utilizado, inicialmente, a avaliação do escopo do trabalho, e sua inserção no âmbito da ciência da informação, pela análise dos resumos dos artigos. Além disso, foram excluídos trabalhos apresentados em congressos, artigos em duplicidade, resenhas, editoriais, resumos de teses e dissertações, traduções e registros que continham as palavras odontologia, gerontologia, deontologia e paleontologia.

A busca na base de dados retornou 278 trabalhos, sem delimitar um período temporal, ou seja, pesquisando-se por todo o período disponível. Desses, 105 estavam entre aqueles excluídos da pesquisa. Assim, analisou-se o total de 173 artigos científicos.

A análise dos dados foi realizada de diferentes maneiras, dependendo das informações buscadas. Após a tabulação dos dados dos trabalhos no Microsoft Excel, foram realizados acertos em razão de diferentes nomes do mesmo autor, relacionando-os com as suas instituições de vínculo.

Para a construção de redes entre autores e instituições, foi utilizado o software Pajek. As redes entre instituições foram construídas com artigos que apresentaram coautores de diferentes instituições, para evitar a poluição do sociograma com atores isolados. Do mesmo modo foi construído o sociograma das relações de coautoria.

A análise das palavras-chave dos artigos foi realizada a partir da frequência de sua utilização, com a ajuda do software Atlas.ti.

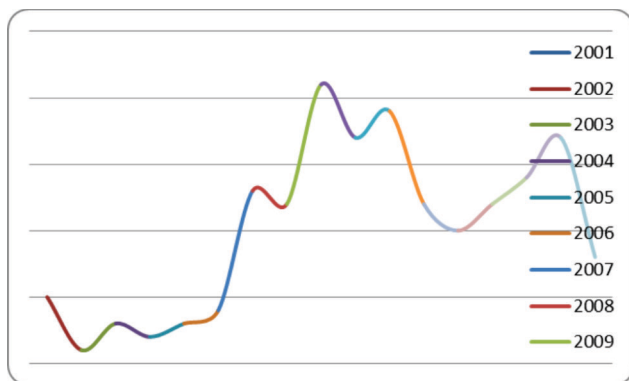
¹ <http://www.brapci.ufpr.br/brapci/>

ANÁLISE DOS DADOS

A primeira vez que o termo ontologia (e correlatos) aparece na base de dados analisada é no ano 2001. Tal dado pode ter influência da importância do tema na web semântica, conforme aponta Zhu *et al* (2015). Ressalta-se que, antes dos anos 2000, a construção do tema, conforme os autores, passa pelas etapas de esclarecimento (1909-1990) e crescimento (1990-2000).

Nota-se o crescimento no número de publicações a partir do ano de 2007, com pico no número de artigos publicados em 2009 (21 artigos). A evolução do número de publicações é ilustrada no gráfico 1. Tal informação também corrobora a divisão realizada por Zhu *et al* (2015), apontando que em 2007 ocorreu uma alteração nos anais da ciência da computação para a base ISI – e isto pode demonstrar aumento do interesse em ontologia pela ciência da computação.

Gráfico 1 – Artigos publicados por ano



Fonte: Os autores

Dentre os 260 autores que publicaram sobre o tema, Maurício Barcellos Almeida foi o autor mais produtivo com 16 artigos, seguido por Maria Luiza de Almeida Campos, com nove artigos publicados. A seguir são apresentados os autores com quatro ou mais artigos publicados (tabela 1). A indicação aponta a produção dos autores, independentemente de coautorias entre eles.

Tabela 1 – Autores mais produtivos

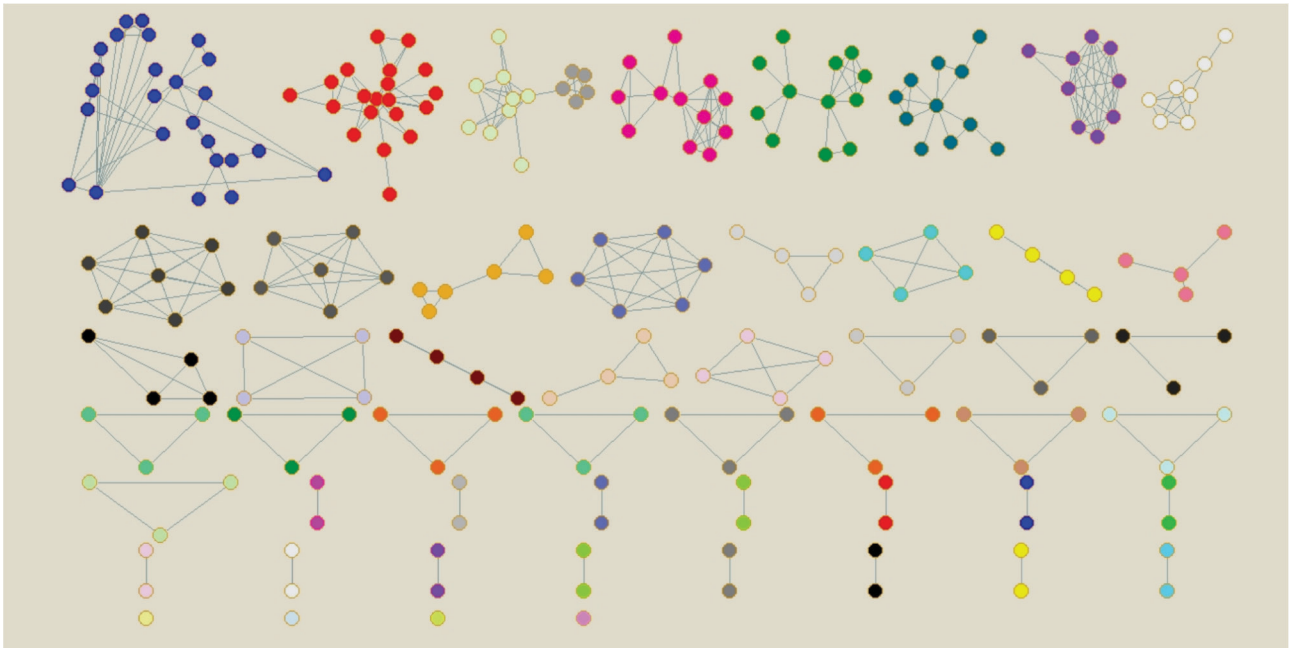
Autor	Instituição	Quantidade
ALMEIDA, Maurício Barcellos	UFMG, PUC Minas	16
CAMPOS, Maria Luiza de Almeida	UFF	9
BRÄSCHER, Marisa	UFSC, UNB	6
CAFÉ, Lígia	UFSC	6
TODESCO, José Leomar	UFSC	6
COELHO, Kátia Cardoso	UFMG	5
RAUTENBERG, Sandro	UNICENTRO	5
SCHIESSL, Marcelo	UNB, Caixa	5
SOUZA, Renato Rocha	FGV, UFMG	5
ALVARENGA, Lídia	UFMG	4
CAMPOS, Linair Maria	UFF, UFRJ	4
FACHIN, Gleisy Regina Bóries	UFSC	4
FERNEDA, Edberto	UNESP	4
FREITAS, Fred	UFPE	4
GOMES, Hagar Espanha	Consultoria, UFF	4
MARCONDES, Carlos Henrique	UFF	4
SANTAREM SEGUNDO, José Eduardo	UNESP, USP	4
STEIL, Andrea Valéria	UFSC	4

Fonte: Os autores

Apesar da quantidade e importância das publicações dos autores, o sociograma das relações entre coautores mostram claramente a autonomia e ausência de relações entre autores (figura 1). As redes de coautoria se encontram isoladas, demonstrando certo afastamento entre grupos de coautores:

Os periódicos que mais publicaram artigos sobre ontologia em ciência da informação podem ser importantes fontes de consulta. *Perspectivas em Ciência da Informação* foi o periódico que mais publicou artigos com a temática (25 publicações), conforme tabela 2:

Figura 1 – Redes de coautoria



Fonte: Os autores

Tabela 2 – Periódicos com mais publicações sobre o tema

Periódico	Frequência	%
Perspectivas em Ciência da Informação	25	14,45
RECIIS: Revista Eletrônica de Comunicação, Informação & Inovação em Saúde	16	9,25
DataGramaZero	15	8,67
Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação	15	8,67
Ciência da Informação	11	6,36
Informação & Informação	10	5,78
Informação & Sociedade: Estudos	10	5,78
Brazilian Journal of Information Studies: Research Trends	9	5,20
Tendências da Pesquisa Brasileira em Ciência da Informação	8	4,62
Transformação	6	3,47
Em Questão	5	2,89
InCID: Revista de Ciência da Informação e Documentação	5	2,89
Liinc em Revista	4	2,31
Perspectivas em Gestão & Conhecimento	4	2,31
Informação & Tecnologia	3	1,73
Informação em Pauta	3	1,73
PontodeAcesso	3	1,73

(Continua)

Tabela 2 – Periódicos com mais publicações sobre o tema (continuação)

Periódico	Frequência	%
Ágora	2	1,16
AtoZ: novas práticas em informação e conhecimento	2	1,16
Revista Brasileira de Biblioteconomia e Documentação	2	1,16
Revista Eletrônica Informação e Cognição	2	1,16
RICI: Revista Ibero-americana de Ciência da Informação	2	1,16
Arquivística.net	1	0,58
Arquivo & Administração	1	0,58
Biblionline	1	0,58
Biblos	1	0,58
Comunicação & Informação	1	0,58
ETD - Educação Temática Digital	1	0,58
Folha de Rosto	1	0,58
Informação Arquivística	1	0,58
Logeion: filosofia da informação	1	0,58
Revista de Informação, Memória e Tecnologia	1	0,58
Revista Digital de Biblioteconomia & Ciência da Informação	1	0,58
Total	173	100

Fonte: Os autores

As publicações sobre ontologia encontram-se distribuídas em 82 instituições, contudo concentram-se em seis instituições, representando 64,73% do total de artigos publicados, conforme apresentado na tabela 3. A Universidade Federal de Santa Catarina (UFSC) destacou-se como a instituição que acumula o maior volume de publicações (33 artigos), seguida de perto pela Universidade Federal de Minas Gerais (UFMG), com 29 artigos publicados.

Tabela 3 – Instituições com mais publicações sobre o tema

Instituição	Artigos
UFSC	33
UFMG	29
UNB	14
UFF	13
UNESP	13
UFPE	10

Fonte: Os autores (dados de pesquisa)

Com a utilização do *software* Pajek, foi construída a rede de relações entre as instituições dos autores que trabalharam juntos na produção de artigos sobre ontologia na ciência da informação.

As redes apresentam as relações entre as instituições cujos autores de trabalhos que constaram na amostra aqui utilizada apresentam vínculos de filiação. Assim, foi possível a representação de suas relações por meio de

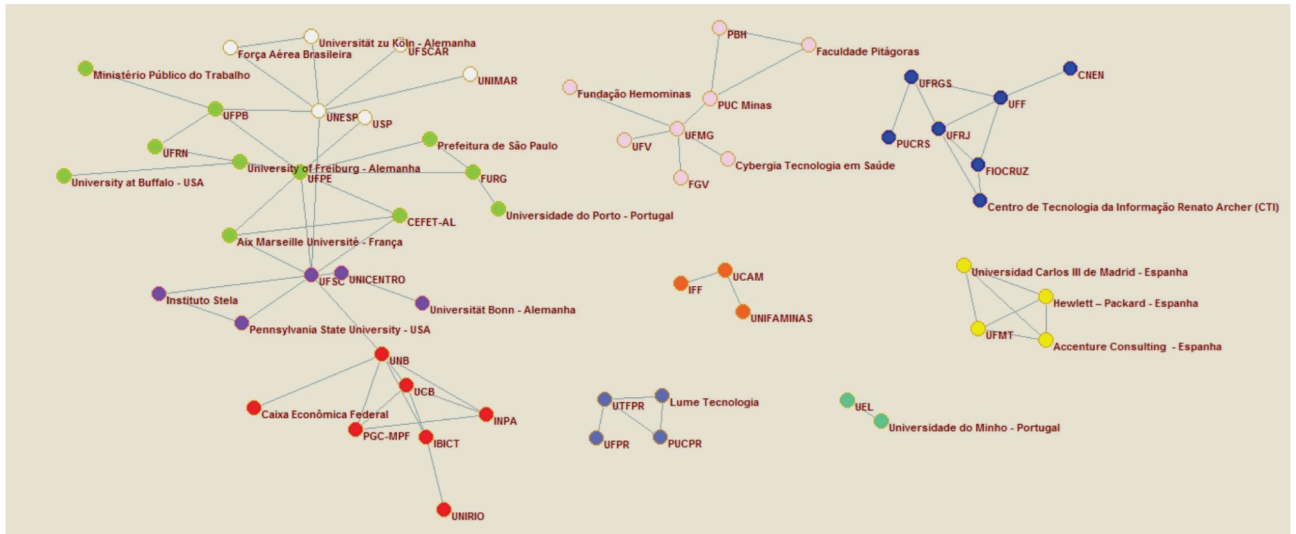
um sociograma (figura 2). Além disso, foi utilizada a ferramenta de determinação de comunidades do próprio *software*, por meio do algoritmo Leuvain.

A análise realizada demonstra a existência de alguns centros de produção de trabalhos sobre ontologia no Brasil, que incluem, inclusive, instituições do exterior. Foram apontadas, pelo método listado, dez conglomerados de instituições que se relacionam em tal produção. Enquanto alguns grupos se encontram isolados, outros apresentam relações que se apresentam como vínculos de instituições com diversos grupos.

Os dez grupos aqui apresentados produziram mais de um quarto dos artigos analisados com inter-relação entre instituições. O total de trabalhos produzidos pelos autores destas instituições, cooperando entre si, é de 53. Neste total estão excluídos os artigos produzidos sem cooperação institucional. Tais dados podem apresentar a possível construção de pólos de estudo sobre ontologia na ciência da informação.

A presença de instituições internacionais aponta no sentido da internacionalização das publicações, pelo relacionamento interinstitucional de entidades nacionais e internacionais. Das 82 instituições cujos autores publicaram artigos sobre ontologia na ciência da informação, 22 delas são instituições estrangeiras, como apresentado no quadro 1.

Figura 2 – Rede de relacionamento entre instituições



Fonte: Os autores

Quadro 1 – Instituições internacionais

Instituição	País
Stuttgart Media University	Alemanha
Universität Bonn	Alemanha
Universität Leipzig	Alemanha
Universität zu Köln	Alemanha
University of Freiburg	Alemanha
University of Stuttgart	Alemanha
Accenture Consulting	Espanha
Hewlett – Packard	Espanha
Universidad Carlos III de Madrid	Espanha
Universidad de Murcia	Espanha
Universidad de Zaragoza	Espanha
Aix Marseille Université	França
Universidade do Minho	Portugal
Universidade do Porto	Portugal
Royal Society of Canada	Canadá
Indiana University	USA
Pennsylvania State University	USA
University at Buffalo	USA
University of Wisconsin	USA
Universidad Nacional Autónoma de México	México
Universidad de La Habana	Cuba
Universidad de la República	Uruguai

Fonte: Os autores

Analisando-se as palavras-chave utilizadas em cada um dos trabalhos, alguns pontos de relevância podem ser levantados. A lista de ocorrências das palavras-chave mais utilizadas nos trabalhos é apresentada na tabela 4, excluídas aquelas sem valor semântico:

Tabela 4 – Palavras-chave

Palavra-chave	Ocorrências
Informação	126
Ontologia	110
Conhecimento	76
Ciência	45
Semantic	41
Web	37
Organização	34
Representação	31
Ontologias	28
Sistemas	26
Recuperação	20
Gestão	15
Científica	12
Classificação	12
Dados	11
Interoperabilidade	10
Modelos	10
Digital	9
Engenharia	9
Metadados	9
Pesquisa	9
Tecnologia	9
Tesouros	9
Comunicação	8
Epistemologia	8
Filosofia	8
Metodologia	8

Fonte: Os autores

Enquanto algumas palavras-chave são esperadas, com sua ocorrência não demonstrando tanta relevância (como, por exemplo, ontologia, ontologias, informação, conhecimento), outras apontam para caminhos trilhados pelos estudos sobre ontologias.

A importância da ontologia na web semântica pode ser demonstrada pela quantidade de vezes que tais palavras-chave aparecem nos artigos estudados. Semântica aparece 41 vezes, enquanto web se apresenta 37 vezes.

Outro ponto importante é a criação de modelos e a possibilidade de interoperabilidade sistêmica que o uso de ontologias pode proporcionar. Tais características das ontologias aparecem dez vezes cada uma nos artigos analisados.

Cabe ressaltar que as palavras-chave demonstram a interdisciplinaridade da temática. Os variados ramos científicos presentes apontam para a sua utilização em diversos domínios.

Por fim, conceitos relacionados com as diferentes definições de ontologia – sempre aplicados à ciência da informação – aparecem listados na tabela 4. A sua relação com tecnologia da informação, filosofia e ciência da informação pode ser demonstrada com o uso das palavras engenharia, digital, filosofia, epistemologia e tesouros.

CONSIDERAÇÕES FINAIS

A bibliometria é importante ferramenta para a determinação de características de um campo de estudo ou uma temática, bem como a evolução e os aspectos que são necessários serem trabalhados na área.

O presente trabalho buscou realizar uma análise exploratória sobre o estudo da temática da ontologia no contexto da ciência da informação no Brasil. Assim, com base nas informações disponibilizadas por uma base de dados brasileira do domínio apontado, foram coletados dados que indicassem para os caminhos da temática no contexto da ciência da informação.

Alguns aspectos são claros, como a necessidade de estudos na área para a possibilidade de interoperabilidade e construção de modelos de representação do conhecimento que possam ser um intermediário entre homem e máquina. Outro aspecto importante é a presença dos diferentes aspectos do estudo de ontologias (filosofia, ciência da informação e tecnologia da informação) como ferramental de desenvolvimento da área.

Apesar disso, os grupos formados no Brasil para produção de trabalhos científicos ainda carecem de relacionamentos, encontrando-se isolados. Tanto os autores como as instituições produtoras de trabalhos científicos necessitam aumentar seu inter-relacionamento para que o fluxo de ideias perpassa e receba contribuições de seus pesquisadores.

Alguns autores já se destacam como referências na área, com vasta produção acadêmica no país sobre o tema. A consolidação de redes de relacionamento pode incrementar a produção nacional sobre a temática.

Entre as limitações do presente trabalho, podem ser elencadas a base de dados utilizada, bem como os artigos incluídos pela leitora dos resumos. Além disso, as palavras utilizadas na pesquisa podem ser insuficientes para a busca sobre toda a temática.

Sugere-se, como trabalhos futuros, a pesquisa sobre a produção de pesquisadores brasileiros no exterior, em caráter complementar a esta pesquisa. A partir daí pode ser traçado o perfil das pesquisas brasileiras sobre ontologias.

REFERÊNCIAS

- 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, 2013. Disponível em: <<http://onlinelibrary.wiley.com/wol1/doi/10.1002/asi.22861/full>>. Acesso em: 16 ago. 2017.
- BRAGA, G. M. Relações bibliométricas entre a frente de pesquisa (research front) e revisões da literatura: estudo aplicado a Ciência da Informação. *Ciência da Informação*, v. 2, n. 1, p. 9-26, 1973. Disponível em: <<http://revista.ibict.br/ciinf/article/view/20/20>>. Acesso em: 16 ago. 2017.
- CARDOSO, A.M.P. Pós-Modernidade e informação: conceitos complementares?. *Perspectivas em Ciência da Informação*, v. 1, n. 1, p. 63-79, jan./jul. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/241/28>>. Acesso em: 16 ago. 2017.
- DING, Y. Semantic web: who is who in the field: a bibliometric analysis. *Journal of Information Science*, v. 36, n. 3, p. 335-356, 2010. Disponível em: <<http://journals.sagepub.com/doi/abs/10.1177/0165551510365295>>. Acesso em: 16 ago. 2017.
- FONSECA, F. The double role of ontologies in information science research. *Journal of the American Society for Information Science and Technology*, v. 58, n. 6, p. 786-793, 2007. Disponível em: <<http://onlinelibrary.wiley.com/wol1/doi/10.1002/asi.20565/full>>. Acesso em: 16 ago. 2017.
- GUARINO, N. Formal ontology and information systems. In: _____. (Ed.). *Formal ontology in information systems: Proceedings of First International Conference (FOIS'98)*, June 6-8, Trento, Italy. Amsterdam: IOS Press, 1998, p. 3-15. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.29.1776&rep=rep1&type=pdf>>. Acesso em: 22 jul. 2017.
- _____. GIARETTA, P. Ontologies and knowledge bases: towards a terminological clarification. In: MARS, N.J.I. (Ed.). *Towards very large knowledge bases: knowledge building & knowledge sharing*, 1995. Amsterdam: IOS Press, 1995. P. 25-32. Disponível em: <https://www.researchgate.net/publication/220041941_Ontologies_and_knowledge_bases_towards_a_terminological_clarification>. Acesso em: 22 jul. 2017.
- MASCARDI, V.; CORDÌ, V.; ROSSO, P. A comparison of upper ontologies. In: WORKSHOP FROM OBJECTS TO AGENTS (WOA), 8., 2007, Genova, Italy. *Electronic proceedings...* Genova, Italy: Seneca, 2007. p. 55-64. Disponível em: <<https://pdfs.semanticscholar.org/4f28/6fd9280449588b9d3781c9c897da28e0cff.pdf>>. Acesso em: 22 jul. 2017.
- MUELLER, S. P. M. Métricas para a ciência e tecnologia e o financiamento da pesquisa: algumas reflexões. *Encontros Bibli*, n. esp., 1º sem., p. 24-35, 2008. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13nesp1p24/1593>>. Acesso em: 22 jul. 2017.
- PRICE, D. J. S. *O desenvolvimento da ciência: análise histórica, filosófica, sociológica e econômica*. Rio de Janeiro: Livros Técnicos e Científicos, 1976.
- ROA, H. N.; SADIQ, S.; INDULSKA, M. Ontology usefulness in human tasks: seeking evidence. In: AUSTRALASIAN CONFERENCE ON INFORMATION SYSTEMS, 25., 2014, Auckland, New Zealand. *Electronic proceedings...* Auckland, New Zealand: ACIS, 2014. Disponível em: <http://aut.researchgateway.ac.nz/bitstream/handle/10292/8178/acis20140_submission_188.pdf?sequence=1>. Acesso em: 16 ago. 2017.
- SMITH, B. Ontology. In: FLORIDI, L. (Ed.). *Blackwell guide to the philosophy of computing and information*. Oxford: Blackwell, 2004. P. 155-166.
- ZHU, Q. et al. Global ontology research progress: a bibliometric analysis. *Aslib Journal of Information Management*, v. 67, n. 1, p. 27-54, 2015. Disponível em: <<http://www.emeraldinsight.com/doi/abs/10.1108/AJIM-05-2014-0061>>. Acesso em: 22 jul. 2017.

Metrics and methods for comparative ontology evaluation

Amanda Hicks

Ph.D, Philosophy, State University of New York at Buffalo
Assistant Professor, Health Outcomes & Policy Faculty,
Institute for Child Health Policy,
University of Florida, Gainesville, Florida, Estados Unidos.
<https://com-hop.sites.medinfo.ufl.edu/files/2015/06/CV-September2016.pdf>
E-mail: aehicks@ufl.edu

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

ABSTRACT

While progress has been made toward describing the need for ontology evaluation and offering proposals concerning what properties to measure and how, work remains to develop ontology evaluation as a rigorous discipline. Ontologies as information artifacts have a variety of aspects that can inform their evaluation, both in terms of what is evaluated and the metrics used. Ontology evaluation as a discipline requires (1) having a systematic account of the different aspects of ontologies and the properties relevant to those aspects, (2) critically developing methods for examining those properties, (3) developing comparative metrics that allow ontology engineers to compare the effects of various modeling choices and allow users to compare the merits of existing ontologies, and (4) charting possible pitfalls of evaluation. This paper considers various properties of ontologies that have been proposed and organizes these properties according to different aspects of ontologies. To begin bringing previous work together and to illustrate where pitfalls and potential solutions might enter into a rigorous evaluation, I offer a more in depth (though still partial) analysis of evaluating the correctness of ontologies. I conclude with a discussion of next steps in systematizing ontology evaluation.

Keywords: Ontologies. Ontologies Evaluation.

Métricas e métodos para a avaliação comparativa de ontologias

RESUMO

Progresso tem sido feito no sentido de descrever a necessidade de avaliação de ontologias, bem como nas propostas para mensurá-las e como mensurá-las, mas a literatura ainda carece de trabalhos sobre a avaliação de ontologias como uma disciplina rigorosa. Ontologias como artefatos de informação apresentam uma variedade de aspectos que podem fornecer subsídios para avaliação, tanto em termos do que avaliar como em termos das métricas adotadas. A avaliação de ontologias é uma disciplina que requer (1) consideração sistemática dos diferentes aspectos das ontologias e das propriedades que são relevantes para tais aspectos; (2) desenvolvimento crítico de métodos para examinar as propriedades mencionadas; (3) desenvolvimento de métricas que permitam aos engenheiros de ontologias comparar os efeitos das diversas decisões de modelagem possíveis, bem como a possibilidade do usuário comparar os méritos de ontologias existentes; (4) identificação das principais práticas que levam a erros de avaliação. O presente artigo considera as várias propriedades das ontologias que têm sido propostas e organiza tais propriedades de acordo com diferentes aspectos ontológicos. Inicia-se apresentando trabalhos relacionados anteriores, ilustrando as práticas que levam a erros, bem como soluções potenciais, para depois oferecer uma análise detalhada (ainda que parcial) da avaliação da correção de ontologias. Conclui-se com uma discussão sobre os próximos passos necessários para sistematizar a avaliação de ontologias.

Palavras-chave: Ontologias. Avaliação de ontologias.

Métricas y métodos para la evaluación comparativa de ontologías

RESUMEN

Progreso ha sido hecho en el sentido de describir la necesidad de evaluación de ontologías, como también en las propuestas para mensurarlas y como mensurarlas, pero la literatura aún carece de trabajos sobre la evaluación de ontologías como una disciplina rigurosa. Ontologías como artefactos de información presentan una variedad de aspectos que pueden proveer subsidios para la evaluación, tanto en términos de que evaluar como en términos de las métricas adoptadas. La evaluación de ontologías es una disciplina que requiere (1) consideración sistemática de los diferentes aspectos de las ontologías y de las propiedades que son relevantes para tales aspectos; (2) desarrollo crítico de métodos para examinar las propiedades mencionadas; (3) desarrollo de métricas que permitan a los ingenieros de ontologías comparar los efectos de las diversas decisiones de modelaje posibles, como también la posibilidad de que el usuario compare los méritos de ontologías existentes; (4) identificación de las principales prácticas que llevan a errores de evaluación. El presente artículo considera las varias propiedades de las ontologías que han sido propuestas y organiza tales propiedades de acuerdo con diferentes aspectos ontológicos. Se inicia presentando trabajos relacionados anteriores, ilustrando las prácticas que llevan a errores, como también soluciones potenciales, para ofrecer después un análisis detallado (aún que parcial) de la evaluación de la corrección de ontologías. Se concluye con una discusión sobre los próximos pasos necesarios para sistematizar la evaluación de ontologías.

Palabras-clave: Ontologías. Evaluación de ontologías.

INTRODUCTION

The call for a rigorous discipline of ontology evaluation is not new (GANGEMI et al., 2005; GANGEMI et al., 2006; GOMEZ-PEREZ, 2001; GUARINO, 2004; HOEHNDORF et al., 2007). While some progress has been made toward describing the need for ontology evaluation and offering proposals concerning what properties to measure and how, we still have work remains to develop ontology evaluation as a rigorous discipline.

Ontologies as information artifacts have a variety of aspects that can inform their evaluation, both in terms of what is evaluated and the metrics used. Ontology evaluation as a discipline requires (1) having a systematic account of the different aspects of ontologies and the properties relevant to those aspects, (2) critically developing methods for examining those properties, (3) developing comparative metrics that allow ontology engineers to compare the effects of various modeling choices and allow users to compare the merits of existing ontologies, and (4) charting possible pitfalls of evaluation.

The next section of this paper considers various properties of ontologies that have been proposed for evaluation – namely ontologies as representational

artifacts, logical theories, mathematical objects, parts of information systems, and community resources for reuse – and organizes these properties according to different aspects of ontologies. To begin bringing previous work together and to illustrate where pitfalls and potential solutions might enter into a rigorous evaluation, I offer a more in depth (though still partial) analysis of evaluating the correctness of ontologies. I conclude with a discussion of next steps in systematizing ontology evaluation.

THE SCOPE OF ONTOLOGIES UNDER CONSIDERATION AND SCOPE OF EVALUATION

While there is disagreement about what should be properly considered an ontology, an assumption of this paper is that, to the extent that this disagreement is fueled by normative considerations of what an ontology *should* be, ontology evaluation as a discipline should take an ecumenical view of the kinds of artifacts it evaluates to further inform the normative debate. Nevertheless, some constraints and a working definition of ontologies need to be offered as a starting point.

The kinds of artifacts under consideration in this paper have a human readable component, a machine readable component, and aim to formally represent something (whether material reality, concepts, linguistic intuitions, or something else). Ontologies are often created to be shared and reused by a community of informaticians, so a discipline of ontology evaluation will need to consider ontology reuse as well. However, for the purpose of this paper we do not consider this to be a necessary condition for an ontology since many of the considerations of evaluating an ontology would apply even to idiosyncratic, proprietary artifacts meant for use by a small number of users.

In this paper, the term ‘ontology’ does not refer exclusively to realist ontologies. This is not an assertion about whether conceptualist ontologies are as good as realist ontologies, nor is it a statement about whether wordnets can *function* as ontologies. Instead it is an agnostic stance intended to avoid begging the question about what makes a good ontology. Ultimately the discipline of ontology evaluation ought be able to meaningfully compare different types of resources, so arguments for adopting one or the other will not be based solely on appeals to common sense or conjecture, but rather will be grounded in some evaluative data to support the claim that one is preferable to another. We, therefore, adopt the definition offered in Neuhaus et al. (2013, p.):

Ontologies are human-intelligible and machine-interpretable representations of some portions and aspects of a domain, where the domain can be portions of the physical world or ways in which human agents mentally represent the physical world. That being said, we do believe that one should always be aware of which kind of ontology one is working with; a representation of concepts should not be mistaken for a representation of material reality.

Hoehndorf et al. (2013) argues that ontologies should always be evaluated as a part of an information system rather than as an ontology alone. I agree that a robust discipline of ontology evaluation must include evaluating ontologies as parts of information systems. To take this a step

further, such evaluation ideally ought to compare the performance of two or more ontologies in the same information system for the same task. However, I do not believe that ontologies cannot be evaluated according to their intrinsic properties. On the contrary, a complete science of ontology will be able to describe how different intrinsic properties of an ontology, including philosophical assumptions, affect the performance of information systems for particular tasks. Such an achievement requires analyzing and evaluating properties of ontologies, measuring their performance, and then synthesizing the results of both steps.

In designing ontology evaluation studies, we ought to have clear answers to the following:

1. What aspect of the ontology is under consideration?
2. What properties of the ontology, whether intrinsic or extrinsic, are being investigated?
3. What method will accurately and reliably capture this property?
4. What metrics can quantify that property, either directly or by proxy?
5. How can the metric be designed to yield accurate and reliable comparisons of evaluation results?

The next section provides a brief overview of 1 and 2 together.

VARIOUS ASPECTS AND PROPERTIES OF ONTOLOGIES

Ontologies are complex artifacts that can be considered from a variety of view points and disciplines. Accordingly, various approaches and proposals for ontology evaluation focus on different aspects of ontologies including ontologies as representational artifacts, as logical theories, as mathematical graphs, as parts of information systems, and as community resources for reuse. Each of these aspects are discussed and used to

organize properties of ontologies that have been proposed for evaluation.

As representational artifacts an ontology can be evaluated for the truth and accuracy of their representations, whether the allowable formal interpretations include all and only the intended interpretations, whether domain experts approve of the content, and breadth and granularity of domain coverage.

The logical system used in the ontology can be evaluated for validity, soundness, completeness, decidability (OBRST et al., 2007), syntactic lawfulness and richness (AMITH; TAO, 2017). The ontology as an axiomatic theory can be evaluated for logical properties such as consistency.

Ontologies that are graphs are mathematical objects, so their topological properties can be measured. Some of the topological metrics can be used as proxies for desirable attributes of ontologies, such as whether there is a single root node, multiple inheritance, depth, and fanned-outness (WALOSZEK, 2012). Waloszek (2012) offers an in depth discussion of this.

As parts of information systems ontologies can be evaluated against the requirements of the information system (NEUHAUS et al., 2013), use cases, their data sources (OBRST et al., 2007), and competency questions (GRÖNINGER; FOX, 1995). Ultimately, ontologies ought to be compared in terms of how they affect the overall performance of the system in a manner that isolates the contributions of various ontologies in the information system. This requires developing information systems in a manner that allows swapping out ontologies to compare results on the same task.

Ontologies as a community resource, including as controlled terminologies, can be evaluated according to their suitability for use and reuse. As such their ranking within a community can be measured and evaluated, including how many other ontologies reuse the current one (AMITH; TAO, 2017), how

frequently the ontology is used in information systems, and how successful it is for integrating data. As resources for use and reuse by humans, ontologies can be evaluated for the intelligibility of terms and definitions in the ontology (AMITH; TAO, 2017; OBRST et al., 2007), including what Amith and Tao (2017) calls “clarity” – the ratio of terms in the ontology that are ambiguous with respect to some source lexicon such as WordNet (MILLER, 1995) – “interpretability” or the ratio of terms in the ontology that have at least one word sense in some source lexicon (AMITH; TAO, 2017), the number duplicate terms in the ontology itself (AMITH; TAO, 2017), and the degree to which there is community consensus that the ontology contains the relevant classes with correct definitions (OBRST et al., 2007). The O² framework proposes usability-measures that quantify the number of annotations on the ontology according to a typology of annotations that promote reuse, e.g., recognition annotations, which describe an ontology’s structure and purpose (GANGEMI et al., 2005). This latter proposal is likely to play a large role in ensuring that ontologies conform to FAIR principles, which aim to make digital artifacts findable, accessible, interpretable, and reusable (WILKINSON et al., 2016).

In what follows, we focus on ontologies as representational artifacts and walk through considerations of measuring the correctness of the representation. We discuss methodological concerns, pitfalls, and solutions that ought to be addressed for a fully-fledged discipline of ontology evaluation.

THE CASE OF CORRECTNESS

When considering ontologies as representational artifacts, the natural question arises of whether they represent their subject matters well. A precise answer to this question requires a clearly defined notion of a good representation and some way of determining and measuring the quality of the representation. Previous discussions of ontology evaluation have dealt with the quality of the representation in terms of fidelity (NEUHAUS et al.,

2013), accuracy (AMITH; TAO, 2017), functional measures (GANGEMI et al., 2006), precision and recall of intended models (GUARINO, 2004), expert agreement (GANGEMI et al., 2005), and coverage (GANGEMI et al., 2005; ROSPOCHER et al., 2012; ZHU et al., 2009). What follows is a guided tour of some considerations that go into evaluating and comparatively measuring the correctness of ontologies.

Neuhaus *et al.* (2013) describes fidelity as a property of ontologies to be evaluated. Fidelity includes correctness of the statements in the ontology, both human readable and machine readable, but fidelity is also intended to capture whether the axioms and documentation are in agreement with each other. This latter criterion is not strictly about the quality of the representational aspect of an ontology, but rather about its usability and accessibility to humans. We certainly consider this important, but an orthogonal issue to the one described here, so in what follows the word ‘correctness’ will be used to describe the accuracy of the statements in the ontology and is distinguished from ‘fidelity’ which also describes agreement of the ontology with documentation.

The questions at hand for evaluating the correctness of the ontology include whether the human readable statements (e.g., definitions, examples, etc.) and axioms are correct. However, it is important to note that the standard of correctness is different for realist, conceptualist, and linguistic ontologies. For a realist ontology, this amounts to asking whether the statements are true statements about the world. For a conceptualist ontology, correctness means that representation corresponds to the conceptualization of a particular person or group, and for wordnets, correctness means whether the statements are readily agreed upon by native speakers of the language (MILLER; FELLBAUM, 1991). Evaluating the Descriptive Ontology for Cognitive and Linguistic Entities (MASOLO et al, 2002) according to a realist conception of truth would be as inappropriate as evaluating the Basic Formal Ontology (ARP et al, 2015) according to whether it conforms to a particular person’s

conception of the world. Evaluating the correctness of an ontology requires first understanding what the ontology is supposed to be a representation of.

SOME PREVIOUS SUGGESTIONS

How should the statements of an ontology be evaluated for correctness? And how can the result be quantified for comparative evaluation? It is widely acknowledged that evaluating ontologies for correctness requires domain experts (NEUHAUS et al., 2013; AMITH; TAO, 2017), which is often approached manually and is therefore both labor intensive and expensive. Neuhaus et al. (2013) suggests some automated methods that include checking for logical consistency, checking whether allowable formal models match intended models, and comparing the structure of the source ontology to some target ontology. Gordon et al. (2013) describe a semi-automatic method for constructing questions for expert review of the ontology discussed further below.

We consider checking for logical consistency is a task that evaluates ontologies as logical theories rather than as representations. For realist ontologies, logical consistency also evaluates the representation since it is a common assumption going back to Aristotle that reality is consistent, and so contradictions must be false. Since human agents often hold contradictory beliefs, a conceptualist ontology can be correct and logically inconsistent.

While allowable models can be automatically generated, it is not clear how these can be tested against intended models in an automated or even semi-automated way since a set of intended models needs to be constructed somehow and Neuhaus et al. (2013) does not describe how to construct that set. If we already had an ontology that was known to define all and only the intended models, the problems of evaluation and of creating a good ontology would already be solved.

Comparing the structure of two ontologies is an interesting proposal, but it is not clear what its value is. For the comparison to be meaningful,

we must have reason to believe that the target ontology is of sufficient quality to function as a gold-standard. If this is the case, it is not clear why one would not simply adopt this ontology rather than create a new one. Also, comparing the structure of ontologies presumably relies on at least some ontology mapping, but this itself is an area that is in need of methods for evaluation. Finally, if the ontologies have different ontological commitments, their correct representations may not be sufficiently isomorphic to produce a meaningful comparison. Without an authoritative digital source of knowledge that can be compared to an ontology (which is to say, with an ontology already known to be correct), human experts are indispensable.

MANUAL EVALUATION

Studies for manual evaluation by domain experts¹ need to have explicit methods and criteria and, wherever possible, metrics developed. For example, when designing an evaluation study, the criteria for correctness need to be carefully considered and clearly explicated to the domain experts. Consider the sentence, “Timolol is an ingredient of Timolol ophthalmic solution.” If simply presented with this sentence and asked whether it is true, most cardiologists would assert that it is true since, by definition, all Timolol ophthalmic solution has Timolol as an ingredient. If an ontology aims to represent linguistic intuitions (or perhaps conceptualizations), this is fine. However, if this is intended as a natural language expression of a description logic axiom in a realist ontology, it should be read as “All Timolol is an ingredient of some Timolol ophthalmic solution”, which is false. Only some Timolol is an ingredient of Timolol ophthalmic solution. This distinction would impact the evaluation results of RxNorm since it contains the triple (LIU et al., 2005):

Timolol ingredient_of ‘Timolol ophthalmic solution’.

¹ (OBRST et al., 2007) describes human evaluation as an approach.

The Bacterial Clinical Infectious Disease Ontology (BCIDO) is an example of a biomedical ontology that was evaluated for correctness by domain experts. Gordon et al. (2013) gives a brief report of ontology evaluation during the development phase of the BCIDO in which a knowledge elicitation technique laddering was used obtain information from infectious disease fellows and the answers were compared to statements in the ontology. For example, the domain experts were asked, “Can you tell me some bacteria that causes acute meningitis?” While the article is sparse on the details of the evaluation, considering possible scenarios in light of this question provides an opportunity to think through some important methodological details. Suppose a domain expert answers the question with “Streptococcus pneumonia.” It is now the task of the ontology engineer to look for an axiom in the ontology that expresses the relation between acute meningitis and Streptococcus pneumonia elicited. Suppose the following class description were in the ontology:

‘Streptococcus pneumoniae’ causes ‘acute meningitis’²

This would entail that every Streptococcus pneumoniae causes acute meningitis, which is false. Most people who are infected with Streptococcus pneumoniae do not develop acute meningitis. Not all Streptococcus pneumoniae infect an organism, and finally, a single bacterium is not sufficient to cause meningitis. So the domain expert’s ontologically naive statement ‘Streptococcus pneumoniae causes acute meningitis’ needs to be appropriately translated to a more accurate statements (or sets of statements) in the ontology such as

‘colony of Streptococcus’ bearer_of ‘infectious agent causing acute meningitis disposition’.

Ben Abacha et al., 2016 reports a semi-automatic approach to evaluating the correctness of statements in an ontology by converting formal statements to

² This axiom is not actually in BCIDO. Instead this class description is one disjunct in disjunctive class description that contains seventeen total disjunctions of the form (causes x). We use this example for simplicity; however, the same critique applies to the actual axioms in BCIDO.

natural language statements with templates. For example, the template that corresponds to the `is_a` relation is as follows:

Is CLASS a type of CLASS?

Notice that this template can lead to many false positives. For example, the question “Is a cat a type of pet?” is likely to elicit the answer, yes. In which case, an `is_a` link between `cat` and `pet` in a realist ontology would not be detected as false. This approach, however, is suitable for a wordnet that is meant to represent linguistic intuitions. These errors can be avoided by presenting natural language statements that involve an explicit quantifier to the domain specialist whenever the semantics of the logical language of the ontology use such quantifiers.

From these considerations, we can abstract the beginning of an evaluation approach the laddering approach for correctness evaluation used in Gordon et al. (2013) with the following steps.

1. employ a knowledge elicitation technique with domain experts;
2. translate natural language sentences to rigorously formalized ontological statements and/or vice versa as necessary;
3. compare the knowledge elicited with the knowledge encoded in the ontology, taking care to.

Next, we discuss how to measure correctness. The results reported in Gordon et al. (2013) simply state that comparison of the results with the ontology “demonstrated agreement with BCIDO class hierarchies”. The vagueness of this summary points to an outstanding question in ontology evaluation. How could we measure this agreement for a more rigorous evaluation? One approach is to compute an accuracy measure as proposed by Amith and Tao (2017) and Burton-Jones et al. (2005) that is the ratio of true statements to the number of statements in the ontology.³

Since the result is a ratio rather than a count, the accuracy measures of two ontologies could be compared. However, to ensure a meaningful comparison, the ontologies ought to be normalized prior to computing conducting the evaluation (VRANDEČIĆ; SURE, 2007). A correct normalization will ensure that statements that are explicit in one ontology but inferred in another are both evaluated. Consider the following three equivalent sets of `is_a` statements:

(1)

`lizard is_a mammal`
`cat is_a mammal`
`mammal is_a animal`

(2)

`lizard is_a mammal`
`lizard is_a animal`
`cat is_a mammal`
`mammal is_a animal`

(3)

`lizard is_a mammal`
`lizard is_a animal`
`cat is_a mammal`
`cat is_a animal`
`mammal is_a animal`

Although each set of statements is logically equivalent to the others (assuming the standard interpretation of ‘`is_a`’ as transitive), (1) has a correctness score of .66, (2) has a correctness score of .75, and (3) has a correctness score of .80. Normalizing ontologies to generate an ontology where all implied statements are explicit and therefore included in the evaluation is necessary for a reliable and comparable fidelity score. Through a process of normalization like that proposed by (VRANDEČIĆ; SURE, 2007), each set of statements would be normalized to (3), which is therefore the set of propositions to evaluate for fidelity, so .80 is the correct fidelity score for (1), (2), and (3).

³ Note that this requires an elicitation technique that also determines whether statements in the ontology are false.

FINAL REMARKS

This paper joins the general call for a rigorous discipline of ontology evaluation. It proposes that we systematize both the intrinsic and extrinsic properties of an ontology that can be evaluated according to the various aspects of an ontology. An initial review of the literature on ontology evaluation has revealed properties that emphasize ontologies as representational artifacts, as logical theories, as mathematical graphs, as parts of information systems, and as community resources for reuse. Further systematizes will likely reveal more categories and more fine grained distinctions. A discipline of ontology evaluation will also involve critically developing methods for examining those properties, developing comparative metrics that allow ontology engineers to compare the effects of various modeling choices and allow users to compare the merits of existing ontologies, and charting possible pitfalls of evaluation. As an example, we discussed methodological issues related to evaluating ontologies for correct representations.

The standard of correctness is different for realist, conceptualist, and linguistic ontologies, and this should be acknowledged in evaluation, and also considerations for applications. We also observed that elicitation techniques of domain knowledge from domain experts should be careful to present the domain knowledge in a manner that reflects the structure of the logical language the ontology is written in. For example, statements in OWL ontologies should be presented in natural language with an explicit quantifier to avoid ambiguity. Finally, metrics ought to be developed that can compare the correctness of ontologies and careful consideration ought to be given to the design of the evaluation study to ensure a comparable result. For example, the knowledge elicitation technique should detect both true and false statements in the ontology, and ontologies ought to be normalized to ensure meaningful comparison.

Future work for developing a systematic discipline of ontology evaluation includes a more complete survey of methods used to carry out ontology evaluations along with a detailed critique of their successes and pitfalls. Metrics ought to be developed that allow comparison and more work needs to be done determining what methods of ontology normalization are optimal for comparative metrics. While this paper has not discussed the many tools that exist for checking the quality of ontologies, undoubtedly the field will need more tools that can support comparing the performance of ontologies in information systems. Finally, data need to be generated that allow us to trace performance errors and successes in information systems to intrinsic properties of ontologies to help guide ontology development and selection.

ACKNOWLEDGEMENTS

I would like to thank Bill Hogan, Selja Seppälä, Andrew Spear, and Brian Stucky for helpful conversations during the writing of this paper. Work on this paper was supported in part by the NIH/NCATS Clinical and Translational Science Awards to the University of Florida UL1 TR000064. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES

- AMITH, M.; TAO, C. Modulated evaluation metrics for drug-based ontologies. *Journal of Biomedical Semantics*, v. 8, n. 1, 2017.
- ARP, R. et al. *Building ontologies with basic formal ontology*. New York: MIT Press, 2015.
- BEN ABACHA, A. et al. Towards natural language question generation for the validation of ontologies and mappings. *Journal of Biomedical Semantics*, n. 7, 2016.
- BURTON-JONES, A. et al. A semiotic metrics suite for assessing the quality of ontologies. *Data & Knowledge Engineering*, v. 55, n. 1, 2005.
- GANGEMI, A. et al. A theoretical framework for ontology evaluation and validation. *SWAP*, v. 166, 2005.
- _____. Modelling ontology evaluation and validation. *Semantic Web*, n. 4011, 2006.
- GOMEZ-PEREZ, A. Evaluation of ontologies. *International Journal of Intelligent Systems*, v. 16, n. 3, 2001.
- GORDON, C. L. et al. Design and evaluation of a bacterial clinical infectious diseases ontology. *AMIA Annual Symposium Proceedings*, 2013.
- GRÖNINGER, M.; FOX, M. S. The role of competency questions in enterprise engineering. In: *BENCHMARKING: theory and practice*. Boston, MA: Springer US, 1995.
- GUARINO, N. Toward a formal evaluation of ontology quality. *Ieee Intelligent Systems*, v. 19, n. 4, 2004.
- HOEHNDORF, R. et al. Evaluation of research in biomedical ontologies. *Briefings in Bioinformatics*, v. 14, n. 6, 2013.
- LIU, S. et al. Rxnorm: prescription for electronic drug information exchange. *IT professional*, v. 7, n. 5, 2005.
- MASOLO, C. et al. *Wonderweb deliverable D17: the wonderweb library of foundational ontologies and the DOLCE ontology*. [S.l.: s.n.], 2002.
- MILLER, G. A. Wordnet: a lexical database for English. *Communications of the ACM*, v. 38, n. 11, 1995.
- _____. FELLBAUM, C. Semantic networks of English. *Cognition*, v. 41, n. 1, 1991.
- NEUHAUS, F. et al. Towards ontology evaluation across the life cycle. *Applied Ontology*, v. 8, n. 3, 2013.
- OBRST, L. et al. The evaluation of ontologies: toward improved semantic interoperability. In: *SEMANTIC web*. Cheung, K.-H. Boston, MA: Springer US, 2007.
- ROSPOCHER, M. et al. Corpus-based terminological evaluation of ontologies. *Applied Ontology*, v. 7, n. 4, 2012.
- VRANDEČIĆ, D.; SURE, Y. How to design better ontology metrics. *The Semantic Web*, 2007.
- WALOSZEK, W. Measures for evaluation of structure and semantics of ontologies. *Metrology and Measurement Systems*, v. 19, n. 2, 2012.
- WILKINSON, M.D. et al. The fair guiding principles for scientific data management and stewardship. *Scientific Data*, n. 3, 2016.
- ZHU, X. et al. A review of auditing methods applied to the content of controlled biomedical terminologies. *Journal of Biomedical Informatics*, v. 42, n. 3, 2009.

Construindo ontologias com a metodologia ontofo- rinfoscience: uma abordagem detalhada das ativida- des do desenvolvimento ontológico

Fabício Martins Mendonça

Doutor em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG – Brasil, com período sanduíche em Universidade do Porto (U.Porto) - Portugal. Professor da Universidade Federal de Juiz de Fora (UFJF) - Juiz de Fora, MG - Brasil.

<http://lattes.cnpq.br/7587726616949092>

E-mail: fabriciommendonca@gmail.com

António Lucas Soares

Doutor em Ciência da Informação pela Universidade do Porto, Porto, Portugal.

Professor Associado, Departamento de Engenharia Informática, INESC TEC, Porto, Portugal.

E-mail: als@fe.up.pt

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

RESUMO

Na utilização de ontologias como sistemas de organização do conhecimento para fins de representação, organização e recuperação da informação, um dos maiores desafios é o processo de construção da ontologia, o qual é complexo e exige conhecimentos técnicos específicos dos desenvolvedores. Nos últimos anos, diversas metodologias foram criadas para auxiliar na construção de ontologias, as quais possuem passos bem estabelecidos mas nem sempre bem detalhados, causando problemas no desenvolvimento de ontologias. O presente artigo descreve uma metodologia para construção de ontologias, denominada OntoForInfoScience, que tem como diferencial o detalhamento das atividades do ciclo de desenvolvimento ontológico, explicando mais claramente termos técnicos, questões lógicas e filosóficas envolvidas na construção de ontologias. Como resultados práticos atuais da OntoForInfoScience, apresenta-se parte do conteúdo de duas ontologias desenvolvidas com tal metodologia: (i) Hemonto: uma ontologia biomédica sobre os componentes do sangue humano utilizados para hematologia e hemoterapia; e (ii) Ontolegis: uma ontologia de domínio jurídico que representa o conhecimento relativo à informação legislativa brasileira sobre o direito médico. Conclui-se que a OntoForInfoScience foi de grande utilidade para a criação de representações ontológicas formais em ambos os domínios modelados, e seus detalhamentos permitiram auxiliar os desenvolvedores em questões lógicas e filosóficas do processo de construção e no entendimento de conceitos técnicos de ontologias.

Palavras-chave: Ontologia. Desenvolvimento de ontologias. Hematologia. Direito médico.

Bulding ontologies with ontoforinfoscience methodology: a detailed approach to the activiyies of ontology development

ABSTRACT

In the use of ontologies as knowledge information systems (KOS) for purposes of representation, organization and information retrieval, one of the biggest challenges is the ontology development process, which is complex and requires specific technical knowledge of the developers. In the recent years, several methodologies have been developed to support the ontologies building, which have steps well established, but not always well detailed for developers, causing problems in the ontologies development. This paper describes a methodology for the ontologies development, called OntoForInfoScience, which has a differential the detail of the activities of the ontology development cycle, explaining more clearly technical terms, logical and philosophical issues involved in the construction of ontologies. As current practical results of OntoForInfoScience, we presented part of the content of two ontologies developed with such methodology: (i) Hemonto: a biomedical ontology on the human blood components used for hematology and hemothrapy; (ii) Ontolegis: a legal domain ontology that represents the knowledge related to Brazilian legislative information on medical law. It is concluded that the OntoForInfoScience was very useful for the construction of formal ontological representations in both domains modeled and their details allowed to help the developers in logical and philosophical questions of the process of development and in the understanding of technical concepts of ontologies.

Keywords: *Ontology. Ontologies development. Hematology. Medical law.*

Construyendo ontologías con la metodología ontoforinfoscience: un enfoque detallado de las actividades del desarrollo ontológico

RESUMEN

En la utilización de ontologías como sistemas de organización del conocimiento para fines de representación, organización y recuperación de la información, uno de los mayores desafíos es el proceso de construcción de la ontología, el cual es complejo y exige conocimientos técnicos específicos de los desarrolladores. En los últimos años, diversas metodologías han sido desarrolladas para auxiliar en la construcción de ontologías, las cuales poseen pasos bien establecidos pero no siempre bien detallados a los desarrolladores, causando problemas en el desarrollo de ontologías. El presente artículo describe una metodología para la construcción de ontologías, denominada OntoForInfoScience, que tiene como diferencial el detalle de las actividades del ciclo de desarrollo ontológico, explicando más claramente términos técnicos, cuestiones lógicas y filosóficas involucradas en la construcción de ontologías. Como resultados prácticos actuales de OntoForInfoScience, se presenta parte del contenido de dos ontologías desarrolladas con tal metodología: (i) Hemonto: una ontología biomédica sobre los componentes de la sangre humana utilizados para hematología y hemoterapia; (ii) Ontolegis: una ontología de dominio jurídico que representa el conocimiento relativo a la información legislativa brasileña sobre el derecho médico. Se concluye que la OntoForInfoScience fue de gran utilidad para la creación de representaciones ontológicas formales en ambos dominios modelados y sus pormenores permitieron auxiliar a los desarrolladores en cuestiones lógicas y filosóficas del proceso de construcción y en el entendimiento de conceptos técnicos de ontologías.

Palabras clave: *Ontología. Desarrollo de ontologías. Hematología. Derecho médico.*

INTRODUÇÃO

Nos últimos anos pôde-se perceber crescimento significativo em pesquisas e aplicações que fazem uso de ontologias como instrumentos de representação do conhecimento dos mais diversos domínios. As possibilidades de uso de ontologias em ciência da informação e em áreas correlacionadas foram reforçadas pelo grande volume e diversidade de dados manipulados na Web, e também pelo avanço no uso de tecnologias semânticas, as quais fazem parte do que se convencionou chamar de Web semântica (Breitman, 2005). No contexto informacional atual, ontologias desempenham papel central, sendo amplamente utilizadas em aplicações das áreas de Organização do Conhecimento, *Business Intelligence*, *Big Data* e *Data Science*.

Os instrumentos utilizados para representação do conhecimento tais como ontologias, tesouros, taxonomias, entre outros, são comumente denominados na atualidade sistemas de organização do conhecimento (SOC), termo traduzido do original em inglês Knowledge Organization Systems (KOS). Os SOCs são instrumentos que tratam da representação do conteúdo de documentos para fins de armazenamento, organização, recuperação e compartilhamento de informação. A utilização de ontologias como um SOC é o seu uso mais comum no âmbito da ciência da informação, com destaque para representação e organização do conhecimento especializado dos mais diversos campos da biomedicina. Smith (2004) afirma que ontologias são utilizadas na padronização dos vocabulários médicos como uma alternativa às terminologias médicas tradicionais, por possuírem maior expressividade na representação do conhecimento e devido ao uso de formalismos lógicos. Além disso, ontologias possibilitam representar o conhecimento de forma mais próxima da realidade, o que é fundamental para a biomedicina (FREITAS E SCHULZ, 2009).

O processo de construção de uma ontologia é complexo, pois envolve a criação de modelos semânticos ou descrições simplificadas da realidade de um dado domínio e também exige

dos desenvolvedores conhecimentos técnicos em modelagem conceitual, em lógica formal e em alguns aspectos filosóficos. Para apoiar os desenvolvedores, nos últimos anos diversas metodologias foram criadas para esse fim, tais como: metodologia de Uschold e King (1995); metodologia de Gruninger e Fox (1995); *Methontology* (Gómez-Perez, Fernandez-Lopes e Vicente, 1996); metodologia NeOn (Suaréz-Figueroa, 2008), entre outras. Em uma análise geral, tais metodologias possuem passos bem estabelecidos do processo de construção, nem sempre bem detalhados aos desenvolvedores, causando problemas como, por exemplo, criação de relações ontológicas muito genéricas e imprecisas, ausência de definições formais para as classes e relações, entre outros.

O presente artigo apresenta uma nova metodologia e mais detalhada para construção de ontologias, denominada *OntoForInfoScience*, e seus resultados práticos alcançados no desenvolvimento de duas ontologias de domínio: (i) a *Hemonto*: uma ontologia biomédica sobre os componentes do sangue humano utilizados para hematologia e hemoterapia; e (ii) a *Ontolegis*: uma ontologia de domínio jurídico que representa o conhecimento relativo à informação legislativa brasileira sobre o direito médico.

De acordo com Mendonça (2015), a *OntoForInfoScience* tem como principal diferencial em relação às outras metodologias disponíveis, um detalhamento das atividades necessárias do ciclo de desenvolvimento ontológico, a fim de auxiliar especialistas em organização do conhecimento, incluindo cientistas da informação, a superar problemas relativos ao jargão técnico e às questões lógicas e filosóficas que envolvem a construção de ontologias. Evidentemente, a solução de tais problemas no contexto da ciência da informação demanda muitas outras ações, como por exemplo, a discussão e inserção de requisitos de formação técnica-acadêmica nas disciplinas da área, as quais são necessárias, mas extrapolam o escopo da presente pesquisa. Ainda sim, a metodologia *OntoForInfoScience* representa uma iniciativa em direção a maior entendimento de termos e

detalhes tcnicos (lgicos e filosfcos) do processo de desenvolvimento de ontologias por parte dos cientistas da informao.

Com relao ao processo de construo das ontologias de domnio *Hemonto* e *Ontolegis*, a metodologia *OntoForInfoScience* possibilitou detalhar o emprego de ontologias de fundamentao, a criao de definies formais para classes (derivadas de definies textuais), a caracterizao dos diferentes tipos de relaes ontolgicas, dentre outras questes tcnicas. Os detalhamentos se mostraram teis para guiar os desenvolvedores de ontologias ao longo de toda a atividade.

O restante do presente artigo est organizado da seguinte forma. A segunda seo apresenta alguns dos principais problemas atualmente identificados na construo de ontologias. A terceira seo descreve, brevemente, as etapas para o desenvolvimento ontolgico atravs da *OntoForInfoScience*. A quarta seo destaca o processo de construo das ontologias *Hemonto* e *Ontolegis*, apresentando parte do contedo de tais ontologias. E, por fim, a quinta seo traz as concluses e consideraes finais sobre a pesquisa apresentada neste artigo.

PROBLEMAS NA CONSTRUO DE ONTOLOGIAS

Mesmo com o crescimento notvel das atividades de pesquisa sobre ontologias e de suas aplicaes prticas na Web nos ltimos anos, algumas necessidades e problemas da rea permanecem em aberto, talvez em funo da imaturidade de ser uma rea de pesquisa e aplicao recente. Dentre os problemas identificados, atualmente, aqueles que foram objeto de estudo desta pesquisa correspondem aos problemas que surgem ao longo do processo de construo de uma ontologia. De fato, o processo de construo de uma ontologia est complexo e exige do seu desenvolvedor uma compreenso apropriada do domnio a ser representado, alm de conhecimentos tcnicos em modelagem conceitual, formalismos lgicos e at mesmo em fundamentos filosfcos.

A fim de apoiar os desenvolvedores no processo de construo de ontologias, diferentes metodologias e mtodos de construo foram elaborados e publicados na literatura da rea. Cronologicamente, houve uma proliferao de metodologias para construo de ontologias entre meados da dcada de 90 e o incio dos anos 2000. Nesse perodo surgiram as metodologias at hoje mais referenciadas por pesquisas na rea, tais como: metodologia de Uschold e King (1995) – *Enterprise Ontology*; metodologia de Gruninger e Fox (1995) – *Toronto Virtual Enterprise (TOVE)*; a *Methontology* (Gmez-Perez, Fernandez-Lopes e Vicente, 1996); o mtodo 101 de Noy e McGuinness (2001); o mtodo CYC de Reed e Lenat (2002); a *On-To-Knowledge Methodology (OTKM)* (Sure, Staab e Stuber 2003); dentre outras. De meados dos anos 2000 at hoje, surgiram outras metodologias de construo bem referenciadas na rea, tais como a metodologia NeOn (Suarz-Figueroa, 2008) e a *Up for ONtology (UPON)* (De Nicola, Missikoff e Navigli, 2009).

Mesmo com a disponibilidade atual de nmero significativo de metodologias para ontologias, diversos problemas relacionados ao seu uso permanecem em aberto. Para apresentar tais problemas optou-se por separ-los em duas classes mais gerais de problemas: (i) a primeira est relacionada com as limitaes atuais das metodologias de construo; e (ii) a segunda engloba os problemas identificados nas ontologias decorrentes de um processo de construo mal conduzido. Os problemas relacionados com as metodologias de construo impactam diretamente a segunda classe de problemas. Entretanto, est importante considerar que nem todos os problemas da segunda classe so culpa das metodologias de construo, tais como: erros decorrentes do projeto ontolgico, de especificao, de no avaliao e no documentao da ontologia, alm de desconhecimento ou falha humana dos desenvolvedores no processo de construo.

Sobre os problemas relacionados s metodologias para construo de ontologias, pode-se afirmar que:

- não há um padrão de construção ou uma metodologia unificada que seja amplamente aceita para o desenvolvimento de ontologias (Uschold e Gruninger, 1996; Jones, Bench-Capon e Visser, 1998; Fernández-Lopez, 1999; Fernández-Lopez e Corcho, 2004; Breitman, 2005; Cardoso, 2007). Tais metodologias, ainda hoje, não são consideradas suficientemente “maduras” como as metodologias empregadas na área de engenharia de software;
 - as metodologias atuais apresentam abordagens e características diversas, sendo direcionadas a propósitos e aplicações específicas (FERNANDÉZ-LOPEZ, 1999), tais como aquelas dedicadas ao domínio dos negócios;
 - em relação a detalhes das atividades e dos procedimentos para sua condução, algumas metodologias e métodos mostram-se superficiais na elucidação dos passos para construção de ontologias, parecendo até considerar que o ontologista já domina o assunto sobre construção de ontologias e não necessita de detalhes acerca de atividades e procedimentos envolvidos (SILVA, SOUZA E ALMEIDA, 2008);
 - algumas abordagens dão mais ênfase a atividades de desenvolvimento, principalmente para implementação da ontologia, como, por exemplo, os métodos CYC e 101, desconsiderando aspectos importantes relacionados a gerenciamento do projeto, a estudo de viabilidade, à manutenção e à avaliação de ontologias (SILVA, SOUZA E ALMEIDA, 2008);
 - o uso recorrente de conceitos específicos da área de ciência da computação não explicados adequadamente aos demais desenvolvedores de ontologias dificulta seriamente o trabalho de construção de ontologias por parte de profissionais que não são da área computacional. Herança múltipla, cardinalidade, tipos de dados (string, inteiro, real, booleano), classe abstrata, classe disjunta são apenas alguns exemplos de conceitos comuns em ciência da computação, necessários na construção de ontologias, não explicados aos desenvolvedores nas metodologias.
- A segunda classe de problemas relacionados à construção de ontologias engloba os problemas de conteúdo de uma ontologia, decorrentes de processos de desenvolvimento mal conduzidos. Dentre tais problemas pode-se identificar os seguintes:
- inclusão de relações que não são genuinamente ontológicas (MUNN E SMITH, 2008);
 - criação e definição de relações de maneira intuitiva, com base em nossa própria experiência de vida (MUNN E SMITH, 2008);
 - criação de instâncias como classes, tornando a hierarquia da ontologia excessivamente especializada (NOY E MCGUINNES, 2001);
 - inclusão de heranças múltiplas: uma subclasse da ontologia que está ligada a mais de uma superclasse (SMITH, 2005);
 - utilização imprópria da relação *is_a* entre classes da ontologia, misturando os diferentes tipos de relação *is_a* que podem existir, tais como instanciação, especialização e equivalência (POVEDA-VILLALÓN, 2010);
 - criação e utilização de relações todo-parte (*has_part* e *part_of*) ambíguas e inconsistentes (BITTNER E DONNELLY, 2007);
 - não caracterização ou não utilização dos diferentes tipos de relação *part_of* existentes no domínio representado pela ontologia (KEET E ARTALE, 2008);
 - Definição de relações muito genéricas (CEUSTERS et al., 2004);
 - Construção de definições formais de classes e de relações de forma imprecisa (SCHULZ, 2006).

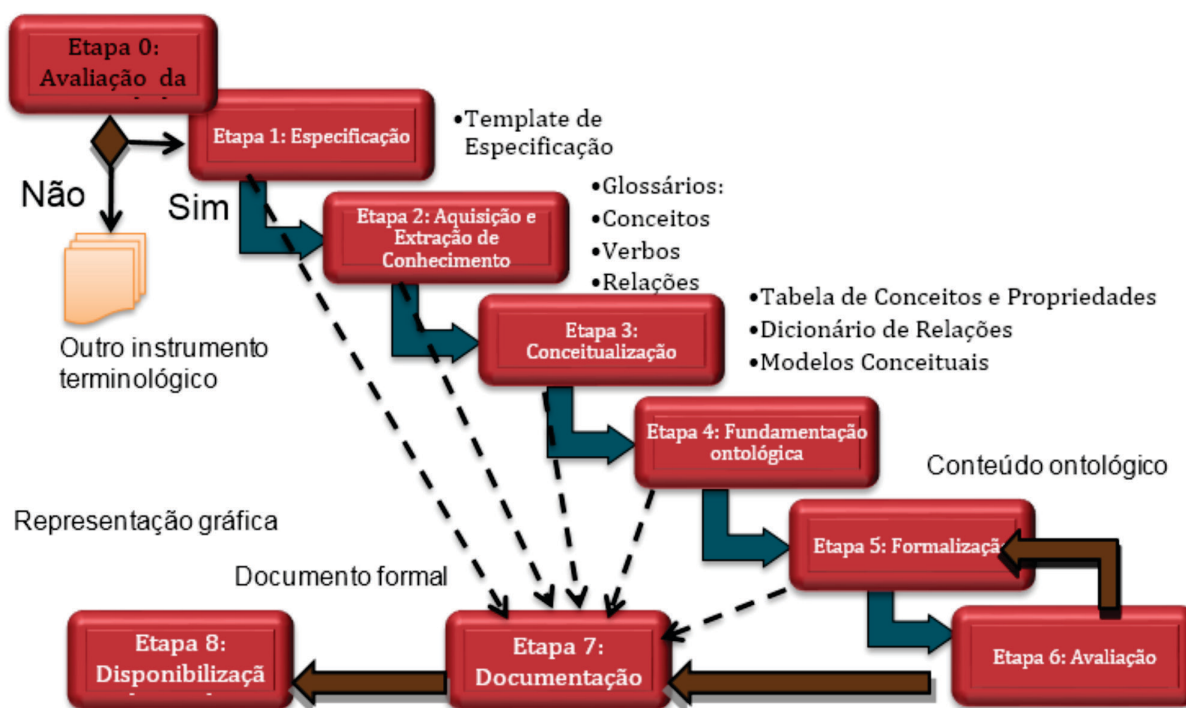
METODOLOGIA ONTOFORINFOSCIENCE

A metodologia *OntoForInfoScience* foi criada com o prop3sito de possibilitar que especialistas em organizao do conhecimento, incluindo cientistas da informao e outros desenvolvedores, possam superar problemas relativos ao jarg3o t3cnico e 3s quest3es l3gicas e filos3ficas no desenvolvimento de ontologias (MENDONÇA, 2015). Para tanto, a *OntoForInfoScience* detalha os passos necess3rios do processo de construo de ontologias em uma linguagem simples e adequada a profissionais diversos, caracterstica que a diferencia das atuais metodologias existentes.

Ainda que apresente uma caracterstica particular, a metodologia *OntoForInfoScience* tamb3m se baseou em algumas etapas de metodologias j3 existentes, especificamente a *Methontology*, o m3todo 101 *Method* e metodologia *NeOn*, a fim de reutilizar algumas etapas e suprir as limita33es presentes em cada uma delas: detalham-se as etapas e passos reutilizados, al3m de realizar adapta33es necess3rias para obter uma linguagem apropriada a todos desenvolvedores de ontologia.

A metodologia *OntoForInfoScience* engloba, ao todo, uma pr3-etapa e oito etapas, conforme apresentado na figura 1. No restante da presente se33o apresenta-se breve descri33o dessas etapas. A descri33o completa est3 disponvel em Mendonça (2015).

Figura 1 – Etapas da metodologia OntoForInfoScience



Fonte: Elaborado pelo autor (2017).

Na pré-etapa, o desenvolvedor realiza uma avaliação prévia sobre a real necessidade da construção de uma ontologia para o problema analisado. Se o objetivo desse desenvolvimento consiste, por exemplo, unicamente na indexação e recuperação de informações em um domínio específico, a construção de um tesouro é suficiente. Entretanto, se o objetivo consiste na representação de aspectos e objetos do mundo real e também a necessidade do uso de relações mais extensíveis do que aquelas contidas em vocabulários controlados, a construção de uma ontologia é fundamental.

Decidida a necessidade de desenvolver a ontologia, inicia-se a sequência de etapas do processo de construção. Na etapa 1 realiza-se a especificação da ontologia através do *template* de especificação, o qual deve conter informações sobre o domínio e escopo, propósito geral, classes de usuários que representam o público-alvo da ontologia, cenários de aplicação para uso da ontologia e o grau de formalidade. Estabelece-se também o tipo da ontologia: ontologia de alto ou médio nível, de domínio ou de tarefa; ontologia leve ou pesada. Por fim, delimita-se o escopo de cobertura da ontologia descrevendo-se o ponto de partida, o limite do domínio coberto e questões de competência.

A etapa 2 consiste na aquisição e extração de conhecimento, que engloba as atividades de seleção dos materiais de referência do domínio e a adoção de métodos de extração do conhecimento. Na *OntoForInfoScience*, essas atividades são conduzidas para combinar diferentes métodos de extração, como: análise textual e manual de textos, extração terminológica automática, método semiautomático de identificação de conceitos do domínio. Além disso, busca-se colaboração, isto é, a aquisição de conhecimento envolvendo diferentes visões do domínio a partir de consulta a especialistas diversos. Os artefatos resultantes da etapa 2 são glossários de conceitos, de verbos e de relações, os quais são utilizados como parâmetros de entrada na etapa 3.

Na etapa 3, de conceitualização, são realizadas atividades como identificação e análise dos conceitos do domínio, que serão incluídos como classes da ontologia. Além disso, realiza-se a organização e

a estruturação do conhecimento a fim de obter relações, propriedades e restrições da ontologia. Na *OntoForInfoScience* esse processo ocorre a partir da transformação dos glossários construídos na etapa anterior em artefatos conceituais da etapa 3, os quais correspondem ao seguintes: i) tabela de conceitos e propriedades, que contém os conceitos identificados e validados, suas definições, seus termos sinônimos, seus valores possíveis e suas propriedades (restrições) a nível conceitual; ii) dicionário de verbos, que contém os verbos identificados e validados como candidatos à relação na ontologia e suas definições textuais; e iii) os modelos conceituais gráficos, que representam as relações conceituais entre os conceitos identificados através de grafos ou estruturas similares.

A etapa 4 corresponde à fundamentação da ontologia desenvolvida, isto é, o desenvolvedor deve pesquisar por ontologias de fundamentação que atuem como ponto de partida e escolher aquela mais apropriada ao objetivo proposto. Ao fazer a escolha, o desenvolvedor deve considerar a abordagem filosófica subjacente à ontologia de fundamentação, para justificar decisões ontológicas na modelagem do domínio. A ontologia de fundamentação selecionada deve ser importada para um editor de ontologias para início implementação.

Na etapa 5 realiza-se a representação formal da ontologia através de uma linguagem lógica. Nessa etapa produz-se a descrição formal do domínio a partir da conceitualização anterior (etapa 3). Portanto, o conhecimento do domínio, tratado anteriormente apenas em nível conceitual, passa a ser tratado em nível ontológico-formal, o que implica uma série de adaptações nas estruturas conceituais a fim de atender às restrições ontológicas e formais. Estão incluídas como atividades: i) construir a taxonomia geral da ontologia, baseada na estrutura da taxonomia da ontologia de fundamentação escolhida; ii) definir propriedades descritivas das classes, as quais envolvem apenas atributos textuais, tais como nomes, sinônimos, definições textuais e anotações; iii) criar definições formais para cada uma das classes da ontologia usando uma linguagem lógica, de modo que a definição formal seja derivada da definição textual elaborada

no passo anterior; iv) definir propriedades de dados das classes, as quais envolvem atributos como tipos de dados, cardinalidade e uso dos quantificadores universais ou existenciais; v) criar instncias das classes da ontologia, a fim de testar o funcionamento da ontologia; vi) especificar relaes ontolgicas, o que consiste na aplicao de um conjunto definido de regras e princpios para a transio das relaes conceituais para o nvel formal-ontolgico. Dentre essas regras, tm-se como exemplos: a restrio a incluso de relaes a apenas aquelas identificadas como entidades da realidade, a caracterizao dos tipos de uma relao ontolgica, tais como tipos *is_a* e tipos *part_of*, dentre outras; vii) definir propriedades das relaes ontolgicas, as quais incluem o nome da relao, a definio semiformal, a propriedades lgicas bsicas, o domnio e a faixa, dentre outras.

A etapa 6 envolve a avaliao da ontologia que, no caso da *OntoForInfoScience*, corresponde a um conjunto de critrios que permitem realizar tanto a validao ontolgica (adequao da ontologia ao domnio – mundo real) quanto a verificao ontolgica (anlise da ontologia quando a correo de sua construo). Exemplos de critrios de validao so natureza ontolgica, universalidade de classes e relaes e grau de representatividade do domnio. Exemplos de critrios de verificao so a no recursividade nas definies, a especificao dos diferentes tipos da relao *part_of*, a definio de relaes inversas e criao de cardinalidades.

Na etapa 7, de documentao, produz-se o documento formal da ontologia. A elaborao da documentao é realizada, praticamente, ao longo de todo o processo de construo. O contedo do documento formal da ontologia engloba o documento de especificao (etapa 1), materiais ou documentos de referncia do domnio (etapa 2), conjunto de modelos conceituais (etapa 3), ontologias reutilizadas (etapas 4 e 5), contedo ontolgico-formal (etapa 5), documento de avaliao da ontologia (etapa 6).

Por fim, a etapa 8 cuida para que a ontologia desenvolvida seja disponibilizada em meio eletrnico de fcil acesso e visualizao aos usurios.

Como passos desta etapa, tem-se: i) a exportao do contedo formal da ontologia (desenvolvido na etapa 5), atravs de algum editor de ontologias para alguma linguagem lgica, em geral a *Ontology Web Language* (OWL), para disponibilizao na web; ii) a apresentao da ontologia em formato grfico para mais fcil entendimento pelos usurios, o que pode incluir tanto um documento esttico, quanto uma interface de busca ao contedo da ontologia.

ONTOLOGIAS DESENVOLVIDAS COM A ONTOFORINFOSCIENCE

Como resultados prticos da *OntoForInfoScience* foram desenvolvidas, at o presente momento, duas ontologias de domnio utilizando a metodologia: (i) *Hemonto*: uma ontologia biomdica de domnio que representa o conhecimento relativo aos componentes do sangue humano (hemocomponentes e hemoderivados) usados para fins teraputicos (MENDONÇA E ALMEIDA, 2013) (MENDONÇA, 2015); (ii) *Ontolegis*: uma ontologia de domnio jurdico que representa o conhecimento relativo a informao legislativa brasileira sobre o Direito Mdico (TORRES, 2017). Ambas as ontologias possuem carter cientfico e social relevante, uma vez que podem ser consideradas ineditas cada qual no seu domnio e representam conhecimento essencial de domnios muito importantes para a populao em geral, servindo como instrumento de apoio aos especialistas de cada uma dessas reas. A presente seo discorre brevemente sobre tais ontologias, evidenciando os resultados alcanados com a aplicao dos passos da *OntoForInfoScience* no processo de construo.

HEMONTO

A *Hemonto* foi a primeira ontologia desenvolvida utilizando a metodologia *OntoForInfoScience*, de tal forma que seu desenvolvimento serviu como primeiro resultado prtico da aplicabilidade das etapas e atividades previstas na *OntoForInfoScience*. Do ponto de vista de representatividade do domnio, a *Hemonto* foi desenvolvida no mbito de

um projeto de escopo maior – o *Blood Project*¹ – que tem por objetivo gerar uma linguagem formal sobre o sangue humano, considerando a carência atual de representações em tal domínio.

Em sua versão atual, a *Hemonto* engloba um total de 209 termos, dos quais 155 são classes e 54 são relações. Seguindo a nomenclatura da ontologia Basic Formal Ontology (BFO) (GRENON E SMITH, 2004), 113 classes são continuantes e 41 classes são ocorrentes. A origem de cada uma dessas classes e relações da *Hemonto* é apresentada na tabela 1, a seguir.

Seguindo as etapas e atividades incluídas na *OntoForInfoScience*, a especificação da ontologia *Hemonto* (etapa 1) foi realizada através do *template* de especificação da metodologia e suas orientações. Dentre os elementos de especificação da *Hemonto*, temos: (i) propósito geral da ontologia: orientar profissionais das áreas de saúde e da informação sobre os procedimentos adequados na extração, manipulação e armazenamento dos componentes do sangue utilizados para fins hemoterapêuticos; (ii) escopo de cobertura da ontologia: como ponto de partida entidades do mundo real, categorizados por meio das classes da ontologia BFO; e o termo “sangue” (*portion of blood*) importado da ontologia FMA. Como limite de cobertura os componentes caracterizados como proteínas e enzimas presentes no sangue humano, os quais são representados e definidos em ontologias específicas desse subdomínio: a Gene Ontology (GO), a Chemical Biomedical Investigations (ChEBI) e a Cell Ontology (CL).

Tabela 1– Origem das classes da HEMONTO

Classes		Relações	
Ontologia de origem	Núm.	Ontologia de Origem	Núm.
HEMONTA (específicas)	62	HEMONTA (específicas)	14

(Continual)

Tabela 1– Origem das classes da HEMONTO (Conclusão)

Classes		Relações	
Ontologia de origem	Núm.	Ontologia de Origem	Núm.
Basic Formal Ontology (BFO)	34	Basic Formal Ontology (BFO)	20
Foundational Model Anatomy (FMA)	22	Relation Ontology (RO)	18
Protein Ontology (PRO)	14	Gene Ontology (GO)	2
Cell Ontology (CL)	9		
Chemical Entities of Biological Interest (ChEBI)	4		
Measurement Method Ontology (MMO)	4		
Gene Ontology (GO)	4		
Ontology for Biomedical Investigations (OBI)	1		
Phenotypic Quality Ontology (PATO)	1		
National Drug File Reference Terminology (NDF-RT)	1		
Total:	155	Total:	54

Total de entidades: 209

Para aquisição e extração do conhecimento do domínio (etapa 2), foram utilizados os seguintes documentos de referência no domínio do sangue, levantados com o apoio de especialistas da área que trabalham na instituição Hemominas²: i) guia de hemocomponentes do Ministério da Saúde brasileiro; ii) padrão terminológico internacional ISBT 28 (ICCBBA, 2010); iii) manual técnico sobre sangue e terapia celular da American Association of Blood Banks (AABB, 2005); iv) livro-texto sobre

² Hemominas é uma fundação centro de hematologia e hemoterapia do estado de Minas Gerais, vinculado à Secretaria do Estado (SES-MG), que tem por finalidade assegurar unidade de comando e direção às políticas estaduais relativas à hematologia e hemoterapia, garantindo à população a oferta de sangue e hemoderivados de qualidade (Hemominas, 2017).

¹ Disponível em <http://mba.eci.ufmg.br/blood/>. Acesso em 26 de junho de 2017.

hematologia clínica Wintrobe's Clinical Hematology 12a edição (Greer et al., 2009). Quanto aos métodos de extração de conhecimento, foram utilizados: i) o software de análise linguística SketchEngine³ para o processo de extração terminológica de termos frequentes encontrados nos documentos; ii) o framework colaborativo Conceptualization Modelling Environment (ConceptME) (SOUSA et al., 2013) para identificação e extração semiautomática de conceitos e relações do domínio; e iii) análise de assunto dos textos dos materiais de referência para identificação de conceitos e relações candidatas.

Na etapa de conceitualização (etapa 3) foram produzidos os seguintes artefatos: (i) um dicionário de conceitos; (ii) um dicionário de verbos; (iii) uma tabela de conceitos e propriedades; (iv) um conjunto de modelos conceituais gráficos do domínio do sangue utilizando as ferramentas *ConceptMe*⁴, *DiagramEditor*⁵ e *OmniGraffle*⁶. Como exemplo desse modelo conceitual, apresenta-se na figura 2 um mapa conceitual que representa o processo de obtenção dos componentes do sangue derivados do plasma. Os conceitos em azul são equivalentes às classes importadas da ontologia FMA e os demais conceitos são específicos da *Hemonto*.

Na etapa 4, fundamentação ontológica, foram adotadas duas ontologias de fundamentação: i) Basic Formal Ontology (BFO) (GRENON E SMITH, 2004), que foi usada como ponto de partida para a construção da *Hemonto* e na definição das classes mais gerais; ii) Relation Ontology (RO) (SMITH et al., 2005), utilizada para especificação de relações ontológicas entre as classes do domínio.

O conteúdo OWL de ambas as ontologias foi importado usando o editor de ontologias Protégé 4.3 para dar início ao desenvolvimento específico da ontologia.

A formalização da *Hemonto* (etapa 5) foi realizada a partir dos seguintes passos: i) construção da taxonomia geral da ontologia, tomando como ponto de partida as entidades mais gerais da BFO e categorizando as entidades específicas do sangue abaixo delas; ii) definição das propriedades descritivas das classes junto ao fornecimento dos atributos textuais correspondentes no Protégé 4.3 (*ID*, *label*, *imported_from*, *hasSynonym*, *definition*, *comments*, etc.); iii) criação de definições formais em OWL-DL para cada classe da ontologia dentro do Protégé (vide exemplo de definição da classe “*portion of plasma*” na figura 3); iv) definição das propriedades de dados das classes para representar as características de cada componente do sangue, tais como: volume apropriado, porcentagem de hematócritos, temperatura de armazenamento, validade, entre outras; v) criação de instâncias das classes usando um conjunto de regras básicas, que envolvem a identificação de instâncias via questões de competência; vi) especificação de relações ontológicas a partir das regras e princípios para transição das relações conceituais para o nível formal-ontológico, tais como tipos de relações *is_a* e tipos de relações *part_of*; vii) definição das propriedades descritivas e lógicas de cada relação da ontologia, utilizando atributos correspondentes no Protégé (*label*, *definition*, *characteristics*, *inverse_of*, *domains*, *ranges*, etc.).

O conteúdo ontológico formalizado nessa etapa 5 foi representado através dos dicionários gerais de classes e relações da ontologia e também como conteúdo do Protégé 4.3, expresso na linguagem OWL/XML. A tabela 2, a seguir, mostra exemplos de algumas entidades definidas formalmente na *Hemonto* e incluídas nos dicionários de classes e relações da ontologia.

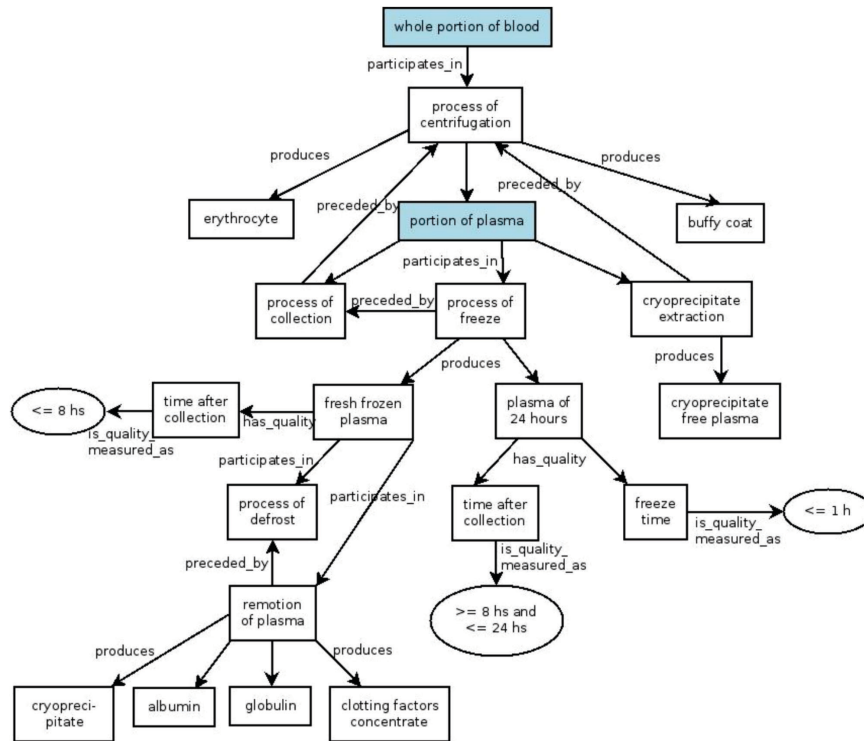
³ Disponível em: <http://Sketch Engine.co.uk/>. Acesso em 26 de junho de 2017.

⁴ Disponível em: http://www.conceptme.pt/conceptme/index.php/Main_Page. Acesso em 26 de junho de 2017.

⁵ Disponível em: <http://dia-installer.de/>. Acesso em 26 de junho de 2017.

⁶ Disponível em: <https://www.omnigroup.com/omnigraffle>. Acesso em 26 de junho de 2017.

Figura 2 – Mapa conceitual do processo de obtenção dos componentes derivados do plasma



Fonte: Elaborado pelo autor (2017).

Figura 3 – Exemplo de definição formal de classe na ontologia

Portion of plasma: definição textual

Da FMA - Portion of plasma: portion of body substance corresponding to the liquid component of blood in which blood cells are immersed.



Anotação da definição textual

Da FMA - Portion of plasma: portion of body substance [FMA/HEMONTTO: *portion of body substance*] **corresponding to** [BFO: *has_quality*] the liquid [FMA/HEMONTTO: *liquid*] **component of** [RO: *part_of*] blood [FMA/HEMONTTO: *portion of blood*] in which blood cells [FMA/HEMONTTO: *blood cell*] **are immersed** [RO: *contains*].



Portion of plasma: definição formal (em lógica OWL-DL)

Portion of plasma is **a** *portion of body substance* and (**has_quality** *liquid*) and (**part_of** *portion of blood*) and (**contains** *blood cell*).

Fonte: Elaborado pelo autor (2017).

Tabela 2 – Formalismos e axiomas lgicos das entidades da HEMONTO

Entidade (classe ou relao)	Propriedades
<p><i>whole portion of blood</i></p> <p>Sinônimos: <i>entire portion of blood; maximum portion of blood; portion of blood in whole body</i></p>	<p>Importado de: FMA: <i>whole portion of blood</i> (FMAID: 263001)</p> <p>Definio textual: <i>Portion of blood collected into an anticoagulant and not further processed.</i></p> <p>Definio formal: <i>is_a 'portion of blood' collected_in some anticoagulant and not (separated)</i></p> <p>Comentrios: <i>A unit of blood collected into an anticoagulant and not further processed unless otherwise specified.</i></p>
<p><i>anticoagulant</i></p> <p>Sinônimos: <i>anticoagulante, anticoagulants</i></p>	<p>Importado de: CHEBI: <i>anticoagulant</i> (CHEBI: 50249)</p> <p>Definio textual: <i>An agent that prevents blood clotting.</i></p> <p>Definio formal: <i>is_a realizable_entity and prevent_of some 'blood coagulation'</i></p>
<p><i>collected_into</i></p>	<p>Origem: HEMONTO</p> <p>Definio semi-formal: C collected_into P – Relao ontolgica identificada entre um continuante C e um ocorrente P, que uma entidade realizvel, tal que o continuante C extraido ou coletado atravs de algum meio ou processo, que corresponde a entidade realizvel P.</p> <p>Domnio: <i>material_entity</i></p> <p>Imagem: <i>realizable_entity</i></p> <p>Relao inversa: <i>has_collected</i></p> <p>Exemplo de uso: <i>whole portion of blood is_a 'portion of blood' collected_in some anticoagulant and not (separated)</i></p>
<p><i>participates_in</i></p>	<p>Origem: RO (RO_ID: 0056)</p> <p>Definio textual: <i>A relation between a continuant and a process, in which the continuant is somehow involved in the process.</i></p> <p>Definio formal: $"x,y(participates_in_{it}(x,y) \equiv mpart_of(x,y)^{ED(x)^{PD(y))}$ onde ED é um endurente ou continuante e PD é um perdurante ou ocorrente.</p> <p>Relao inversa: <i>has_participant</i></p> <p>Domnio: <i>continuant</i></p> <p>Imagem: <i>occurent</i></p> <p>Exemplo de uso: <i>plasma-extractors participates_in process of separation of the plasma this blood clot participates_in this blood coagulation</i></p>

Na etapa 6 foi realizada a avaliação do conteúdo ontológico da *Hemonto* com base no conjunto de critérios para validação e verificação ontológicas. Tais critérios abrangem aspectos relativos ao compromisso ontológico, especificação da ontologia, validação por especialistas, expansibilidade, completude, integridade, consistência, precisão e documentação.

A documentação da *Hemonto* foi construída ao longo de todo processo de desenvolvimento e concluída na etapa 7, visando produzir um documento formal. Já na etapa 8, realizou-se a disponibilização da ontologia por meio da linguagem OWL/XML, obtida pela exportação através do Protégé. Além disso, usando o plug-in OWL-Doc deste editor de ontologias, gerou-se um documento em formato textual da *Hemonto* para acesso geral de seu conteúdo por parte de quaisquer tipos de usuários.

ONTOLEGIS

A ideia para a construção de uma ontologia para representação da informação legislativa sobre o Direito Médico – a *Ontolegis* – emergiu da necessidade de se organizar o conhecimento altamente especializado das leis brasileiras que regulamentam o exercício legal da medicina, e que devido ao seu volume excessivo e a terminologia empregada, afetam sua aplicação prática (TORRES, 2017). Uma ontologia em tal domínio contribui para organização da informação jurídica relacionada e, consequentemente, para a transparência das ações do Estado em uma sociedade democrática.

De maneira mais específica, França (2014) destaca que os temas principais tratados no Direito Médico são: o exercício legal da medicina, a prescrição no processo administrativo, o segredo médico, o erro médico, a omissão de socorro, a responsabilidade médica, o Código de Defesa do Consumidor, os transplantes de órgãos e tecidos, o aborto e a eutanásia; os quais são representados por entidades do domínio jurídico dentro da *Ontolegis*.

Torres (2017) observa que a escolha da *OntoForInfoScience* para o desenvolvimento da *Ontolegis* se deu em função de tal metodologia

apresentar maior detalhamento de etapas para a construção de uma ontologia, ser direcionada para cientistas da informação, estar em português e fazer referência à ontologia de fundamentação BFO.

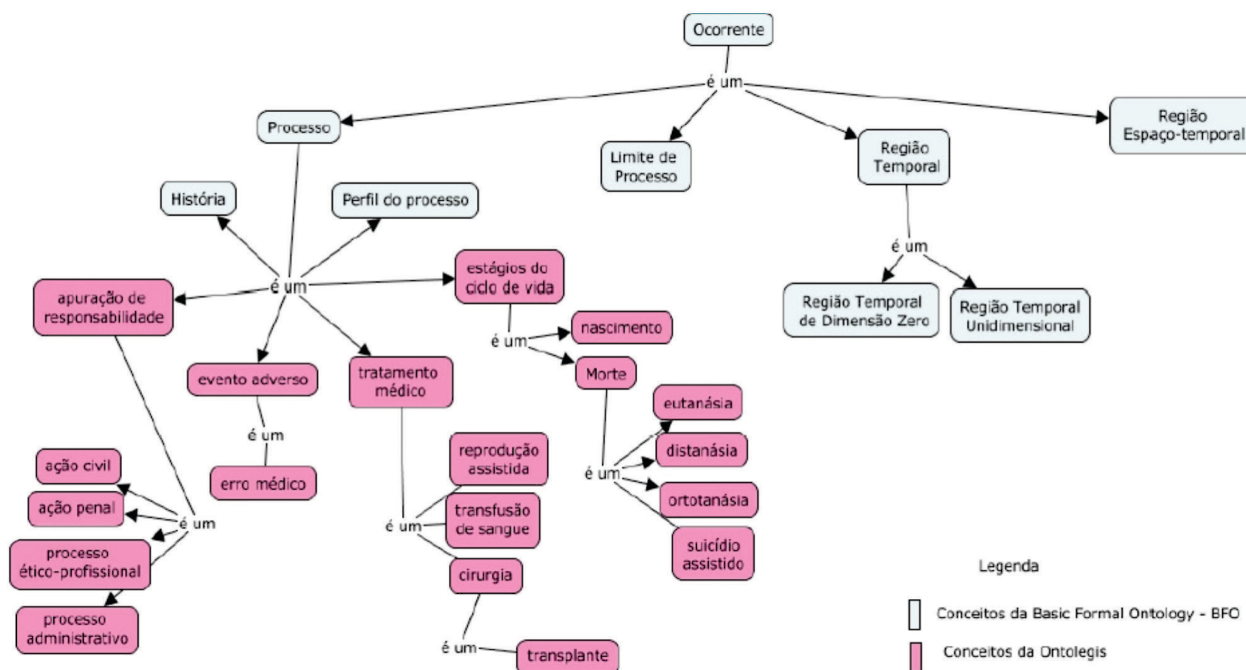
Seguindo as etapas e atividades prescritas na *OntoForInfoScience*, a *Ontolegis* foi especificada (etapa 1) utilizando o *template* de especificação da metodologia, o qual define informações sobre a ontologia tal como seu propósito geral: a *Ontolegis* tem por objetivo possibilitar a sistemas de informação, pacientes, profissionais de saúde e operadores do Direito respostas às questões jurídicas sobre as relações existentes entre médicos e pacientes, planos de saúde, clínicas, hospitais, laboratórios e políticas públicas de saúde (TORRES, 2017).

Para aquisição do conhecimento do domínio tratado (etapa 2), foi formado um *corpus* de textos relevantes ao Direito Médico extraído da legislação disponível no portal do Conselho Federal de Medicina (CFM), o qual engloba as leis: 12.845, 12.842, 11.105/2005, 11000/2004, 10211/2001, 9656/1998, 9436/1997, 9434/1997, 9263/1996, 8501/1992, 8080/1990, 8142/1990, 6839/1980, 6815/1980, 6681/1979, 5991/1973, 3999/1961, 3268/1957, 4113/1942. A extração do conhecimento do *corpus* formado foi realizada combinando a análise de assunto (uma análise humana) com o uso de uma ferramenta de extração automática – o software Sobek⁷.

A etapa de conceitualização (etapa 3) da *Ontolegis* produziu os seguintes artefatos de representação: (i) uma tabela de conceitos e propriedades; (ii) uma tabela de conceitos e valores; (iii) um dicionário de verbos; e (iv) um conjunto de modelos conceituais gráficos do domínio do Direito Médico, que incluem, por exemplo, taxonomias e mapas conceituais. A figura 4, a seguir, apresenta a taxonomia das entidades ocorrentes da ontologia, tal que os conceitos em branco foram importados da BFO-PT e aqueles em rosa são específicos da *Ontolegis*.

⁷ Software gratuito de mineração de textos desenvolvido pela Universidade Federal do Rio Grande do Sul (UFRGS), disponível em <http://sobek.ufrgs.br/about.html>. Acesso em 21 de junho de 2016.

Figura 4 – Taxonomia das entidades ocorrentes da *Ontolegis*



Fonte: Torres (2017).

Na etapa de fundamentao ontolgica (etapa 4), foram avaliadas trs ontologias de fundamentao para reutilizao: a Basic Formal Ontology (BFO), a Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) e o International Committee for Documentation – Conceptual Reference Model (CIDOC-CRM), tendo sido escolhida a BFO para o desenvolvimento da *Ontolegis*. Torres (2017) destaca tambm que houve dificuldade inicial com a aplicao da BFO, pelo fato de ela ser apresentada em ingls e a *Ontolegis* em portugus, j que a ltima é voltada à informao legislativa brasileira. Para resolver a questo, a autora criou uma verso da BFO em portugus denominada BFO-PT.

A etapa de formalizao da *Ontolegis* (etapa 5) foi conduzida com a realizao das seguintes atividades: (i) importao do arquivo OWL da ontologia BFO 2.0 utilizando o Protégé 5.0; (ii) traduo do contedo OWL importado para a lngua portuguesa criando a BFO-PT; (iii) construo da taxonomia geral da *Ontolegis*, utilizando a ferramenta *OntoBee*⁸

para pesquisa às classes do domnio em repositrios ontolgicos existentes e informaes coletadas de dicionrios especializados e vocabulrios controlados (Descritores em Cincias da Saude (DeCS), EuroVoc e tesouro das Naes Unidas) para complementar as informaes dos conceitos/ classes do domnio; (iv) definio das propriedades descritivas das classes da ontologia atravs do atributo *annotations* do Protégé; (v) criao de 29 tipos de relaes semnticas do domnio de Direito Mdico, tal que 17 foram reutilizadas da Relation Ontology (RO) e outras 12 criadas como relaes especficas do domnio tratado (ver exemplo da tabela 3, a seguir); (vi) especificao das informaes descritivas de cada relao criada a partir do atributo *object properties* do Protégé.

⁸ Disponvel em <http://www.ontobee.org/>. Acesso em 28 de Junho de 2017.

Tabela 3 – Exemplos de relações da *Ontolegis*

Relação	Exemplo de uso
É assistido por	Paciente é assistido por hospital.
É autorizado por	Doação de órgãos é autorizado por pacientes ou familiares.
É fiscalizado por	Médico é fiscalizado por Conselho Federal de Medicina.

A etapa seguinte consistiu na avaliação (etapa 6) do conteúdo ontológico da *Ontolegis*, utilizando como base o conjunto de critérios para validação e verificação ontológicas da metodologia *OntoForInfoScience*. Entre os critérios e sua avaliação na ontologia temos, por exemplo: (i) conectividade do domínio e contexto: todos conceitos incluídos na *Ontolegis* estão conectados ao domínio do Direito Médico, com exceção de algumas classes importadas da BFO; (ii) consulta aos especialistas: não houve o envolvimento direto de especialistas no processo de validação da *Ontolegis*, a não ser por meio de consulta a textos especializados.

A documentação da *Ontolegis* (etapa 7) foi criada no decorrer do desenvolvimento da ontologia, tal que seu conteúdo atual engloba um total de 122 classes, 142 subclasses e 1431 registros de anotações (*annotations*) das classes e relações da ontologia.

Por fim, Torres (2017) afirma que a etapa de disponibilização (etapa 8) da *Ontolegis* não foi realizada, uma vez que a ontologia se encontra em um estágio de teste de investigação científica, ainda não colocada em prática para aplicação de seus usuários.

CONSIDERAÇÕES FINAIS

Sabe-se que o processo de construção de uma ontologia é complexo e exige do seu desenvolvedor uma compreensão apropriada do domínio a ser representado, além de conhecimentos técnicos em modelagem conceitual, formalismos lógicos e alguns fundamentos filosóficos. Nesse sentido, um dos principais problemas que profissionais e pesquisadores em ciência da informação enfrentam

no desenvolvimento de ontologias é a presença de termos técnicos provenientes da computação, lógica e filosofia. O presente artigo descreveu o processo de construção de ontologias a partir da metodologia *OntoForInfoScience*, que tem como diferencial o detalhamento das atividades de construção de ontologia além do que é apresentado em metodologia tradicionais. Os resultados práticos desse processo de construção foram apresentados através de duas ontologias de domínio: *Hemonto* e *Ontolegis*.

Em uma análise geral da utilização da *OntoForInfoScience* no desenvolvimento das ontologias apresentadas pode-se afirmar que ela mostrou-se eficiente quanto ao propósito de gerar uma representação formal adequada em ambos os domínios. Tal metodologia utiliza uma linguagem simples, que permite a compreensão das atividades executadas ao longo do processo de construção, sem perder a capacidade de expressividade do uso de linguagens lógicas.

Algumas funcionalidades da *OntoForInfoScience* que são úteis e contribuem para o desenvolvimento de ontologias são: i) explicação sobre o significado e o preenchimento das propriedades descritivas e lógicas das classes e relações; ii) incentivo o uso de ontologias de fundamentação como ponto de partida do desenvolvimento ontológico; iii) detalhamento do processo de criação de definições textuais e formais das classes de uma ontologia; iv) explicação sobre o significado de uma relação ontológica e de que forma é possível caracterizar os tipos de relação ontológica para usá-los adequadamente na representação do domínio; v) incentivo à colaboração como um aspecto essencial na etapa de conceitualização, com o uso de ferramentas colaborativas; vi) explicação de formas para pesquisa por termos relacionados ao domínio para importação na ontologia.

Considerando tais funcionalidades e sua aplicabilidade prática, a metodologia *OntoForInfoScience* foi de grande utilidade para os propósitos de desenvolvimento tanto da *Hemonto* quanto da *Ontolegis*.

Por limitaes de espao, o aspecto mais importante da *OntoForInfoScience* – o detalhamento das atividades de desenvolvimento ontolgico - no pode ser apresentado em sua totalidade aqui, estando disponvel em Mendonça (2015).

Quanto ao contedo da *Hemonto* e seus resultados parciais, pode-se afirmar que a ontologia desenvolvida caracteriza-se por conter uma representao formal mais abrangente no domnio do sangue do que aquela disponvel nas terminologias biomdicas, tais como: FMA, GALEN, UMLS, GO, dentre outras. Ainda assim, a *Hemonto* uma ontologia em desenvolvimento, devido ao seu amplo escopo de cobertura. Sabe-se que o processo de construo de uma ontologia uma processo iterativo e sua validao em diferentes comunidades de especialistas, diferentes corpos mdicos, fundamental para avaliao de sua utilidade prtica. Nesse sentido, a *Hemonto* precisa de maior validao ps-desenvolvimento entre especialistas da rea, alm de revises e atualizaes em seu contedo ontolgico.

No que diz respeito a *Ontolegis*, ela representa importante iniciativa para organizao do conhecimento altamente especializado das leis brasileiras que regulamentam o exerccio legal da medicina, inserido no mbito do Direito Mdico. Entretanto, como tal ontologia ainda se encontra em um estgio de investigao cientfica e no aplicada na prtica, faz-se necessrio a continuidade do desenvolvimento da *Ontolegis*, especialmente no que se refere ao seu processo de axiomatizao. Tal medida ir contribuir para melhoria do nvel de formalizao da ontologia e permitir que ela sirva de instrumento de indexao, recuperao de informao e minerao de textos dos documentos legislativos da rea de Direito Mdico. Cabe ressaltar tambm que a construo da *Ontolegis*, a partir da *OntoForInfoScience*, contribuiu de maneira importante para testar a aplicabilidade prtica de tal metodologia, conduzida em um domnio diferente do biomdico e por pesquisadores que no so os prprios autores da metodologia, como o caso da *Hemonto*.

Por fim, ressalta-se como um dos propsitos principais da *OntoForInfoScience* ser um instrumento de apoio na expanso das possibilidades de desenvolvimento de ontologias, estendendo-a a profissionais que no so engenheiros do conhecimento. Apesar desse propsito, sabe-se que uma soluo mais abrangente e completa para as dificuldades dos cientistas da informao, no que se refere ao uso de termos tcnicos e questes lgicas e filosficas inerentes ao desenvolvimento de ontologias, passa necessariamente pela discusso e insero de requisitos de formao tcnica-acadmica nas disciplinas relacionadas ao tema em cincia da informao. Trabalho este que o bastante amplo e ultrapassa o escopo desta pesquisa.

REFERNCIAS

- AMERICAN ASSOCIATION OF BLOOD BANKS – AABB. *Technical manual*. 17th ed. Maryland, 2011.
- BITTNER, T.; DONNELLY, M. Logical properties of foundational relations in bio-ontologies. *Artificial Intelligence in Medicine*, v. 39, n. 3, p. 197-216, 2007.
- BREITMAN, K. *Web semntica: a internet do futuro*. Rio de Janeiro: LTC, 2005.
- CARDOSO, J. The semantic web vision: where are we?. *IEEE Intelligent Systems*, p. 22- 26, Sept./Oct. 2007.
- CEUSTERS, W. et al. Mistakes in medical ontologies: where do they come from and how can they be detected?. *Studies Health Technology Information*, n. 102, p. 145-164, 2004.
- DE NICOLA, A.; MISSIKOFF, M.; NAVIGLI, R. A software engineering approach to ontology building. *Information Systems*, n. 34, p. 258-275, 2009.
- FERNÁNDEZ, M. et al. Building a chemical ontology using methontology and the ontology design environment. *Intelligent Systems*, v. 14, n. 1, p. 37-46, Jan./Feb.1999.
- _____; CORCHO, O. Methodologies and methods for building ontologies. In: _____. *Ontological engineering*. London: Springer, 2004. P. 107-153.
- FRANÇA, GENIVAL V. Apresentao à 1a edio. In: DIREITO mdico. 3. ed. Rio de Janeiro: GZ Editora, 2014.

Construindo ontologias com a metodologia ontoforinfoscience:
uma abordagem detalhada das atividades do desenvolvimento ontológico

- FREITAS, F.; SCHULZ, S. Pesquisa de terminologias e ontologias atuais em biologia e medicina. *RECIIS*, v. 3, n. 1, p. 8-20, mar. 2009.
- FUNDAÇÃO CENTRO DE HEMATOLOGIA E HEMOTERAPIA DE MINAS GERAIS - HEMOMINAS. *Home*. 2017. Disponível em: <<http://www.hemominas.mg.gov.br/>> Acesso em: 26 jun. 2017.
- GÓMEZ-PÉREZ, A.; FERNÁNDEZ, M.; VICENTE, A. J. Towards a method to conceptualize domain ontologies. In: ECAI WORKSHOP ON ONTOLOGICAL ENGINEERING, 1996, Budapeste. *Proceedings...* [S.l.: s.n.], 1996.
- GREER, J. P. et al. *Wintrobe's clinical hematology*. 12th ed. Philadelphia: Lippincott Williams & Wilkins, 2009.
- GRENON, P.; SMITH, B. SNAP and SPAN: towards dynamic spatial. *Spatial Cognition & Computation*, v. 4, n.1, p. 69-104, 2004.
- GRUNINGER, M.; FOX, M. S. Methodology for the design and evaluation of ontologies. In: WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWLEDGE SHARING, 1995. *Proceedings...* [S.l.: s.n.], 1995.
- ICCBB. ISBT 128 Standard. *Standard Terminology for Blood, Cellular Therapy and Tissue Product Descriptions*, v. 3, n. 33, Jan. 2010.
- JONES, D.; BENCH-CAPON, T.; VISSER, P. *Methodologies for ontology development*. 1998. Disponível em: <<http://cweb.inria.fr/Resources/ONTOLOGIES/methodo-for-ontodev.pdf>>. Acesso em: 13 jun. 2017.
- KETT, C.; ARTALE, A. Representing and reasoning over a taxonomy of part-whole relations. *Applied Ontology*, v. 3, n. 1-2, p. 91-110, 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 (Hemonto)*. 2015. Tese (Doutorado)- Universidade Federal de Minas Gerais, Belo Horizonte, Brasil, 2015.
- MUNN, K.; SMITH, B. *Applied ontology: an introduction*. Heusenstamm, Germany: Ontos Verlag, 2008.
- NOY, N. F.; MCGUINNESS, D. L. *Ontology Development 101: a guide to creating your first ontology*: Stanford Knowledge Systems Laboratory Technical report KSL-01-05 and Stanford Medical Informatics technical report SMI-2001-0880. [S.l.: s.n.], 2001.
- POVEDA-VILLALÓN, M.; SUÁREZ-FIGUEROA, M. C.; GÓMEZ-PÉREZ, A. A double classification of common pitfalls in ontologies. In: WORKSHOP ON ONTOLOGY QUALITY AT THE INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT, 17., 2010, Portugal. *Proceedings...* P. 1-12. 2010.
- REED, S. L.; LENAT, D. B. *Mapping ontologies into Cyc*, 2002. Disponível em: <http://www.cyc.com/doc/white_papers/mapping-ontologies-into-cyc_v31.pdf>. Acesso em: 13 jun. 2017.
- ROSSE, C.; MEJINO, J. L. V. A reference ontology for biomedical informatics: the foundational model of anatomy. *Journal of Biomedical Informatics*, v. 36, p. 478-500, 2003.
- SCHULZ, S.; KUMAR, A.; BITTNER, T. Biomedical ontologies: what part-of is and isn't. *Journal of Biomedical Informatics*, v. 39, p. 350-361, 2006.
- SILVA, D. L.; SOUZA, R. R.; ALMEIDA, M. B. Ontologias e vocabulários controlados: comparação de metodologias para construção. *Ciência da Informação*, v. 37, n. 3, p. 60-75, set./dez. 2008.
- SMITH, B. *Ontology and information systems*, 2004. Disponível em: <<http://www.ontology.buffalo.edu/ontology/>> Acesso em: 13 jun. 2017.
- _____; CEUSTERS, W. Ontological realism: a methodology for coordinated evolution of scientific ontologies. *Applied Ontology*, v. 5, p. 139-188, 2010.
- _____. et al. Relations in biomedical ontologies. *Genome Biology*, v. 6, 2005.
- _____. The OBO foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, v. 25, n. 11, p. 1251-1255, 2007.
- SOUZA, C.; PEREIRA, C.; SOARES, A. Collaborative elicitation of conceptual representations: a corpus-based approach. In: ADVANCES in information systems and technologies. Heidelberg: Springer Berlin Heidelberg, 2013. P. 111-124.
- SUÁREZ-FIGUEROA, M. C. et al. *NeOn D5.4.1: NeOn methodology for building contextualized ontology networks*. Disponível em: <<http://www.neon-project.org/>>. Acesso em: fev. 2008.
- SURE, Y.; STAAB, S.; STUBER, R. *On-To-Knowledge Methodology (OTKM)*. 2003. Disponível em: <http://www.sfu.ca/~mhatala/iat881/papers/2003_ontohandbook_otkmethodology.pdf>. Acesso em: 20 jun. 2017.
- TORRES, SIMONE. *Modelagem de domínios em Sistemas de Organização do Conhecimento (SOC): uma investigação em tesauros e ontologias para a informação legislativa*. 2017. Tese (Doutorado em Ciência da Informação)- Universidade Federal de Minas Gerais, Belo Horizonte, 2017.
- USCHOLD, M.; GRUNINGER, M. Ontologies: principles, methods and applications. *Knowledge Engineering Review*, v. 11, n. 2, p. 93-136, 1996.
- _____; KING, M. *Towards a methodology for building ontologies*. 1995. Disponível em: <<http://citeseer.ist.psu.edu/uschold95toward.html>> Acesso em: 20 jun. 2017.

Estudo comparativo das estruturas semânticas em diferentes sistemas de organização do conhecimento

Gercina Ângela de Lima

Pós-doutorado pela Universidad Carlos III (UC3M) - Madrid, Espanha. Pós Doutorado pela Universidade de São Paulo (USP) - São Paulo, SP - Brasil. Doutora em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG). Professora da Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/3183050056105009>

E-mail: limagercina@gmail.com

Benildes Coura Moreira dos Santos Maculan

Doutora em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG) - Brasil. Professora da Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/5336218259257800>

E-mail: benildes@gmail.com

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

RESUMO

A elaboração de Sistemas de Organização do Conhecimento (SOCs) pressupõe o estabelecimento de relações semânticas entre seus conceitos, visando à criação de uma estrutura conceitual coerente. Diferentes SOCs possuem distintos níveis de representação relacional, que são dependentes do tipo de estrutura que cada instrumento requer para representar a base do conhecimento. Nota-se, entretanto, que existe um problema no entendimento sobre quais são os objetivos e propósitos de diferentes tipos de SOCs, principalmente sobre sua elaboração e uso. Este artigo pretende apresentar um estudo comparativo sobre a construção de SOCs do tipo taxonomia, sistema de classificação, tesouros e ontologias, a partir da apresentação de uma análise prática, com o intuito de demonstrar as características de suas estruturas conceituais, usos, potencialidades e limitações, verificando e demonstrando como as relações semânticas vão sendo incorporadas aos instrumentos.

Palavras-chave: Sistemas de Organização do Conhecimento. Relações semânticas. Estrutura conceitual.

Comparative study of semantic structures in different systems of knowledge organization

ABSTRACT

The elaboration of Knowledge Organization Systems (KOS) is based on the establishment of semantic relations between their concepts, aiming the creation of a coherent conceptual structure. Different KOS have different levels of relational representation, which are dependent on the type of structure that each tool requires to represent the knowledge base. It is observed, however, that there is a problem in the comprehension of what are the objectives and purposes of different types of KOS, especially their elaboration and use. This paper intends to present a comparative study on the construction of KOS about the type of taxonomy, classification system, thesauri and ontologies, from the presentation of a practical analysis. It has the purpose to demonstrate the characteristics of KOS conceptual structures, uses, potentialities and limitations, verifying and demonstrating how the semantic relationships are being incorporated into the tools.

Keywords: Knowledge Organization Systems. Semantic relationships. Conceptual structure.

Estudio comparativo de estructuras semánticas en diferentes sistemas de organización del conocimiento

RESUMEN

La elaboración de Sistemas de Organización del Conocimiento (SOCs) presupone el establecimiento de relaciones semánticas entre sus conceptos, buscando la creación de una estructura conceptual coherente. Diferentes SOCs poseen distintos niveles de representación relacional, que son dependientes del tipo de estructura que cada instrumento requiere para representar la base del conocimiento. Se observa, sin embargo, que existe un problema en el entendimiento sobre cuáles son los objetivos y propósitos de diferentes tipos de SOCs, principalmente sobre su elaboración y uso. Este artículo pretende presentar un estudio comparativo sobre la construcción de SOCs del tipo taxonomía, sistema de clasificación, tesauros y ontologías, a partir de la presentación de un análisis práctico, con el propósito de demostrar las características de sus estructuras conceptuales, usos, potencialidades y limitaciones, verificando y demostrando cómo las relaciones semánticas se van incorporando a los instrumentos.

Palabras clave: *Sistemas de Organización del Conocimiento. Relaciones semánticas. Estructura conceptual.*

INTRODUÇÃO

Sistemas de Organização do Conhecimento (SOCs), tais como taxonomias, sistemas de classificação, tesauros e ontologias, são instrumentos de representação do conhecimento; têm sido estudados na ciência da informação (CI) para organizar e recuperar informações, assim como para outros propósitos. Assim, na área de CI, entender como esses instrumentos são conceitualmente construídos auxilia a compreensão sobre o tema da organização do conhecimento.

Alguns estudiosos (SOERGEL, 1997; VICKERY, 1997; GILCHRIST, 2003; KLESS; SIMON; KAZMIERCZAK, 2012; KLESS et al., 2015; ALMEIDA, 2013) têm realizado estudos comparativos entre ontologias e outros tipos de SOCs, porém a maioria deles acontece em relação ao tesauro. Pressupõe-se que isso ocorra porque o tesauro, assim como a ontologia, pode ser utilizado como base de conhecimento de um domínio. Além disso, o tesauro parece possuir também semelhanças com as ontologias na forma em que as relações semânticas são estruturadas, em especial, no que diz respeito às relações *é-um* e *parte-de* nas ontologias e algumas relações hierárquicas (BT/NT) nos tesauros. Entretanto, há pouca literatura que compara ontologias e os outros tipos de instrumentos.

No âmbito da formação de alunos do curso de biblioteconomia, há disciplinas que trabalham com os diferentes tipos de SOCs. Contudo, na formação em pós-graduação para a área de ciência da informação, é usual ter-se alunos de outras diferentes áreas do conhecimento (computação, engenharia, sistemas, entre outros), que contam apenas com os fundamentos conceituais e com as definições existentes encontradas na literatura e em normas de construção dos instrumentos, para os quais não é simples o entendimento das características conceituais dos diferentes tipos de SOCs. Assim, há a necessidade de se fazer conhecer as características das estruturas conceituais, potencialidades e limitações desses instrumentos pelos profissionais da área.

Para suprir essa lacuna, foi desenhada a disciplina *Representação da Informação*, que vem sendo aplicada a diferentes turmas desde 2013. Ela tem como objetivo estudar a modelagem conceitual em diferentes SOCs, analisando a representação das estruturas conceituais, sobretudo quanto às relações semânticas estabelecidas, visando identificar elementos comparativos entre eles. Para viabilizar esse empreendimento, a proposta da disciplina é um trabalho prático de desenvolvimento de quatro diferentes instrumentos (taxonomia, sistema de classificação, tesauros e ontologias) de representação do conhecimento.

Partiu-se do pressuposto de que, por meio da prática, em conjunto com a fundamentação teórica, e à medida que os alunos fossem realizando as etapas da modelagem de cada instrumento, suas características ficariam mais evidentes.

Este artigo apresenta uma compilação de análises observadas a partir de uma amostra selecionada dentre os trabalhos apresentados pelos alunos, com exemplos específicos que permitiram uma comparação de construção das estruturas desses instrumentos. Demonstram-se, também, as características das estruturas conceituais e as relações semânticas que ocorrem em cada um dos quatro SOCs, assim como os procedimentos metodológicos empregados na construção deles.

Não é objetivo deste artigo aprofundar na descrição dos instrumentos, mas demonstrar o nível de representação relacional que acontece no momento da sua construção, apresentando uma análise comparativa quanto às suas semelhanças e diferenças, com apontamento para possibilidades de usos e limitações.

SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO

Para Vickery (2012, p. 203), os SOCs são “instrumentos complementares que ajudam o usuário a encontrar seu caminho no texto”, e, como ferramentas auxiliares, têm a finalidade de facilitar uma pesquisa genérica ou mais específica, assim como a escolha dos termos de pesquisa. Os princípios teóricos para a elaboração de SOCs são advindos das teorias da classificação, da teoria do conceito e da terminologia, que orientam os agrupamentos de conceitos pelas suas características, de acordo com seus graus de semelhanças e diferenças. As bases da classificação remontam a Aristóteles, em que um todo pode ser dividido em partes, sucessivamente, formando classes e subclasses. Essa lógica da divisão para representar o conhecimento parte do princípio classificatório de gênero e espécie para formar uma estrutura conceitual hierárquica para representar o conhecimento.

O núcleo principal desse processo é a escolha dos conceitos e seus relacionamentos e sob quais aspectos se quer representar determinado domínio. Pode-se dizer, então, que os conceitos são a base para a construção da estrutura conceitual, e são as relações entre eles que os mantêm reunidos dentro de um determinado domínio e princípio de representação. Para Green (2008), os relacionamentos são o coração da organização do conhecimento:

Tentamos localizar informações relacionadas para a necessidade de um usuário; tentamos localizar documentos contendo palavras que possam ser usadas para transportar a informação procurada; e/ou recuperamos documentos indexados por um descritor de assunto relevante ou classificado em uma categoria de assunto relevante. Usamos relacionamentos para navegar entre assuntos. Ajudamos a localizar documentos que juntos dizem mais do que a soma de suas partes (GREEN, 2008, p. 158, tradução nossa).

Existem vários tipos de relações semânticas (ISO 25964-1, 2011), porém, neste trabalho, vamos considerar os três tipos básicos utilizados na estruturação de um SOC: (1) de equivalência, que ocorrem entre termos sinônimos (total; parcial, etc.); (2) hierárquicas, com níveis de superordenação e subordinação, formando renques¹ e cadeias²; (3) associativas, entre termos que mantêm algum tipo de ligação que não de superordenação e subordinação.

A expressividade semântica dos relacionamentos e a forma de representá-los variam segundo o sistema a ser construído, porém, os quatro sistemas (taxonomia, sistemas de classificação, tesauro e ontologia), descritos na próxima seção neste artigo contemplam a noção de hierarquia entre conceitos.

¹ O renque é constituído de conceitos subordinados a um mesmo conceito, ou seja, conceitos coordenados; são conceitos ‘irmãos’. O renque é, portanto, uma série horizontal de conceitos. Tutorial Elaboração de tesauro documentário. Disponível em: <<http://www.conexaorio.com/bitit/tesauro/>>. Acesso em: 10 ago. 2017.

² Como as cadeias são séries verticais de conceitos, é necessário estabelecer a ordem em que cada conceito deverá estar em relação aos outros conceitos. Tutorial Elaboração de tesauro documentário. Disponível em: <<http://www.conexaorio.com/bitit/tesauro/>>. Acesso em: 10 ago. 2017.

ESTRUTURA E RELAÇÕES NOS SOCS

Nesta seção, serão descritas as características estruturais dos SOCs, na seguinte ordem: (1) taxonomia, (2) sistemas de classificação, (3) tesouro e (4) ontologia.

ESTRUTURA E RELAÇÕES EM TAXONOMIAS

A norma ISO 25964-2 (2013) sugere que a unidade básica de uma taxonomia seja denominada “categoria”, porém, na literatura, ela é também designada como “rótulo” (MARTINEZ et al., 2004, p. 106; MOREIRO GONZÁLEZ, 2011, p. 52), nome que será adotado neste trabalho. A taxonomia é composta de um conjunto de termos arranjados em hierarquias representando o domínio modelado, que organiza conceitos segundo as suas semelhanças e diferenças. Entre suas funções, estão o mapeamento do conhecimento de um domínio e o estabelecimento de um rótulo para as informações disponibilizadas, sendo um mecanismo de acesso em ambiente digital.

Uma taxonomia: (a) possui uma lista estruturada de termos, que representam conceitos; (b) apresenta apenas relações hierárquicas (gênero/espécie e todo-parte); (c) não oferece definições dos termos; (d) permite navegação (*navigation*) e busca (*browsing*) por tópicos (rótulos) e, assim, a organização e a recuperação de informação (*searching*) em bibliotecas digitais, portais institucionais, corporativos e menus de e-commerce, seja na Web seja em intranets; (e) comporta agregação de metadados e dados; (f) proporciona um modelo conceitual e mapa conceitual do domínio modelado; (g) não inclui códigos (notações), que são mais usuais em sistemas de classificação (CAMPOS; GOMES, 2007; MACULAN et al., 2009; MACULAN; LIMA; PENIDO, 2011). Além disso, ela pode ser customizada para refletir a terminologia, a cultura, as metas e os objetivos de uma organização, para a gestão organizacional (ISO 25964-2, 2013).

Lambe (2007) ainda ressalta que a taxonomia hierárquica é mais usual, mas podem ser criados outros tipos de estruturas: a poli-hierárquica ou a facetada. A última segue os mesmos

princípios utilizados na construção de sistemas de classificação facetados e de tesouros, a partir da Teoria da Classificação Facetada desenvolvida por Ranganathan (1967). Contudo, para atender ao escopo deste estudo, será considerado o tipo de taxonomia mais usual, o hierárquico.

Uma taxonomia é um instrumento pré-coordenado, pois estabelece a formação dos vários rótulos (assuntos simples ou compostos) no momento de sua construção. Logo, cada rótulo pode ser representado por um conceito único, por exemplo: “Direitos humanos”; pela combinação de conceitos: “Governo, cidadãos e direitos”; uma frase ou expressão: “Enfrentando câncer”, “Busca por ensaios clínicos?” ou “Cuide-se no sol”, de forma que seja mais intuitivo, comunicativo ou que antecipe a demanda do usuário (ISO 25964-2, 2013, p. 61). A norma ainda aponta que os rótulos podem não ser um assunto dentro de uma categoria: (a) público destinado: “Crianças e jovens”; (b) tempo ou local: “Arquivo”; (c) formato do documento: “comunicado de imprensa”; (d) uma tarefa: “Apresentação de reivindicação”, e que haja o uso concomitante desses tipos de rótulos.

No âmbito da ciência da informação, Vickery (1975, p. 10) definiu a taxonomia como um instrumento utilizado para “organizar todas as entidades de um universo em uma simples hierarquia: uma grande árvore na qual cada entidade pertence apenas a uma única classe, cada espécie a um único gênero, e assim por diante”. Ademais, Martinez et al. (2004) afirmam que, em um sentido amplo, a taxonomia cria ordem e rótulos que auxiliam localizar a informação relevante, e, de forma mais específica, “é o ordenamento e rotulação de metadados, que permite organizar sistematicamente a informação primária” (MARTINEZ et al., 2004, p. 106). Assim, uma taxonomia pode atender a objetivos diversos e em qualquer domínio e contexto de uso.

De modo geral, as relações empregadas na modelagem de uma taxonomia são as relações hierárquicas. Segundo Campos e Gomes (2007), a taxonomia de tipo canônica vai estabelecer relações de gênero-espécie, em unidades sistemáticas. Para

as autoras, nas taxonomias de domínio e nas de processos e tarefas gerenciais, aplica-se o “princípio classificatório policotômico [...] [que] possibilita uma organização que representa classes de conceitos com um princípio de divisão (coisas e seus tipos, processos e seus tipos), não priorizando nenhum dos aspectos” (*idem*, p. 5). Assim, as autoras afirmam que resultará em uma simples hierarquia de especificação, representando diversos aspectos do conhecimento modelado (períodos temporais, fenômenos, objetos, fatos, tarefas, atividades).

ESTRUTURA E RELAÇÕES EM SISTEMAS DE CLASSIFICAÇÃO BIBLIOGRÁFICA

Os sistemas de classificação são construídos a partir de signos estruturados e normalizados, podendo ser gerais, quando buscam cobrir todos os campos do conhecimento, ou especializados, quando representam um assunto específico. Ademais, os sistemas de classificação podem ser analíticos, quando sistematizam fenômenos físicos, viabilizando seu entendimento, ou podem ser documentais (de assuntos ou bibliográficos) e implicam a classificação de documentos, em qualquer tipo de suporte e mídia, visando facilitar, principalmente, a localização do item em um dado acervo.

A estrutura dos sistemas de classificação usualmente é constituída por: (1) um esquema de classificação, agrupando elementos por suas semelhanças; (2) uma notação da classificação, que representa os cabeçalhos; e (3) um índice, que facilita a busca do usuário. Segundo Guinchat e Menou (1994), os sistemas de classificação possuem cabeçalhos que permitem representar os conceitos e objetos de um campo determinado do conhecimento, ordenados de forma sistemática, em função de um ou de vários critérios materiais e intelectuais. Portanto, todo sistema de classificação possui uma divisão por classes, que é dividida numa subclasse e, assim, sucessivamente, até chegar ao mais específico (TRISTÃO; FACHIN; ALARCON, 2004). Essa sequência, que respeita a ordem natural das coisas, é composta de classes subordinadas, classes superordenadas e classes coordenadas. A divisão

por classes se baseia em: (1) princípio de divisão: critérios preestabelecidos, obedecendo a uma característica específica de divisão; (2) classes simples e classes compostas: a classe simples identifica uma única característica, na qual não cabem mais subdivisões, e a classe composta identifica mais de uma característica, na qual podem caber mais subdivisões.

Existem diversificados sistemas de classificação de assuntos (ou classificações bibliográficas), dentre os quais se citam: (1) Classificação Decimal de Dewey, de 1876; (2) Classificação Expansiva, de 1891; (3) Classificação da Biblioteca do Congresso, de 1902; (4) Classificação Decimal Universal, de 1905; (5) Classificação de Assuntos, de 1906; (6) Classificação dos Dois Pontos, de 1933; (7) Classificação Bibliográfica, de 1935; e (8) Classificação Internacional, de 1961.

A Classificação dos Dois Pontos, criada por Ranganathan, é um sistema de classificação por facetas, bastante difundido na atualidade. A Teoria da Classificação Facetada criada por ele e a técnica da análise facetada permitem a análise dos assuntos de um documento em facetas, privilegiando os seus diferentes aspectos, podendo-se estabelecer relações entre esses aspectos. Essa característica imprime maior flexibilidade ao sistema de classificação, possibilitando receber novos conceitos e, assim, adequar-se às constantes mudanças e evoluções do conhecimento. De modo geral, entretanto, as principais relações empregadas na modelagem de um sistema de classificação são as mesmas relações hierárquicas estabelecidas em uma taxonomia, já descritas na seção anterior.

ESTRUTURA E RELAÇÕES EM TESAUSOS

Um tesouro é composto de um conjunto de termos descritores (preferidos e não preferidos) relacionados semântica e genericamente entre si, permitindo diferentes tipos de organização (alfabética, por campos semânticos, entre outros). Os termos descritores devem ser padronizados quanto à sua forma e significado, compondo-se de palavras ou grupos de palavras. É um sistema pós-coordenado,

pois adota termos descritores que designam conceitos simples, que podem ser combinados, no processo de indexação, ou combinados pelo usuário no momento da busca. Ressalta-se, porém, que o tesauro pode ter maior ou menor nível de pré-coordenação na formação de descritores.

O tesauro é criado a partir de um conjunto de regras e princípios de contextualização, tendo a função de representar, de forma abreviada, o conhecimento de um domínio, com níveis diferenciados de controle da terminologia e padronização, para dar maior consistência à indexação. Em geral, irá atuar na representação de recursos informacionais, na análise e síntese de textos, assim como na formulação de estratégias de buscas por informações. Não é comum a disponibilização de tesouros aos usuários para a formulação da busca; contudo, esse tipo de uso facilita a recuperação, uma vez que orienta o usuário na escolha dos termos.

Os tesouros possuem três elementos básicos: (1) um léxico; (2) uma rede paradigmática: indica relações essenciais e estáveis entre descritores, numa rede lógico-semântica; (3) uma rede sintagmática: indica relações contingentes entre descritores, válidas apenas em determinado contexto de uso, através de regras de sintaxe. A elaboração de tesouros deve prever a definição de cada um dos termos descritores preferidos e, também, o elemento “Nota de Escopo”, que fornecem orientação sobre seu emprego, significado de um descritor e a relação com outros termos. Os tesouros podem ser apresentados por uma visualização gráfica, em estrutura alfabética sistemática, que permite observar as relações entre os descritores que compõem o grupo (*cluster*) semântico do conceito em questão.

Na elaboração de um tesauro terminológico (ou conceitual), a sua abrangência conceitual será definida pela natureza do serviço, pelo seu objetivo e propósito, levando em consideração o domínio (garantia literária), as especificidades das necessidades informacionais do usuário (garantia do usuário), assim como do seu comportamento de busca e o contexto (garantia da instituição) no qual o serviço é oferecido.

Os fundamentos de construção possuem suas bases na Teoria da Classificação Facetada, Teoria do Conceito e Terminologia.

As relações semânticas estabelecidas em um tesauro são, basicamente: (a) hierárquicas (gênero-espécie; todo-parte); (b) de equivalência (sinônimos); (c) associativas (ligações semânticas). De forma geral, esses relacionamentos são explicitados a partir de símbolos tais como: (1) USE: indica sinonímia e que há outro termo preferencial; (2) UF (usado para): indica sinonímia e que é o termo preferencial; (3) BT: indica a classe mais geral do termo; (4) NT: indica termo específico; (5) RT: indica termo em relações associativas diferentes.

ESTRUTURA E RELAÇÕES EM ONTOLOGIAS

A ontologia é estudada em diversas áreas do conhecimento e tem sua origem na filosofia; e a definição de ontologia como uma ciência se deve a Aristóteles (1978). Na ciência da informação e na ciência da computação, a ontologia é estudada como instrumento de representação do conhecimento, em geral, como um artefato computacional. Nesse contexto, uma ontologia descreve formalmente termos e relações que podem existir entre eles dentro de determinado domínio, inteligível pelo computador.

Os estudos sobre a ontologia ganharam mais destaque no contexto da Web Semântica, por possibilitar o entendimento da linguagem natural, integração de fontes de informação, interoperabilidade semântica e por servir como base de conhecimento em várias aplicações. Um ponto comum existente em todas as ontologias é a dimensão semântica, que é inerente à sua estrutura, sempre com a inclusão de um vocabulário, com as especificações dos termos, possibilitando a identificação das categorias fundamentais do domínio em questão, como esses termos relacionam entre si, e quais relacionamentos serão possíveis de serem criados (GRUNINGER et al., 2008).

Vickery (1997, p. 284) foi o primeiro autor na área de ciência da informação a estudar sobre ontologia; ele a define como “uma formalização sistemática

de definições de conceitos, relacionamentos e regras que capturam o conteúdo semântico de um domínio em um formato legível por máquina.” Já Smith (2003) define ontologia como:

um dicionário de termos formulados em uma sintaxe canônica e com definições comumente aceitas, concebidas para produzir um lexical ou uma estrutura taxonômica de representação do conhecimento que pode ser compartilhada pela comunidades de diferentes sistemas de informações (SMITH, 2003, p. 6, tradução nossa).

Assim, pode-se afirmar que a ontologia é uma lista de conceitos ou entidades dentro de um domínio específico, os quais podem ser estruturados de forma hierárquica, por meio de relações semânticas explicitadas formalmente, em meio informatizado. A estrutura da ontologia nem sempre é a mesma, e, para Gruber (1992), os componentes básicos são: (1) classes, as quais são organizadas em estrutura conceitual hierárquica; (2) relações, que representam a relação entre os conceitos; (3) axiomas, que representam sentenças verdadeiras; (4) instâncias, que representam os dados. Todos esses componentes devem possuir uma representação explicitamente formal para que a ontologia seja processável por máquina.

Existem várias relações que podem ser representadas em uma ontologia, sendo as mais comuns: a relação classe-classe (a relação “é um”), a relação instância-classe (“é instância de”) e a relação instância-instância (a relação “é parte de”). As ontologias são consideradas linguagens mais semanticamente ricas, em comparação com os outros tipos de SOCs, porque todas as relações entre os termos podem ser explicitadas, inclusive de equivalências e as associativas. Além disso, não existem limites para as variações das relações; assim podem ser feitas deduções e inferências que não são possíveis de serem realizadas nos outros SOCs.

Entre as tecnologias utilizadas na construção de ontologias, estão as linguagens *Resource Description Framework* (RDF), *Resource Description Framework Schema* (RDF-S) e *Web Ontology Language* (OWL). A RDF descreve metadados permitindo criar triplas

para fazer o relacionamento sujeito-predicado-objeto, permitindo representar a semântica contida neles. Já o RDF-S define as *tags* especificando as classes, propriedades e relacionamentos que podem ser utilizados nas triplas, que incluem as *tags* e a sua estrutura hierárquica (taxonomia). Já a OWL foi construída a partir da RDF-S e da DARPA *Agent Markup Language* (DAML) + *Ontology Inference Layer* ou *Ontology Interchange Language* (OIL), sendo considerada a linguagem de ontologias na Web Semântica. A OWL é uma linguagem computacional lógica, que possibilita melhor interoperabilidade entre os conteúdos, satisfazendo os formalismos exigidos para a representação dos recursos de informação. Segundo Isotani e Bittencortt (2015), as principais características da OWL são a descrição explícita e formal de conceitos num dado domínio (classes), exibindo suas propriedades, características e atributos, e a possibilidade de expressar o raciocínio lógico da descrição para a Web semântica.

DESCRIÇÃO DOS PROCEDIMENTOS E RESULTADOS

Este estudo se caracteriza como descritivo, que permitiu estabelecer relações de dependência entre variáveis (características dos quatro instrumentos: taxonomia, sistema de classificação, tesouros e ontologias), que serviram de base para as análises sobre as semelhanças e diferenças entre elas. Assim, o estudo tem caráter comparativo, de natureza qualitativa, demonstrando o nível semântico da representação dos relacionamentos entre termos nas estruturas dos quatro instrumentos.

O estudo teve por base as discussões e análises realizadas no âmbito da disciplina Representação da Informação, oferecida no Programa de Pós-Graduação em Ciência da Informação (PPGCI), da Escola de Ciência da Informação da Universidade Federal de Minas Gerais (UFMG). Para o trabalho prático de construção dos quatro instrumentos, foram selecionados, aleatoriamente, cinco documentos acadêmicos do tipo teses e dissertações, defendidos no PPGCI/ECI/UFMG,

para a elaboração de instrumentos de representação do conhecimento, a partir do assunto contido nesses documentos. Definiu-se o público-alvo como sendo alunos de graduação e professores.

A metodologia para os processos classificatórios empregados na construção dos instrumentos foi respaldada na Teoria do Conceito e na Teoria da Classificação Facetada. A técnica da análise facetada foi aplicada utilizando-se os princípios dos três planos propostos por Ranganathan: (1) plano das ideias: (a) análise de assunto e identificação das características que constituem aquele assunto; (b) decisão sobre as classes básicas (mais abstratas) para agrupamento dos assuntos em facetas e subfacetas, realizado a partir dos princípios das cinco categorias Personalidade, Matéria, Energia, Espaço, e Tempo (PMEST) de Ranganathan; (c) subdivisão e agrupamento dos assuntos conforme suas características comuns, separando os de características diferentes; (d) arranjo dos elementos dentro dos grupos e subgrupos; (2) plano verbal: utilizado para determinar o(s) termo(s) que designam cada conceito; (3) plano notacional: utilizado somente na construção do sistema de classificação, criando-se notações simbólicas (com números e/ou letras), de acordo com as regras estabelecidas.

A amostra para a comparação apresentada neste artigo foi um dos trabalhos entre os cinco criados no âmbito da disciplina já mencionada. A escolha da amostra foi intencional, pois se buscou selecionar o trabalho que melhor pudesse exemplificar em sua estrutura as características dos quatro instrumentos construídos.

Em síntese, a metodologia incluiu: (1) seleção do documento a ser analisado; (2) captura de conhecimento: análise no documento para identificação dos assuntos representativos do conteúdo; (3) organização dos assuntos com a subdivisão e agrupamento, alinhados com cada tipo de instrumento; (4) identificação das relações semânticas, em conformidade com a necessidade de cada instrumento; (5) apresentação da estrutura dos instrumentos. A seguir, apresentam-se os procedimentos de construção de cada instrumento.

PERCURSO METODOLÓGICO E ELABORAÇÃO DOS INSTRUMENTOS

1. Seleção do trabalho: para avaliação dos quatro SOCs construídos:

SILVA, A.R. *Estudo dos princípios de categorização na biblioteconomia e ciência da informação*: Ranganathan: entre a teoria clássica e a abordagem cognitiva contemporânea. 2010. 195 f. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2010. Disponível em: <<http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/ECID-8ELHMA>>. Acesso em: 10 ago. 2017.

2. Descrição da construção dos Sistemas de Organização do Conhecimento-SOC

2.1. Construção da taxonomia

2.1.1. (2.1.1) Tipo de taxonomia: de domínio, expressando os assuntos do documento.

2.1.2. Captura de conhecimento: levantamento dos assuntos do documento, com a leitura e a marcação dos termos considerados significativos do conteúdo do documento e o contexto de uso. Resultados da taxonomia: foram levantados 99 termos.

2.1.3. Normalização dos rótulos pré-coordenados: a escolha dos rótulos obedeceu aos seguintes princípios: (a) comunicabilidade: linguagem do usuário (garantia do usuário); (b) utilidade: representativo do documento; (c) estimulação: termos intuitivos e fáceis de entender; (d) compatibilidade: termos alinhados à área de conhecimento do documento.

2.1.4. Criação da estrutura classificatória: mapeamento dos assuntos do documento a partir dos princípios das categorias PMEST e, depois, as

classes e subclasses foram organizadas a partir do plano das ideias e do plano verbal. No primeiro, foram criadas cadeias (hierárquicas genéricas: tipo de; hierarquias partitivas: parte de e todo de) e renques. Para a ordem dos conceitos dentro das cadeias, utilizaram-se os cânones: extensão decrescente (crescente em relação à intensão) e modulação (sequência dos elos). Para os renques, utilizaram-se os cânones: exaustividade, exclusividade (elementos mutuamente exclusivos), sequência útil (do geral para o particular) e sequência consistente (criar classes semelhantes de forma homogênea). A ordenação dos elementos das classes e subclasses foi feita em ordem alfabética. A validação pelo especialista foi realizada pelas autoras deste artigo.

2.2. Construção do sistema de classificação

2.2.1. Captura de conhecimento: foi iniciada a partir da taxonomia já elaborada, conforme os princípios de criação de um sistema classificatório.

2.2.2. Determinação das classes básicas (maior abstração): feita a partir da análise da taxonomia criada, resultando em quatro classes básicas: (a) Entidades compreendem os princípios, objetos imateriais, objetos materiais; (b) Atividades compreendem as operações, estados, processos; (c) Propriedades compreendem as quantidades, qualidades, relações; (d) Dimensões compreendem tempo, posição, espaço.

2.2.3. Criação da notação: inclusão apenas de números ordinais, em ordem crescente, separados por ponto, conforme a endentação. Não houve elaboração de índice, pois não fazia parte do escopo da proposta do trabalho.

2.2.4. Sintaxe para formação de notação: uso do PMEST na ordem de citação.

2.3. Construção do tesauro

2.3.1. Captura de conhecimento: também foi iniciada a partir da taxonomia já elaborada conforme os princípios de criação de um sistema classificatório.

2.3.2. Determinação das classes básicas: utilizaram-se as mesmas quatro classes básicas da taxonomia criada: (a) Entidades; (b) Atividades; (c) Propriedades; (d) Dimensões.

2.3.3. Glossário: foram elaboradas definições para os termos descritores.

2.3.4. Estabelecimento de relações: foram identificadas as relações hierárquicas da taxonomia (gênero-espécie, todo-parte) que foram representadas no tesauro pelas relações *Broader Term* (BT) e *Narrow Term* (NT). Depois, pelo teor das definições, foram estabelecidas as relações associativas utilizando-se o símbolo *Related Term* (RT). Em seguida, com nova análise do documento e consulta a dicionários da área, foram estabelecidas as relações de equivalência, utilizando-se os símbolos *Use For* (UF).

2.3.5. Forma de apresentação: alfabética, juntamente com a parte sistemática.

2.4. Construção da Ontologia

2.4.1. Captura e conhecimento: foi iniciada a partir da taxonomia e do tesauro já elaborados, conforme os princípios de criação de um sistema classificatório.

2.4.2. Determinação das classes de conceitos: utilizou-se a linguagem OWL, por meio da ferramenta Protégé³.

³ A Ferramenta Protégé é um dos softwares mais referenciados e utilizados na construção de ontologias. Foi desenvolvida na Stanford Center for Biomedical Informatics Research na Escola de Medicina da Universidade de Stanford, EUA.

2.4.3. Criação das entidades e estabelecimento de relações: ampliação das relações possíveis entre as classes do domínio, além das relações entre termos definidas pelo tesouro, sendo “use” e “use para”; assim como as propriedades “Termo Geral”, “Termo Específico” e “Termo Relacionado”, foram acrescentadas as relações “requires” e “isRequiredBy”, “isPartOf” e “hasPart”, “isBasedOn” e “isBasesFor”, “references” e “isReferencedBy”. Essas relações são definidas no padrão SCORM, que consiste em um modelo de referência para a construção e publicação de objetos de aprendizagem na web, por meio de sistemas de gerenciamento de conteúdo para educação a distância (NEVES, D.E; BRANDÃO, W. C.; ISHITANI, L., 2016).

2.4.4. Esquematização dos relacionamentos: as classes de conceitos e relações definidas nos itens 2.4.2 e 2.4.3 foram formalmente descritas por meio das propriedades e das relações estabelecidas na linguagem OWL.

2.4.5. Forma de apresentação: uma vez formalizada e descrita, a ontologia pôde ser visualizada na ferramenta *Protegé*, a partir do arquivo interoperável no formato OWL/RDF. Assim, foi possível gerar uma visualização gráfica navegável assim como realizar consultas sobre a ontologia por meio da linguagem SPARQL (*Protocol and RDF Query Language*).

3. Escolha da Classe: amostra para a análise e demonstração das relações semânticas: foi selecionada a Classe: Cognição.

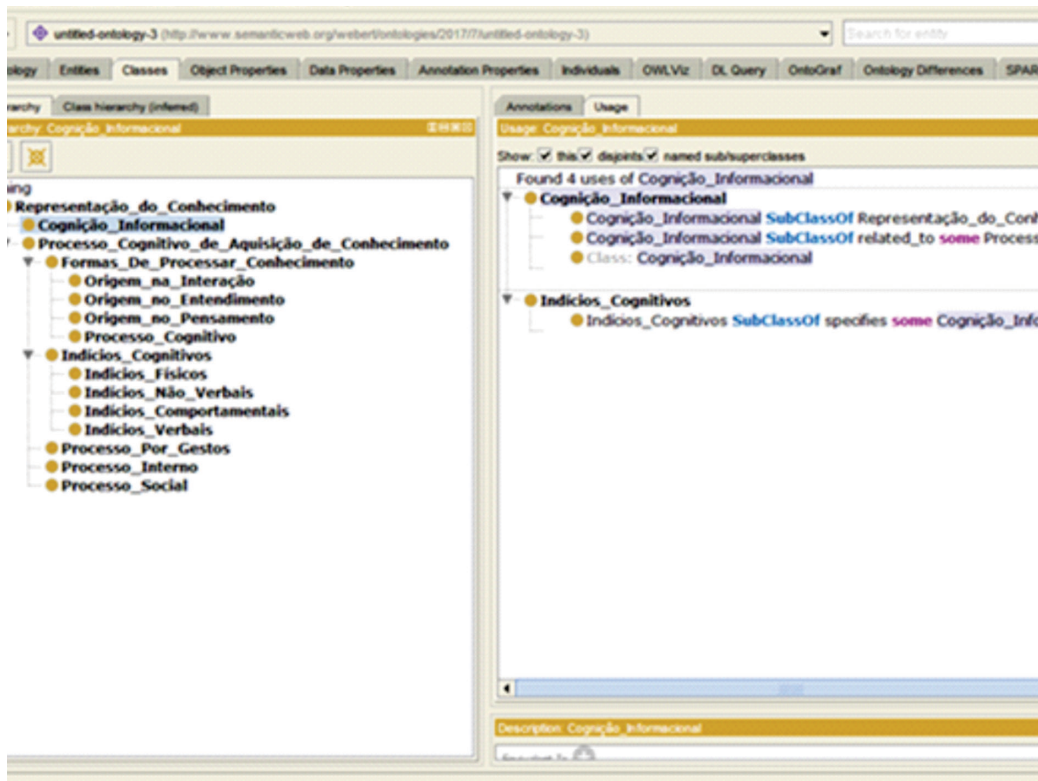
Após elaboração dos SOCs, as classes apresentaram-se nas seguintes estruturas:

Figura 1 – Estruturas da taxonomia, sistema de classificação e tesouro

<p>Taxonomia: Processo cognitivo de aquisição de conhecimento Processo social Processo interno Processo por gestos Indícios cognitivos Indícios verbais Indícios comportamentais Indícios não-verbais Indícios físicos Formas de processar conhecimento Origem no pensamento Origem no entendimento Origem na interação</p>	<p>Sistema de Classificação 1 ATIVIDADES 1.1 Ciências Cognitivas 1.1.1 Teorias cognitivas 1.1.2 Processos Cognitivos 1.1.2.1 Acordo cognitivo; Processo social 1.1.2.2 Enfoque cognitivo; Processo interno; Disposição para o aprendizado 1.1.2.3 Gesto categorial; Linguagem corporal; Expressividade não-verbal; Movimentos de cabeça, olhos, mãos e posição do corpo; Expressões faciais 1.1.2.4 Indícios cognitivos 1.1.2.4.1 Indícios verbais; Aderência de conceitos; Assimilação; Aprendizagem de conceitos 1.1.2.4.2 Indícios comportamentais; Critério para atribuição; Mudança cognitiva 1.1.2.4.3 Estrutura cognitiva; Padrões de ações física e mental; Estrutura do conhecimento; Conflitos; Reestruturação; Estágios do desenvolvimento; Conhecimento novo; Assimilação e acomodação 1.1.2.4.4 Limites entre categorias; Conhecimento prévio 1.1.2.4.5 Indícios físicos; Estereótipos 1.1.2.4.5.1 Função socio-cognitiva 1.1.2.4.5.2 Função socioafetiva 1.1.2.5 Componentes cognitivos 1.1.2.5.1 Origem no pensamento: Memorização 1.1.2.5.2 Origem no entendimento: Percepção de relações; Aprendizagem 1.1.2.5.3 Origem na interação: Abordagem de Vigotsky 1.1.2.5.3.1 Afetiva 1.1.2.5.3.2 Moral; Juízos 1.1.2.5.4 Comportamento reprodutivo</p>	<p>Tesouro COGNIÇÃO INFORMACIONAL NE É um tipo de representação do conhecimento que envolve elementos como a percepção, o acordo cognitivo, o gesto categorial, o pensamento e os indícios cognitivos. A noção de conhecimento é primária e advém das Ciências Cognitivas, de diferentes teorias cognitivas, se estruturando no paradigma cognitivo, sobretudo com Piaget e Vigotsky. TG Representação do conhecimento TE Gesto categorial Indícios cognitivos Pensamento Percepção TR Ciências cognitivas Paradigma cognitivo Piaget Teoria cognitiva Vigotsky</p>
---	--	---

Fonte: Elaborado pelas autoras (2017).

Figura 2 – Estrutura da ontologia



Fonte: Elaborado pelas autoras (2017).

A taxonomia foi elaborada pensando-se mais em um *menu* de busca em um *site* sobre o processo cognitivo de conhecimento. Já tesouro, como a intenção é indexar documentos, a modelagem está mais voltada para a cognição informacional, pois o documento é da área da CI. O sistema de classificação foi modelado de modo mais geral (enciclopédico) com o intuito de classificar assuntos em âmbito mais amplo, para organização de recursos informacionais (acervos). A ontologia foi modelada a partir dos interesses dos usuários e o que eles esperam encontrar dentro da lógica da base do conhecimento deste domínio, no contexto da Web Semântica.

A estrutura de uma taxonomia é bastante similar à estrutura de um sistema de classificação, a não ser pelo fato de que não é da natureza de uma taxonomia possuir notações. Tal como ocorre nos sistemas de classificação, também nas taxonomias acontece a repetição de termos. Mas, ainda que os termos sejam reproduzidos, eles têm uma dimensão

conceitual diferente. Ressalta-se que, com isso, uma característica tanto das taxonomias quanto dos sistemas de classificação é que o rótulo ou cabeçalho, fora da estrutura, não pode identificar o significado de forma unívoca.

Nas taxonomias e nos sistemas de classificação, não há o controle de equivalências, como existe nos tesouros. Contudo, com certeza, os rótulos e cabeçalhos poderiam ser substituídos por algum termo equivalente (sinônimos, quase-sinônimos, variações linguísticas), o que pode ocasionar alguma frustração no usuário, caso o termo escolhido não seja intuitivo para ele. Uma das vantagens do tesouro em relação à taxonomia e aos sistemas de classificação, é a possibilidade de expressar o conjunto completo de relações associativas entre conceitos e não apenas as relações hierárquicas que compõem esses instrumentos.

Ainda que bastante completas, as relações semânticas explicitadas nos tesouros não são capazes de alcançar o nível descritivo e de formalidade

encontrado em ontologias consistentes, uma vez que nestas as relações estabelecidas conduzem à possibilidade de inferência pelo sistema. Nas ontologias, os conceitos concretos, que representam as propriedades de domínio e seus relacionamentos, são distinguidos dos conceitos abstratos, que são representados como características e atributos. No tesouro, na taxonomia e no sistema de classificação, esse tipo de especificação não é requerido e não pode ser determinado. Assim, observa-se que a ontologia possibilita uma representação formal, com base em lógica, o que permite a este artefato fazer inferência sobre o domínio modelado, criando novos conhecimentos.

Os sistemas de organização do conhecimento possuem características específicas para sua aplicação em determinados contextos, e alguns exigem normas e metodologias diferenciadas de construção e atualização de acordo com suas finalidades e aplicações específicas.

CONSIDERAÇÕES FINAIS

Este artigo apresenta resultados do trabalho prático realizado com os alunos da disciplina Representação da Informação. Esse trabalho partiu do pressuposto de que, por meio da prática, em conjunto com a fundamentação teórica, e à medida que os alunos fossem realizando as etapas da modelagem de cada instrumento, suas características ficariam mais evidentes, e seus objetivos e propósitos, mais claros. Conforme relato dos alunos da disciplina, à medida que eles foram realizando as etapas da modelagem, agrupando os termos, a partir de suas relações semânticas, eles puderam distinguir com maior clareza as características, os objetivos e os propósitos de cada um dos instrumentos que estavam sendo construídos.

Ressalta-se que a integração entre os alunos de diferentes áreas do conhecimento, que geralmente matriculam-se nesta disciplina, com formação em biblioteconomia, em ciências da computação, engenharia, sistemas de informação, letras e ciências cognitivas, entre outras, auxilia na elaboração

dos instrumentos, pois se cria uma interação de conhecimentos que facilita a realização de todas as etapas dos procedimentos metodológicos na criação dos instrumentos, desde a modelagem até a implementação da ontologia.

A proposta de utilizar a prática como forma de complementação, ou em conjunto com os fundamentos conceituais encontrados na literatura, e com as recomendações das normas de construção de tesouros e ontologias, a ISO 25964-1 (2011) e a 25964-2 (2013), foram eficientes, pois se considera que o objetivo proposto foi atingido, já que facilitou o entendimento dos alunos quanto às semelhanças e diferenças existentes nestes sistemas de organização do conhecimento, conforme proposta do conteúdo programático da disciplina Representação da Informação.

REFERÊNCIAS

- 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. 2013. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.22861/full>>. Acesso em: 11 ago. 2017.
- ARISTÓTELES. *Metafísica*: livro alfa. São Paulo: Abril Cultural, 1978. (Coleção Os Pensadores).
- CAMPOS, M.L.A.; GOMES, H.E. Taxonomia e classificação: a categorização como princípio. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2007, Salvador, Bahia. *Anais...* Salvador: UFBA. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT2--101.pdf>>. Acesso em: 11 ago. 2017.
- GILCHRIST, A. Thesauri, taxonomies and ontologies: an etymological note. *Journal of Documentation*, v. 59, n. 1, p. 7-18, 2003. DOI: 10.1108/00220410310457984.
- GREEN, R. Relationship in Knowledge Organization. *Knowledge Organization*, v.35, n.2/3, p.150-159, 2008. Disponível em: <https://cluster10-files.instructure.com/files/25425605/download?download_fid=1&verifier=T2jUGuHtAb6PBobt8Vajx9w4n3Fa1FBwKM2r0vd9>. Acesso em: 11 ago. 2017.
- GRUBER, T. What is an Ontology?, 1992. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 11 ago. 2017.

- GRUNINGER, M. Ontology, taxonomy, folksonomy: understanding the distinctions. *Applied Ontology*, n. 3 p. 191–200. 2008. Disponível em: <<https://mor.nlm.nih.gov/pubs/pdf/2008-ao-mg.pdf>>. Acesso em: 11 ago. 2017.
- GUINCHAT, C.; MENO, M. *Introdução geral às ciências e técnicas da informação e documentação*. Brasília, DF: IBICT, 1994.
- INTERNATIONAL STANDARD ORGANIZATION. *ISO 25964-1: thesauri and interoperability with other vocabularies*. Part 1: thesauri for information retrieval. Geneve: International Standard Organization, 2011.
- INTERNATIONAL STANDARD ORGANIZATION. *ISO 25964-2: thesauri and interoperability with other vocabularies*. Part 2: interoperability with other vocabularies. Geneve: International Standard Organization, 2013.
- ISOTANI, S.; BITTENCORTT, I.I. *Dados abertos conectados*. São Paulo, SP: Novatec, 2015. Disponível em: <<http://ceweb.br/livros/dados-abertos-conectados/>>. Acesso em: 11 ago. 2017.
- KLESS, D. et al. Thesaurus and Ontology Structure: Formal and Pragmatic Differences and Similarities. *Journal of The Association for Information Science and Technology*, v. 66. n. 7, p. 1348–1366, 2015. Disponível em : <<http://onlinelibrary.wiley.com/doi/10.1002/asi.23268/abstract>>. Acesso em: 11 ago. 2017.
- KLESS, D.; KAZMIERCZAK, E.; MILTON, S. K. Relationships and Relata in Ontologies and Thesauri: Differences and Similarities. *Applied Ontology*, v. 7, n. 4, nov., 2012, p. 401–428. Disponível em: <<http://content.iospress.com/articles/applied-ontology/ao118>>. Acesso em: 11 ago. 2017.
- LAMBE, P. *Organising knowledge: taxonomies, knowledge and organizational effectiveness*. Oxford, England: Chandos, 2007.
- MACULAN, B.C.M.S. et al. Taxonomia, folksonomia, acessibilidade e usabilidade: proposta de interseção na área de organização do conhecimento, com foco na recuperação de informação. In: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO - SECIN, 3., Londrina. *Anais...* Londrina: UEL, 2009. p. 1-26. Disponível em: <<http://eprints.rclis.org/23854/>>. Acesso em: 11 ago. 2017.
- MACULAN, B.C.M.S.; LIMA, G.A.B.O.; PENIDO, P. Taxonomia facetada como interface para facilitar o acesso à informação em bibliotecas digitais. *Revista ACB: Biblioteconomia em Santa Catarina, Florianópolis*, v. 16, n. 1, p. 234-249, jan./jun. 2011. Disponível em: <<https://revista.acbsc.org.br/racb/article/view/746>>. Acesso em: 11 ago. 2017.
- MARTÍNEZ, A. et al. Las categorías o facetas fundamentales: una metodología para el diseño de taxonomías corporativas de sitios web argentinos. *Ciência da Informação*, Brasília, DF, v. 33, n. 2, p. 106-111, 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n2/a11v33n2.pdf>>. Acesso em: 11 ago. 2017.
- MOREIRO GONZÁLEZ, J.A. *Linguagens documentárias e vocabulários semânticos para a web: elementos conceituais*. Salvador: EDUFA, 2011.
- NEVES, D.E; BRANDÃO, W.C.; ISHITANI, L. Metodologia para recomendação e agregação de Objetos de Aprendizagem no padrão SCORM. RBIE, v.24, n.1, 2016. Disponível em: <<http://www.wladimirbrandao.com/docs/publications/2016@RBIE.pdf>>. Acesso em: 11 ago. 2017.
- RANGANATHAN, S.R. *Phelogema library classification*. Bombay, Ásia: Publishing House, 1967, 540 p.
- SMITH, B. Ontology: Philosophical and computational. In: FLORIDI, L. (Ed.). *The Blackwell Guide to the Philosophy of Computing and Information*. New Jersey: Blackwell Publishing.
- SOERGEL, D. *Functions of a thesaurus: classification: ontological knowledge base*. [S.l.]: College of Library and Information Services, University of Maryland, 1997. Disponível em: <<http://www.umiacs.umd.edu/~oard/teaching/878/spring99/readings/functclass.pdf>>. Acesso em: 11 ago. 2017.
- TRISTÃO, A.M.D.; FACHIN, G.R.B.; ALARCON, O.E. Sistema de classificação facetada e tesouros: instrumentos para organização do conhecimento. *Ciência da Informação*, Brasília, DF, v. 33, n. 2, p. 161-171, ago. 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n2/a17v33n2.pdf>>. Acesso em: 11 ago. 2017.
- VICKERY, B.C. *Classification and indexing in science*. 3. ed. Londres: Butterworths, 1975.
- _____. On knowledge organisation. In: FACETS OF KNOWLEDGE ORGANIZATIONS, 2., 2011. *Proceeding of the ISKO UK Second Biennial Conference*, Londres: Emerald, 2012. 416 p.
- _____. Ontologies. *Journal of Information Science*, v. 23, n. 4, p. 277-286, 1997. DOI: 10.1177/016555159702300402.
- ZENG, M.L. Knowledge organization systems (KOS). *Knowledge Organization*, Frankfurt, v. 35, n. 2-3, p. 160-182, 2008.

Guidelines for writing definitions in ontologies

Selja Seppälä

Ph.D., University of Geneva, Switzerland
Postdoctoral Associate, University of Florida, Gainesville, United States.
<http://seljaseppala.wordpress.com>
E-mail: sseppala@ufl.edu

Alan Ruttenberg

M.S. Massachusetts Institute of Technology
Director of Clinical Data Exchange, School of Dental Medicine,
University at Buffalo, United States.
<http://alan.ruttenbergs.com>
E-mail: alanruttenberg@gmail.com

Barry Smith

Ph.D., University of Manchester, United Kingdom.
Professor of Philosophy, University at Buffalo, Buffalo, United States.
<http://ontology.buffalo.edu/smith/>
E-mail: phismith@buffalo.edu

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

ABSTRACT

Ontologies are being used increasingly to promote the reusability of scientific information by allowing heterogeneous data to be integrated under a common, normalized representation. Definitions play a central role in the use of ontologies both by humans and by computers. Textual definitions allow ontologists and data curators to understand the intended meaning of ontology terms and to use these terms in a consistent fashion across contexts. Logical definitions allow machines to check the integrity of ontologies and reason over data annotated with ontology terms to make inferences that promote knowledge discovery. Therefore, it is important not only to include in ontologies multiple types of definitions in both formal and in natural languages, but also to ensure that these definitions meet good quality standards so they are useful. While tools such as Protégé can assist in creating well-formed logical definitions, producing good definitions in a natural language is still to a large extent a matter of human ingenuity supported at best by just a small number of general principles. For lack of more precise guidelines, definition authors are often left to their own personal devices. This paper aims to fill this gap by providing the ontology community with a set of principles and conventions to assist in definition writing, editing, and validation, by drawing on existing definition writing principles and guidelines in lexicography, terminology, and logic.

Keywords: Definitions. Ontology. Textual definitions. Definitions in ontologies. Guidelines. Design patterns. Applied ontology.

Diretrizes para criação de definições em ontologias

RESUMO

Ontologias têm sido usadas cada vez mais para promover a reutilização de dados científicos, permitindo que dados heterogêneos sejam integrados via uma representação normalizada e única. As definições desempenham um papel central no uso de ontologias, seja por humanos seja por computadores. Definições textuais permitem que ontologistas e curadores de dados entendam o significado pretendido dos termos da ontologia, e usem esses termos de forma consistente com o contexto. Definições lógicas permitem que máquinas verifiquem a integridade da ontologia e raciocinem sobre os dados anotados em termos ontológicos, para que as inferências promovam descoberta de conhecimento. Portanto, é importante não apenas incluir em ontologias diversos tipos de definições, tanto em linguagens formais como em linguagens naturais, mas também garantir que essas definições atinjam padrões de boa qualidade e sejam úteis. Embora ferramentas como o Protégé possam auxiliar na criação de definições lógicas bem estruturadas, produzir boas definições em linguagem natural ainda é, em grande medida, uma questão de engenheiros humanos, a qual é apoiada, na melhor das hipóteses, apenas por poucos princípios gerais. Por falta de diretrizes precisas, os autores de definições são muitas vezes deixados à sorte com seus próprios princípios pessoais. Este artigo pretende preencher essa lacuna, fornecendo à comunidade ontológica um conjunto de princípios e convenções para auxiliar na escrita, edição e validação de definições, a partir de princípios e diretrizes existentes em lexicografia, terminologia e lógica.

Palavras-chave: Definições. Ontologia. Definições textuais. Definições em ontologias.

Directrices para creación de definiciones en ontologías

RESUMEN

Ontologías han sido usadas cada vez más para promover la reutilización de datos científicos, permitiendo que datos heterogêneos sean integrados vía una representación normalizada es única. Las definiciones desempeña un papel central en el uso de ontologías, sea por humanos sea por computadoras. Definiciones textuales permiten que ontologistas y curadores de datos entiendan el significado pretendido de los términos de la ontología, y usen esos términos de forma consistente con el contexto. Definiciones lógicas permiten que máquinas verifiquen la integridad de la ontología y raciocinen sobre los datos anotados en términos ontológicos, para que las inferencias promuevan descubierta de conocimiento. Por lo tanto, es importante no solamente incluir en ontologías diversos tipos de definiciones, tanto en lenguajes formales como en lenguajes naturales, pero también garantizar que esas definiciones atinjan patrones de buena calidad y sean útiles. A pesar de que herramientas como el Protégé puedan auxiliar a la creación de definiciones lógicas bien estructuradas, producir buenas definiciones en lenguaje natural aún es, en gran medida, una cuestión de ingenieros humanos, la cual es apoyada, en la mejor de las hipótesis, solamente por pocos principios generales. Por falta de directrices precisas, los autores de definiciones son muchas veces dejados a la suerte con sus propios principios personales. Este artículo pretende llenar esta laguna, suministrando a la comunidad ontológica un conjunto de principios y convenciones para auxiliar en la escrita, edición y validación de definiciones, a partir de principios y directrices existentes en lexicografía, terminología y lógica.

Palabras-clave: Definiciones. Ontología. Definiciones textuales. Definiciones en ontologías.

INTRODUCTION

Ontologies are being used increasingly to promote reusability of scientific and other sorts of information and to address semantic interoperability issues by integrating heterogeneous data under a common, normalized representation. Definitions play a central role in the use of ontologies both by humans and by computers. Textual definitions allow ontologists and data curators to understand the intended meaning of ontology terms and to use these terms in a consistent fashion across multiple heterogeneous contexts. Definitions formalized in a logical language such as the Web Ontology Language (OWL) allow machines to check the integrity of ontologies and reason over data annotated with ontology terms to make inferences that promote knowledge discovery in areas such as biomedicine.

Therefore, it is important not only to include in ontologies definitions of both types, but also to ensure that these definitions can be created in a way that will contribute in a reliable fashion to their being useful.

According to a recent survey we conducted on definition practices in ontologies (SEPPÄLÄ, 2013; SEPPÄLÄ; RUTTENBERG, 2013), ontologists often lack adequate training in definition writing. This may partly explain the disparities in definition coverage observed in the OBO Foundry ontologies. It may also help to explain why so many ontologies lack either or both forms of definition (SCHLEGEL; SEPPÄLÄ; ELKIN, 2016). While tools such as Protégé can assist in creating well-formed logical definitions, producing good natural language definitions is still more a matter of human ingenuity supported at best by just a small number of general definition writing principles.

Our aim here is to fill this gap by providing the ontology community with a set of principles and conventions to assist in definition writing, editing, and validation. They draw from existing proposals from the disciplines of applied ontology, terminology, lexicography, and logic (ARP; SMITH; SPEAR, 2015; 2009; KELLEY, 1998;

LANDAU, 2001; NDI-KIMBI, 1994; PAVEL; NOLET, 2001; SMITH, 2013; SVENSÉN, 1993; SWARTZ, 1997; VÉZINA et al., 2009).¹

SOME PRELIMINARIES ON DEFINITIONS IN GENERAL

To make good use of the guidelines which follow, it is important to specify in which way the term *definition* is used. It is also important to know about the functions of definitions in ontologies. We briefly describe these aspects to provide background for the definition writing guidelines which follow.²

ONTOLOGICAL DEFINITIONS

The purposes of ontological definitions are different from those of definitions in standard dictionaries. In particular, the latter are required to define *all* lexical items, including those – for example ‘of’, ‘because’, ‘the’ – which do not refer but rather serve structural roles in natural language sentences. Ontological definitions, in contrast, are for ‘terms’, singular noun phrases which are content words that form part of a domain-specific vocabulary used by a group of experts to communicate about entities to which the terms refer.

HOW IS THE WORD ‘DEFINITION’ USED?

The word definition can be used in a number of ways. Here we focus on two: (i) *intension* and (ii) representational artifact. The former is the content or meaning of the definition and has a counterpart on the side of reality – the set of things, at any given time, to which the definition applies – which is called the *extension* of a definition. The latter is the form of the definition, the natural language text and potentially the accompanying axioms in some logical formalism that express this meaning. Good definition writing hinges on recognizing that these are the two sides of a single coin.

¹ The present guidelines follow the structure and to some extent the wording of the lexicographic definition writing guidelines prepared for the wordnet community (SEPPÄLÄ, Forthcoming).

² The preliminaries section is largely based on (SEPPÄLÄ et al., 2016; SEPPÄLÄ; RUTTENBERG; SMITH, 2016; SEPPÄLÄ; SCHREIBER; RUTTENBERG, 2014; SMITH, 2013), which discuss these aspects in more detail.

STRUCTURE OF A TEXTUAL DEFINITION

When used to refer to the natural language text of a definition, the term ‘definition’ itself can denote different forms: a sentence and a sentence fragment. Broadly, a definition has the canonical form X is a Y that Zs as in example (1) adapted from the definition of ‘ligament’ (synonym of ‘skeletal ligament’) in the Uberon multi-species anatomy ontology (UBERON).

(1) A **ligament** is a dense regular connective tissue connecting two or more adjacent skeletal elements.

Definitions in this form have a three-part structure:

1. a **definiendum** [X], i.e., the defined term;
2. a **definiens** [a Y that Zs], i.e., the part that expresses the definition content and that is called a **definition** in dictionaries;
3. a **copula** [is] that expresses an equivalence between definiendum and definiens.

In example (1), ‘ligament’ is the definiendum, which is connected to the definiens “a dense regular connective tissue connecting two or more adjacent skeletal elements” with the copula “is”. Thus, in the wider meaning, ‘definition’ signifies a whole consisting of definiendum, copula, and definiens and takes the linguistic form of a sentence, as in (1).

More narrowly, it denotes the **definiens** alone and takes the linguistic form of a sentence fragment, as in (2):

(2) Dense regular connective tissue connecting two or more adjacent skeletal elements. (UBERON_0008846)

The guidelines presented in this paper are concerned with the formulation of the definiens. In dictionaries, definiendum and definiens appear in distinct entry fields and the copula is usually implicit, as in (3):

(3) ligament: *Dense regular connective tissue connecting two or more adjacent skeletal elements.* (UBERON_0008846)

The **definiens** – in italics – is subdivided into at least two parts:

- The **genus** – *connective tissue* – which constitutes the head of the definition (the Y part) and tells us what kind of thing the defined term denotes. When the genus is the immediate superordinate term it is called the **genus proximus**.
- One or more **differentia(e)** (the Z part(s) or **distinguishing features**) – *dense, regular and connecting two or more adjacent skeletal elements* – that tell us what characterizes the things referred to by the definiendum and what distinguishes them from the things referred to by the genus term and by all other terms under the same genus.

Example (3) thus tells us that “connective tissue” is the genus of the definition of ‘ligament’ and that a ligament is connective tissue of a certain type. The definition has three differentiae, “dense”, “regular” and “connecting two or more adjacent skeletal elements”. These together tell us that ‘ligament’ is more specific than ‘connective tissue’ and that ‘ligament’ differs from other subtypes of connective tissue in virtue of the fact that it is dense and regular and that it has the function of connecting two or more adjacent skeletal elements.

Each part serves as a condition for determining which things are members of the defined term’s extension. In ontologies, definitions include **necessary conditions** that apply to all the members of the extension, but which may also apply to members of other term extensions, and, whenever possible, jointly **sufficient conditions**, which allow a user to determine whether a given entity is a member of the extension.

(4) triangle: *A polygon with three edges and three vertices.* (adapted from WIKIPEDIA, 2017)

This definition tells us that every triangle is necessarily a polygon that has three edges and three vertices, and is sufficient for us to know that anything that is a polygon with three edges and three vertices is a triangle. Together, these

conditions form a *definition by necessary and jointly sufficient conditions*, also called a *classical definition*.

A definition that contains only necessary conditions that are not jointly sufficient to rule out instances that are not members of the extension is called a *partial definition*. For example, defining a bird as an animal that lays eggs produces a definition that applies to all instance of birds but also to amphibians.

FUNCTIONS OF DEFINITIONS

Irrespective of the type of resource in which they appear and their context of use, definitions have two primary functions:

- a *cognitive* function: to augment or reconfigure (constrain or correct) our knowledge of the world relating to a given term, thus allowing us to better understand a term's meaning in a specific context of use and create new meaningful semantic connections and interpretations;
- a *linguistic* function: to describe or prescribe what we should understand when one or more speakers use the defined term in a specific context of use.

For example, at the cognitive level, the definition of the term 'ligament' in (3) tells us *what a ligament is* (some physical object), *what it does* (it has a connective function), and *what its properties are* (it has the physical qualities of being dense and regular). At the linguistic level, this definition tells us how the UBERON developers intended us to understand and use the term.

To fulfill the mentioned functions, the content and form of definitions have to be adapted to each context of use (for example teaching a child to speak, teaching undergraduates in a classroom, or annotating data with an ontology of microbiology) and to the target audience, which in ontologies is for example human curators of databases or developers of natural language processing software. This explains why ontologies have definition types different from those of dictionaries and

other resources. In ontologies, specifying the intended meaning of a term involves clarifying and disambiguating this meaning so that it can be interpreted by a computer. Definitions in ontologies *stipulate* and *disambiguate* the meanings of terms in consistent and non-circular ways.

For ontologies, the primary context of use requires the inclusion of logical definitions since these are the sorts of definitions needed by a reasoner to perform logical operations, for example, with data annotated with ontology classes. Textual definitions are needed to facilitate ontology development and application by human users. In order to ensure consistency in ontology development and use, the textual and logical definitions of a term must convey the same type of content. This need for parallel semantic content along with the constrained logical properties conferred by the use of necessary conditions give logical definitions additional derived functions that are useful in definition writing and checking:

- *consistency checking and inference*: necessary conditions allows us to check whether an instance is consistent with the classes of which it is asserted to be a member, and to infer that the instance has all the properties that are necessary conditions, even if not explicitly stated.
- *instance classification*: sufficient conditions allow inferences to be made to the effect that given entities are instances of the class that is being defined;
- *taxonomic schematization*: the axioms of a class's logical definition can be used as a template for the axioms of its subclasses, as well as for the contents of the associated textual definitions;
- *regularizing expression and interpretation of facts*: the controlled vocabulary used in axioms allows us to check the intended meaning of a natural language expression in the textual definition.
- *avoiding circularity*: for any given ontology, there should be certain terms accepted as primitive (that are not defined in the ontology); other definitions

are then built in a step-by-step manner from these identified primitives (which may be defined in some superordinate ontology).

It is important to note that the differences between ontology and dictionary definitions result in varied sets of definition correctness criteria.

DEFINITION WRITING GUIDELINES

The guidelines we present in this paper are relevant for writing good definitions in ontologies and emphasize textual definitions. The principles are numbered for easy reference and, whenever possible, illustrated with definitions from existing ontologies.³ The examples include a term (sometimes called a class label), the definition, and the source of the definition – which may be identified through a URI⁴. Our proposed solutions are based on the original definitions, unless otherwise specified.

1. Guideline: CONFORM TO CONVENTIONS

Authors should conform to the usual linguistic and lexicographic conventions of the English language (where relevant extended by discipline specific idiolects).

- ‘Definition’ in what follows means: ‘definiens’. Thus, the definition should not include the term being defined (the definiendum) and it should not include the copula *is*; it should be limited to the definiens, a logically connected sentence fragment – see the definition of ‘data set’ in example (5).
- A definition should avoid punctuation marks other than commas, but should end with a period if the first letter is capitalized. Uses of other sorts of punctuation – e.g., parentheses,

colons, slashes (as in ‘and/or’), or semi-colons – are to be avoided, as they detract from the requirement that the definition should be a single logically unitary and unambiguous sentence fragment.

- Definitions should be written in a natural language supplemented by the technical terms and symbols used in specific disciplines where necessary.
- Nouns are divided into ‘count’ and ‘mass’ according to whether they can be pluralized (‘cow’ and ‘datum’ are examples of the former; ‘cattle’ and ‘information’ of the latter). Definitions of count nouns should start with an article (‘a’, ‘an’, ‘the’).

1.1. Guideline: HARMONIZE DEFINITIONS

Harmonize the definitions in the ontology in order that they all conform to a single set of typographical conventions and editorial guidelines. Example (5) illustrates a case where the needed consistency is lacking. The definitions of ‘data set’ and ‘measurement datum’ in the Information Artifact Ontology (IAO) should either all be limited to the *definiens* (recommended) or all contain the *defined term and the ‘is’ copula*.

- | | | |
|-----|---|---|
| (5) | ✗ | data item: <i>A data item is an information content entity that...</i> (IAO_0000027) |
| | ✓ | data set: <u>A data item that is an aggregate of other data items of the same type that...</u> (IAO_0000100) |
| | ✗ | measurement datum: <i>A measurement datum is an information content entity that...</i> (IAO_0000109) |

Where it is agreed by the ontology editors that a specified limited vocabulary will be used in definitions, the wording of the definitions should also be checked to ensure conformance.

³ Note that the sources may have been edited since we accessed them on BioPortal (<http://biportal.bioontology.org>, accessed July 14, 2017) and OntoBee (<http://www.ontobee.org>, accessed July 14, 2017).

⁴ The ontologies referenced in the examples are listed at the end of the guidelines. All the CURIEs are prefixed with <http://purl.obolibrary.org/obo/> unless otherwise specified in the ontology reference list.

2. Guideline: PRINCIPLES OF GOOD PRACTICE

The following are principles for working with definitions that should be followed to promote good practice.

- Very high-level ontologies will have many terms — such as ‘entity’ — which are so general that they are not capable of being defined without circularity; these should be marked as ‘primitive’; they should be accompanied by elucidations and examples of use.
- Definitions should be unique (i.e., no two terms in a single ontology should share what is, logically speaking, the same definition).⁵
- When referring to other classes, use the controlled vocabulary specified in the ontology (i.e., the terms in the ontology labels) to produce consistently written definitions across the ontology.
- Include a single definition per class. This means that there should be a single ‘definition’ annotation property (IAO_0000100), which should contain a single definition (see also principle 6.2).
- Cite your sources and do it in a separate annotation property, preferably using the IAO ‘definition source’ annotation property (IAO_0000119).
- Credit authorship, preferably using the IAO ‘term editor’ annotation property (IAO_0000117).

3. Guideline: USE THE GENUS-DIFFERENTIA FORM

A definition should have the genus-differentia form, where the genus anchors the defined entity to a known entity at a higher level of generality, and the differentia (or differentiae) picks out those cases within this higher level

that fall under the term defined. In example (6), only the first definition of mammal is correctly structured. It has a *genus* and three *differentiae*; the second definition only has a genus followed by example subtypes; the last definition lacks a genus (including the definiendum instead) and has one differentia.

- (6) ✓ mammal: **a vertebrate that has hair, gives live birth, and nurses its young**
 ✗ mammal: **a vertebrate like, a cat, dog, or whale**
 ✗ mammal: *Mammals nurse their young.*

3.1. Guideline: INCLUDE EXACTLY ONE GENUS

A definition should always have one and only one genus (7). The genus is a superordinate term that tells us with what kind of thing we are dealing. Definitions of things like colors are expressed in the nominal form and start with a genus such as “a quality”. Similarly, definitions of processes and other entities that unfold in time, which are often expressed by verbs are, in ontologies, defined with nominal phrases.

In example (7), the definition of ‘recombinant vector’ in the Ontology for Biomedical Investigations (OBI) does not have a genus. This can be fixed by adding its *genus proximus*, which in OBI is ‘processed material’.

- (7) ✗ recombinant vector: *A recombinant vector is created by a recombinant vector cloning process, ... (OBI_0000731)*
 ✓ recombinant vector: **A processed material** created by a recombinant vector cloning process, ...

3.1.1. Guideline: USE THE GENUS PROXIMUS

Use the *genus proximus*, that is, the closest parent term. This ensures that the genus is specific enough and that all the terms at the same level have the same genus.

⁵ *OBO Foundry Principles: Textual Definitions* at <http://obofoundry.org/principles/fp-006-textual-definitions.html> [accessed: July 14, 2017].

As an ontology is developed and new classes interposed between existing levels, the *genus proximus* may also change and the corresponding definitions need to be updated. Example (8) shows two definitions of sibling classes: the definition of ‘data set’ includes the correct *genus proximus* (‘data item’), whereas the definition of ‘measurement datum’ does not, as it includes the parent of its *genus proximus* (‘information content’).

- (8) ✗ **data item:** *A data item is an information content entity that...* (IAO_000027)

data set: *A data item that...* (IAO_0000100)

measurement datum: *A measurement datum is an information content entity that...* (IAO_0000109)

- ✓ **data item:** *An information content entity that...*

data set: *A data item that...*

measurement datum: *A data item that...*

3.1.2. Guideline: AVOID PLURALS

The genus of a definition should have the same syntactic properties as the definiendum. According to the naming conventions contained in the OBO Foundry Principles,⁶ which are also in line with good terminology practice accepted elsewhere, terms in ontologies should be written “as if writing in plain English”. This includes writing them in the singular form. Therefore, the genus of a definition should also be in the singular form. In example (9), the definition of ‘myeloablative agonist’ in the National Cancer Institute Thesaurus (NCIt) has a plural genus, which can be replaced by its singular form. Note that the rest of the definition needs to be adapted to accommodate this change.⁷

⁶ OBO Foundry Principles: Naming Conventions at <http://obofoundry.org/principles/fp-012-naming-conventions.html> [accessed: June 21, 2017].

⁷ In the corrected version of this example, we also edited the rest of the definition to comply with the single fragment principle.

- (9) ✗ Myeloablative Agonist: **Agents that destroy bone marrow activity. They are used to prepare patients for bone marrow or stem cell transplantation.** (NCIt:C1711)

- ✓ myeloablative agonist: **An agent that destroys bone marrow activity and that is used to prepare patients for bone marrow or stem cell transplantation.**

3.13. Guideline: AVOID CONJUNCTIONS AND DISJUNCTIONS

The genus of a definition should be a single word: not a conjunction (10) and also not a disjunction (11). A conjunction or disjunction in a genus can be avoided by using a more general genus (10) and, if applicable, by adding a differentia specifying the things denoted by the conjunction or disjunction, as in (11). In example (10), we avoid the conjunction by using a more general term and adding the elements of the conjunction as differentia – note that the new genus might have to be added to the ontology if it falls within its scope. In example (11), we avoid the disjunction by using a parent term.

- (10) ✗ Cell Culture System: **Systems and reagents for the propagation of cells in tissue culture.** (NCIt:C19147)

- ✓ cell culture system: **A system comprising reagents and other systems which together enable propagation of cells in tissue culture.**

- (11) ✗ analyte: **The sample or material being subjected to analysis.** (NCIt:C128639)

- ✓ analyte: **A spe/cimen being subjected to analysis.**

3.1.4. Guideline: AVOID CATEGORIZERS

The role of the genus of a definition is to categorize the denoted thing under a more general type. Therefore, it is not necessary to add an extra categorizing expression, such as “type of” and “sort of”. These can usually be omitted, as in (12).

- (12) ✘ immunogen: **A type of antigen capable on its own of inducing an immune response.** (CHEBI_60816)
- ✓ immunogen: **An antigen capable, on its own, of inducing an immune response.**

4. Guideline: AVOID USE/MENTION CONFUSION

A definition should tell us about the thing that is denoted by the definiendum, not about the term itself or the class that represents the defined thing. Making this confusion is called the *use/mention confusion*. It is illustrated in examples (13) and (14).

- (13) ✘ miosis AE: *Miosis AE is an eye AE that is a term with various definitions, which generally include constriction of the pupil.* (OAE_0001998)
- ✓ miosis AE: *An eye AE that generally includes constriction of the pupil.*
- (14) ✘ mammal: **Representation of an animal that feeds its young with milk.** (ARP; SMITH; SPEAR, 2015)
- ✘ mammal: **A term that refers to an animal that feeds its young with milk.**
- ✓ mammal: **An animal that feeds its young with milk.**

The first two examples in (14) confound the class and the term with the thing that they represent, respectively.

5. Guideline: INCLUDE NECESSARY AND, WHENEVER POSSIBLE, JOINTLY SUFFICIENT CONDITIONS

Definitions should avoid non-defining information. They must consist exclusively of necessary conditions which should, whenever possible, be jointly sufficient.

5.1. Guideline: AVOID ENCYCLOPEDIA INFORMATION

Definitions should avoid including information that goes beyond the minimum useful information for specifying the intended meaning of a term (15). They should also avoid

including information about the use of a term (16). In ontologies, any type of information that goes beyond what is or should be expressed by the axioms may be considered encyclopedic. This kind of information might be useful for understanding a term's intended meaning or its use or is deemed otherwise useful for the target audience but it is not defining information and so should not be included in a definition. Often, an indication that a piece of information is encyclopedic is that it cannot be expressed as a necessary condition. Such information should be included in a separate annotation property, such as 'editor note' (IAO_0000116) or 'example of usage' (IAO_0000112).

- (15) ✘ spiracle: *An orifice of the tracheal system by which gases enter and leave the body. **Ants have 9 or 10 spiracles on each side of the body. The prothoracic spiracles have been lost, so the first opening occurs on the mesothroax.*** (SIBO_0000502)
- ✓ spiracle: *An orifice of the tracheal system by which gases enter and leave the body. **Editor note:** Ants have 9 or 10 spiracles on each side of the body. The prothoracic spiracles have been lost, so the first opening occurs on the mesothroax.*
- (16) ✘ complex: *Set of interacting molecules that can be copurified. **This term and its children should be used only at PARTICIPANT level.*** (MI_0314)
- ✓ complex: *Set of interacting molecules that can be copurified. **Editor note:** This term and its children should be used only at PARTICIPANT level.*

5.2. Guideline: AVOID NEGATIVE TERMS

A definition that tells us what something *is not* is uninformative; it should tell us what something *is*. Therefore, a definition should not include negative features, which should be avoided by stating positive characteristics of the defined entity (17). However, there are at least two

exceptions to this principle: (i) when the defined entity is inherently negative (18) and (ii) when the negative feature is a distinguishing feature between two sibling classes (19).

- (17) ✗ mineral: *composed of matter other than plant or animal* (WordNet 3.1)
 ✓ mineral: *Composed of solid inorganic matter.*
- (18) ✓ non-parametric test: *A statistical hypothesis test that is not based on any parameterized family of probability distributions.* (OBCS_0000236)
 ✓ bachelor: *A man who is not married.*
- (19) ✓ paracentric inversion: *A chromosomal inversion that does not include the centromere.* (SO_1000047)
 pericentric inversion: *A chromosomal inversion that includes the centromere.* (SO_1000046)

5.3. Guideline: AVOID DEFINITIONS BY EXTENSION

Definitions by extension (20) list instances, i.e., members of the definition's extension, and do not tell us what the thing being defined is, unless we are already familiar with the listed instances. To be maximally informative, use definitions by intension that specify characteristics of the defined thing, as in the corrected version of (20).

- (20) ✗ planet: *Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune.*
 ✓ planet: *A celestial body in orbit around the Sun that has a nearly round shape and has cleared the neighborhood around its orbit⁸.*

⁸Adapted from the 2006 definition of planet by the International Astronomical Union (IAU).

6. Guideline: ADJUST THE SCOPE

Definitions should have the appropriate scope.

6.1. Guideline: A DEFINITION SHOULD BE NEITHER TOO BROAD NOR TOO NARROW

A definition that is too broad includes things that are not in the extension of the term being defined (21). A definition that is too narrow fails to include things that are in the extension of the term being defined (22).

- (21) ✗ bird: *an animal that lays eggs* (KELLEY, 1998)
 ✓ bird: *warm-blooded egg-laying vertebrate with feathers and with wings that evolved from forelimbs* (adapted from WordNet 3.1)

In example (21), the first definition of bird would be too broad since it includes all oviparous animals, i.e., fish, birds, reptiles and insects. By contrast, the above definition has an adequate scope.

- (22) ✗ antidote: *a substance that counteracts snakebite* (KELLEY, 1998)
 ✓ antidote: *a remedy that stops or controls the effects of a poison* (WordNet 3.1)

In example (22), the first definition of antidote would be too narrow since it excludes other types of poison from its extension. By contrast, the WordNet 3.1 definition of antidote has an adequate scope.

6.2. Guideline: DEFINE ONLY ONE THING WITH A SINGLE TEXTUAL DEFINITION

A definition should define only one thing. Therefore, a class should have a single textual definition. Two cases that should be avoided are often found: (i) nested definitions, where definitions contain definitions of other terms embedded within them (23), and (ii) multiple definitions, where a single definition annotation property contains definitions in addition to the primary definitional sentence fragment (24). Such cases might indicate that a new class should be added to the ontology. If the additional definition is

of an entity that falls within the scope of the ontology, we recommend either (i) removing the additional definition from the annotation property and adding it as the definition of a new term, then using that in the first definition, as in (24), (ii) if the class already exists in the ontology, use the class in the primary definition and delete the additional definition (23). If the additional definition is of an entity that is out of the scope of the ontology, we recommend either (i) if the term exists in another ontology, using MIREOT (COURTOT et al., 2011) to import the term and then using it, (ii) adding it into a separate editor note, (iii) deleting the additional definition, or (iv) replacing the term deemed to be unknown to the reader with its definition, as in the second solution in (23).

Whenever the term is already in the ontology, a cross-reference is preferable since it preserves the link to the full definition of the potentially problematic entity introduced in the definition, as well as to the term that denotes it, both of which are lost when the term is replaced by its definition. Replacing the term by its definition may cause issues if the corresponding class is deprecated and the definition must be updated. It is more difficult to find an embedded definition than it is to identify a term in a definition.

(23) The WordNet 3.1 definition of *cytolytic*, which contains a nested definition of *cytolysis*, can be edited as follows:

- ✗ cytolytic: *of or relating to **cytolysis**, the **dissolution or destruction of a cell*** (WordNet 3.1)
- ✓ cytolytic: *of or relating to **cytolysis***
- ✓ cytolytic: *of or relating to **the dissolution or destruction of a cell***

(24) ✗ cell measurement: *Any quantification of a morphological or physiological parameter of one or more cells. **A cell is a membrane-enclosed protoplasmic mass constituting the smallest structural unit of an organism that is capable of independent functioning.*** (CMO_0000227)

- ✓ cell measurement: *Any quantification of a morphological or physiological parameter of one or more **cells**.*

cell: A membrane-enclosed protoplasmic mass constituting the smallest structural unit of an organism that is capable of independent functioning.

7. Guideline: AVOID CIRCULARITY

For a definition to be informative, it should avoid circularity. Circularity can manifest itself in two ways: within the same definition or within the system of definitions.

Circularity within the same definition occurs when a class is defined in terms of itself using one of the labels attached to the class (25) or a synonym thereof (26), or some grammatically derived form, as in (27) when *fearful* is not separately defined independently of *fear*.

- (25) ✗ training objective: *A **training objective** which is fulfilled by the provision of some training* (OBI_0000962)
- ✓ ctraining objective: *An **objective** which is fulfilled by the provision of some training.*
- (26) ✗ large: *The attribute of something that is **big**.* (KELLEY, 1998)
- ✓ large: *The attribute something **has when it has a measurable quality that is above average.*** (adapted from the definition of *large* in WordNet 3.1)
- (27) ✗ fear: *The state of being **fearful**.* (LANDAU, 2001)
- ✓ fear: ***an emotion experienced in anticipation of some specific pain or danger*** (WordNet 3.1)

Circularity within the system of definitions occurs when terms are defined in terms of each other, forming a circular pair (28) or a circular chain of definitions.

- (28) ✗ **training objective**: *A training objective which is fulfilled by the provision of some **training*** (OBI_0000962)
- training process**: *A process that achieves a **training objective*** (OBI_0000962)

In example (28), the pair of circular definitions tell us that a training objective is fulfilled by some training (process) and that a training process achieves a training objective. It does not tell us what either of these things really are.

8. Guideline: INCLUDE JOINTLY SATISFIABLE FEATURES

The stated conditions must be jointly satisfiable (ARP; SMITH; SPEAR, 2015), that is

- the entity defined must have instances (29),
- there should be no logical contradictions involved (29). (ARP; SMITH; SPEAR, 2015)

(29) × round square: **A geometric figure that is simultaneously round and square shaped.**

9. Guideline: USE APPROPRIATE DEGREE OF GENERALITY

A definition should be general.

9.1. Guideline: AVOID GENERALIZING EXPRESSIONS

Avoid the use of expressions such as “usually” and “generally” since either

- a definition is already a statement of general knowledge about a typical case (e.g., in biology) (30); or
- the feature containing the expression is not a defining feature, but rather an encyclopedic piece of information that can be moved to a separate ‘editor note’ annotation property (31).

(30) × increased activity of parathyroid: *increased function of this paired endocrine gland that normally produces parathyroid hormone (PTH) that regulates calcium and phosphorous metabolism* (MP_0003432)

✓ increased activity of parathyroid: *Increased function of the paired endocrine gland that produces parathyroid hormone (PTH) that regulates calcium and phosphorous metabolism.*

(31) × Tumor Lysis Syndrome: **A syndrome resulting from cytotoxic therapy, occurring generally in aggressive, rapidly proliferating lymphoproliferative disorders. It is characterized by combinations of hyperuricemia, lactic acidosis, hyperkalemia, hyperphosphatemia and hypocalcemia.** (MESH:D015275)

✓ tumor lysis syndrome: *A syndrome resulting from cytotoxic therapy, characterized by combinations of hyperuricemia, lactic acidosis, hyperkalemia, hyperphosphatemia and hypocalcemia.*

Note: Tumor lysis syndrome occurs generally in aggressive, rapidly proliferating lymphoproliferative disorders.

9.2. Guideline: AVOID EXAMPLES AND LISTS

Avoid using examples and expressions that specify or enumerate examples (or counterexamples) of things within the definition, such as “etc.”, “for example”, and “such as”, whether it applies (i) to members of the definition’s extension (32) or (ii) to various types of things falling under the extension of a distinguishing feature, as in (33). In the first case, the example(s) should be included in a separate annotation property, such as ‘example’ (IAO_0000112). In the second case, when possible, replace the examples in the definition with a more general term, the extension of which includes all the examples, as in (33).

(32) × cellular_organism: *An organism of microscopic or submicroscopic size, especially a bacterium or protozoan* (NCRO_0000483)

- ✓ cellular organism: *An organism of microscopic or submicroscopic size.*
Examples: a bacterium; a protozoan

In example (32), bacterium and protozoan are examples of cellular organisms introduced by the expression “especially”.

- (33) ✗ patient questionnaire: *A questionnaire that comprises a set of **questions about a patient, such as height, weight, race, biological sex, clinical history, etc.**, which will be filled by the human subject.* (OBIB_0000020)
- ✓ patient questionnaire: *A questionnaire that comprises a set of **demographic and medical questions**, which will be filled by the human subject.*

In example (33), “etc.” marks the presence of examples listed to illustrate the types of questions that are part of a questionnaire (i.e., the extension of a differentia). These examples can be replaced by more general terms.

9.3. Guideline: AVOID INDEXICAL AND DEICTIC TERMS

Avoid indexical and deictic terms, such as ‘today’, ‘here’, and ‘this’ when they refer to (the context of) the author of the definition or the resource itself. Such expressions often indicate the presence of a non-defining feature or a case of use/mention confusion. Most of the times, the definition can be edited and rephrased in a more general way (34).

- (34) ✗ hypersecretion of corticotropin-releasing hormone: *excessive release of **this factor**, which normally stimulates the pituitary to release adrenocorticotrophic hormone, from the hypothalamus* (MP_0001753)
- ✓ hypersecretion of corticotropin-releasing hormone: *Excessive release of corticotropin-releasing hormone from the hypothalamus.* Editor note: corticotropin-releasing hormone normally stimulates the pituitary to release adrenocorticotrophic hormone

9.4. Guideline: AVOID SUBJECTIVE AND EVALUATIVE STATEMENTS

Definitions should avoid any kind of subjective and evaluative language, as in (35) where the words ‘delicious’ and ‘beautiful’ are removed from the definition without any change in the informative content.

- (35) ✗ cranberry bean: *Also called shell bean or shellout, and known as borlotti bean in Italy, the cranberry bean has a large, knobby beige pod splotched with red. The beans inside are cream-colored with red streaks and have a **delicious** nutlike flavor. Cranberry beans must be shelled before cooking. Heat diminishes their **beautiful** red color. They’re available fresh in the summer and dried throughout the year.* (FOODON_03411186)

- ✓ cranberry bean: *A bean that has a large, knobby beige pod splotched with red, that is cream-colored with red streaks, and has a nutlike flavor.*

Synonyms: shell bean; shellout; borlotti bean

Editor note: Cranberry beans must be shelled before cooking. Heat diminishes their beautiful red color. They’re available fresh in the summer and dried throughout the year.

10. Guideline: DEFINE ABBREVIATIONS AND ACRONYMS

Abbreviations and acronyms should be defined not explicated (36). In most cases, the full form of the term should be added as the preferred term using the ‘editor preferred term’ (IAO_0000111) annotation property, and the abbreviation or acronym added as a synonym label using ‘alternative term’ (IAO_0000118). Note that in (36) ‘laser’ can be kept as the preferred term since the acronym has become lexicalized.

- (36) ✘ laser: *A **laser** (acronym for light amplification by the stimulated emission of radiation) is a light source that...* (OBI_040064)
- ✓ laser: *A light source that...*
Alternative term: light amplification by the stimulated emission of radiation

Editor note: “laser” is an acronym

11. Guideline: Match Textual and Logical Definitions

Since textual and logical definitions are meant to specify the intended meaning of an ontology’s terms, they should *in principle* include the same type of information, although in some cases the logical definition might have additional axioms that help reasoning but which otherwise might be considered inessential. To ensure consistency in ontology development and use, and to promote cohesion across definitions, it is recommended that the parts of a given textual definition match the parts of the corresponding logical definition as in (37). However, the parts need not appear in the same order.

- (37) ✓ bacteremia: *An infection that has as part bacteria located in the blood.* (IDO_0000506)

bacteremia
EquivalentTo
infection and
(has_part some
(infectious agent and Bacteria and
(located_in some blood)))

11.1. Guideline: Proofreading Definitions

Always check the spelling and grammar using appropriate tools. In example (38), the definition of ‘anesthesiology residency program’, ‘specialty’ is misspelled and ‘administration’ is spelled incorrectly in two different ways (see words in bold). This could be easily rectified by running a spellchecker.

- (38) ✓ anesthesiology residency program:
*A medical residency in the medical **speciality** that focuses on the **administeration** of medication for the temporary general or local suppression of sensory or motor nerve function during some health care encounter or on making decisions regarding the **adminstration** of such medication.*
(OOST_00000232)

FINAL REMARKS

The objective of these guidelines was to provide the applied ontology community with the relevant principles and conventions for good definition practices in ontologies. These guidelines were also meant to give basic foundational background knowledge on definitions to help readers understand the rationale behind these principles. Finally, these definition writing guidelines have the added benefit of serving as a verification tool for existing ontology definitions.

ACKNOWLEDGEMENTS

This work was supported in part by the Swiss National Science Foundation (SNSF), by the NIH/NCATS Clinical and Translational Science Awards to the University of Florida UL1 TR000064, and by the University at Buffalo UL1 TR001412. The content is solely the responsibility of the authors and does not necessarily represent the official views of the SNSF, the National Institutes of Health, or the NCTE. Many thanks also to Amanda Hicks, Mark Jensen, Daniel R. Schlegel, and Patrick Ray for their useful discussions and comments and their contributions with examples.

REFERENCES

ARP, R.; SMITH, B.; SPEAR, A. D. *Building ontologies with basic formal ontology*. Cambridge, MA: MIT Press, 2015.

COURTOT, M. et al. Mireot: the minimum information to reference an external ontology term. *Applied Ontology*, v. 6, n. 1, p. 23-33, 2011.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION - ISO. *Terminology work: principles and methods (iso 704:2009)*. Geneva: ISO, 2009.

KELLEY, D. *The art of reasoning*. Third edition. New York, London: W.W. Norton & Company, 1998.

LANDAU, S. I. *Dictionaries: the art and craft of lexicography*. 2nd edition. Cambridge: Cambridge University Press, 2001.

NDI-KIMBI, A. Guidelines for terminological definitions: the adherence to and deviation from existing rules in bs/iso 2382: data processing and information technology vocabulary. *Terminology*, v. 1, n. 2, 1994. p. 327-350.

PAVEL, S.; NOLET, D. *Handbook of terminology*. Canada: Public Works and Government Services - Translations Bureau, 2001.

SCHLEGEL, D. R.; SEPPÄLÄ, S.; ELKIN, P. L. Definition coverage in the obo foundry ontologies: the big picture. In: *Proceedings of the International Workshop on Definitions in Ontologies (ICBO BioCreative 2016)*, August 1-4, Corvallis (ORG), USA, 2016.

SEPPÄLÄ, S. Survey on defining practices in ontologies: Report summary. In: *Proceedings of the International Workshop on Definitions in Ontologies (DO 2013)*, 2013, Montreal, Canada. 2013.

_____. Definition writing guidelines for cili. *Linguistic Issues in Language Technology (LiLT)*, Special Issue on Linking, Integrating and Extending Wordnets. Forthcoming.

_____; RUTTENBERG, A. *Survey on defining practices in ontologies: report*, 2013. Available at: <<http://definitionsinontologies.weebly.com/survey-report.html>>. Accessed on: 30 nov. 2017.

_____.; _____.; SCHREIBER, Y. Definitions in ontologies. *Cahiers de lexicologie*, n. 109, p. 173-206, 2016.

_____.; SMITH, B. The functions of definitions in ontologies. In: *Formal ontology in information systems: Proceedings of the 9th international conference (FOIS 2016)*, v. 283. FERRARIO R.; KUHN, W. (ED). Annecy, France: IOS Press, 2016. p. 37-50.

_____.; SCHREIBER, Y.; RUTTENBERG, A. Textual and logical definitions in ontologies. In: *Proceedings of The First International Workshop on Drug Interaction Knowledge Management (DIKR 2014)*, 1., *The Second International Workshop on Definitions in Ontologies (IWOOD 2014)*, and *The Starting an OBI-based Biobank Ontology Workshop (OBIB 2014)*, October 6-7, Houston, TX, USA. 2014.

SMITH, B. Introduction to the logic of definitions. In: *Proceedings of the International Workshop on Definitions in Ontologies (DO 2013)*, in *Proceedings of the 4th International Conference on Biomedical Ontology Workshops (ICBO 2013)*, v. 1061, July 7, Montreal, Canada. 2013.

SVENSÉN, B. *Practical lexicography: principles and methods of dictionary-making*. Oxford England; New York: Oxford University Press, 1993.

SWARTZ, N. *Definitions, dictionaries and meanings*. 1997. Available at: <<http://www.sfu.ca/~swartz/definitions.htm>>. Accessed on: 14 jul. 2017.

VÉZINA, R. et al. *La rédaction de définitions terminologiques*. Montréal: Office québécois de la langue française, 2009.

WIKIPEDIA. *Triangle*: wikipedia, the free encyclopedia, 2017. Available at: <<https://en.wikipedia.org/w/index.php?title=Triangle&oldid=785338335>>. Accessed on: 31 jul. 2017.

ANNEX

ONTOLOGIES REFERENCED IN THE EXAMPLES

Unless otherwise specified below, all the cited identifiers of terms from these ontologies are CURIEs,⁹ with the prefix <http://purl.obolibrary.org/obo/>.

- CHEBI: Chemical Entities of Biological Interest
- CMO: Clinical Measurement Ontology
- DDANAT: Dictyostelium discoideum anatomy
- EFO: Experimental Factor Ontology
(prefix <http://www.ebi.ac.uk/efo/>)
- FOODON: FoodOntology (FoodOn)
- IAO: Information Artifact Ontology
- MESH: Medical Subject Headings (MeSH)
(prefix MESH=<http://purl.bioontology.org/ontology/MESH>)
- MI: Molecular Interactions
- MP: Mammalian Phenotype Ontology
- NCI: National Cancer Institute Thesaurus
(prefix NCI=<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#>)
- NCRO: Non-Coding RNA Ontology
- OAE: Ontology of Adverse Events
- OBCS: Ontology of Biological and Clinical Statistics
- OBI: Ontology for Biomedical Investigations
- OBIB: Ontology for Biobanking
- OOST: Ontology of Organizational Structures of Trauma centers and Trauma systems
- SIBO: Social Insect Behavior Ontology
- SO: Sequence Types and Features Ontology
- UBERON: Uberon multi-species anatomy ontology

⁹ CURIE Syntax 1.0: A syntax for expressing Compact URIs, W3C Working Group Note 16 December 2010 (<https://www.w3.org/TR/curie/>, accessed July 14, 2017)

Elaboração de modelos de domínio em ontologias: a abordagem onomasiológica e a função da definição

Maria Luiza de Almeida Campos

Pós-Doutora pela Universidade de São Paulo (USP), São Paulo, Brasil.

Pós-Doutora pela Fundação Instituto Oswaldo Cruz (Fiocruz), Rio de Janeiro, Brasil.

Doutora em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT - UFRJ), Rio de Janeiro, Brasil.

Professora da Universidade Federal Fluminense. Departamento de Ciência da Informação e do Programa de

Pós-Graduação em Ciência da Informação (UFF – PPGCI), Niterói, Brasil.

E-mail: maria.almeida@pq.cnpq.br

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

RESUMO

A captura do conhecimento para a elaboração de modelos de domínios pode adotar uma abordagem semasiológica de perspectiva linguística (partindo do termo para o conceito) ou a abordagem onomasiológica, de perspectiva extralinguística (partindo do conceito para o termo). No que tange às ontologias, a adoção da abordagem semasiológica, na maioria das vezes, no que concerne à atividade de captura de conhecimento, reflete um procedimento de análise de textos eletrônicos, partindo da linguagem natural e do texto escrito em seus diversos formatos. Já a abordagem onomasiológica considera primeiramente que se chega ao termo através do extralinguístico, ou seja, a partir do entendimento do referente no âmbito de um domínio de especialidade; o termo seria o término de uma atividade de conotação sobre o referente. Este artigo aborda a metodologia apoiada na abordagem onomasiológica elaborada por I. Dahlberg, e demonstram-se os papéis que as definições possuem nesta atividade. A definição, neste sentido, é entendida como produto de um acordo estabelecido entre o modelizador/ontologista e o especialista do domínio; ela reflete o modelo pretendido de representação do mundo modelado.

Palavras-chaves: Modelo de domínio. Ontologia. Abordagem onomasiológica. Abordagem semasiológica. Captura do conhecimento.

Domain models development in ontologies: the onomasiological approach and the function of the definition

ABSTRACT

The knowledge capture to elaborate domain modeling can have a semasiological approach from a linguistic perspective (from the term to the concept) or the onomasiological approach of an extralinguistic perspective (from the concept to the term). Regarding the ontologies, the use of the semasiological approach in the majority of the times, what involves the activity of knowledge capture reflects a procedure of electronic texts analysis starting with natural language and written text in its various formats. The onomasiological approach, on the other hand, considers at first that the term is reached through the extralinguistic, that is, from the understanding of the referent within a specialty domain; the term would be the end of an activity of connotation about the referent. In this paper we expose the methodology supported by the onomasiological approach elaborated by I. Dahlberg, and it demonstrates the roles that the definitions have in this activity. The definition in this case is understood as the product of an agreement established between the modeler/ontologist and the domain expert; It reflects the intended model of representation of the modeled world.

Keywords: Domain model. Ontology. Onomasiological approach. Semasiological approach. knowledge capture.

Elaboración de modelos de dominio en ontologías: el enfoque onomasiológico y la función de la definición

RESUMEN

La captura del conocimiento para la elaboración de modelos de dominios puede adoptar el enfoque semasiológico de perspectiva lingüística (partiendo del término para el concepto) o el enfoque onomasiológico, de perspectiva extralingüística (partiendo del concepto para el término). En lo que atañe a las ontologías, la adopción del enfoque semasiológico, la mayoría de las veces, en lo que refiere a la actividad de captura de conocimiento, refleja un procedimiento de análisis de textos electrónicos, partiendo del lenguaje natural y del texto escrito en sus diversos formatos. Ya el enfoque onomasiológico considera primero que se llega al término a través del extralingüístico, o sea, a partir del entendimiento del referente en el ámbito de un dominio de especialidad; el término sería la terminación de una actividad de connotación sobre lo referente. En este artículo, abordamos la metodología apoyada en el enfoque onomasiológico elaborado por I. Dahlberg y demostramos los papeles que las definiciones poseen en esta actividad. Se entiende la definición, en este sentido, como producto de un acuerdo establecido entre el ontólogo y el especialista del dominio; ella refleja el modelo pretendido de representación del mundo modelado.

Palabras clave: *Modelo de dominio. Ontología. Enfoque onomasiológico. Enfoque semasiológico. Captura del conocimiento.*

INTRODUÇÃO

Ao longo dos últimos anos, as ontologias informacionais vêm sendo consideradas como sistemas formais para representar conhecimento em determinado domínio, atuando assim como um modelo formal para sistematizar a expressão do conhecimento que manipulam. Desde a década de 1990, foram conceituadas como uma especificação explícita de uma conceituação (GRUBER, 1993a). As ontologias são, portanto, consideradas como um corpo de conhecimento sistematizado em dado domínio, no qual entidades e relações que se manifestam são representadas através de dada formalização.

Sua finalidade, como muito bem apresentada por Bachimont (2006), atende a todos os que necessitam sistematizar a expressão do conhecimento que manipulam. Formalmente, uma ontologia é uma declaração de uma teoria lógica definindo um conjunto de primitivos de representação com os quais modelamos o domínio do conhecimento; tais primitivos são classes, atributos e relações. As definições explicitam este entendimento, enquanto os axiomas formais restringem a interpretação. A ontologia fornece uma linguagem de comunicação com o agente (SANTOS, 2010).

Os modelos de domínios resultantes dessas operações de conceituação em ontologias informacionais podem ser definidos como modelos que representam o conhecimento de uma área de aplicação (área de assunto, de missão ou de problema) e que devem servir como uma representação de consenso (ou modelo de referência) de uma conceituação compartilhada por determinada comunidade.

No âmbito da inteligência artificial, na década de 1990, Davis; Shrobe e Szolovits, (1993) iniciam uma reflexão sobre a variedade de representações possíveis que podem ser produzidas a partir de uma interpretação sobre o mundo, acentuando a importância de uma visão compartilhada sobre o conhecimento que se quer representar. Para Davis, apesar de a representação do conhecimento ser um dos conceitos mais centrais e familiares em inteligência artificial, a questão fundamental consiste em defini-lo, reforçando assim a posição de Newell (1982) sobre a imperfeição na representação do conhecimento (CAMPOS, 2010).

Nesse sentido, consideramos que determinar uma representação significa optar por um conjunto de compromissos ontológicos que sejam pragmaticamente possíveis de serem representados e, no caso de ontologias informacionais, possíveis

de serem representados computacionalmente. Logo, a representação do modelo de domínio em uma ontologia informacional requer uma escolha compartilhada de conceituações como uma forma de pensar sobre um mundo fechado e computacionalmente representável.

Nessa perspectiva, este artigo propõe discutir teoricamente um dos primeiros procedimentos para a elaboração de ontologias de domínio – o levantamento de *corpus* terminológico consistente que descreva este domínio. De modo geral, as propostas metodológicas pressupõem a existência ou mesmo a captura de uma estrutura de entidades **já existentes** sugerindo, a partir dessa estrutura, critérios para verificar a explicitação de propriedades e relações entre elas (FERNANDEZ; GOMEZ-PERZ; JURISTO, 1997; DING & FOO, 2002; FERNANDEZ-LÓPEZ, 2001; GANGEMI; PISANELLI; STEVE, 1999; GRUNINGER; FOX, 1995; GUARINO; WELTY, 2002; METAXIOTIS; PSARRAS; ASKNOUNIS, 2001; MIZOGUCHI, 2014; PASLARU-BONTAS, 2007; SURE; STAAB; STUDER, 2002; USCHOLD; KING, 1995). Entretanto, temos observado a necessidade de estudos que venham a discutir procedimentos metodológicos para a formação de *corpus* terminológico em ontologia de domínio. Esse *corpus*, consistentemente constituído, agregando as relações e axiomas, permite que se criem condições para a implementação de raciocínio automático (CAMPOS; ROCHA, 2017).

Em domínios de especialidade, as definições explicitam o conteúdo semântico de uma entidade. Tendo em vista que as entidades são descritas e definidas umas em relação às outras através de enunciados definitórios (DAHLBERG, 1981), consideramos que tais conteúdos presentes nas definições complementam a descoberta deste *corpus* terminológico.

A determinação de definições como um ponto de partida para a sistematização do *corpus* terminológico em ontologia de domínios se sustenta, neste estudo, por meio de uma abordagem onomasiológica. Esta abordagem é uma das orientações metodológicas para a elaboração de terminologias e tem sido recuperada

pela investigação na área da inteligência artificial e pelo desenvolvimento das ontologias. No entanto, as questões que envolvem a elaboração e a sistematização/classificação de domínios de conhecimento pertencem também à ciência da informação, através das teorias e metodologias visando à organização do conhecimento (VICKERY, 1997).

No campo da ciência da informação e dos estudos em linguagens documentárias, temos Dahlberg como uma das representantes da abordagem onomasiológica no campo da atividade tesaurográfica (DAHLBERG, 1980). Em contraposição à abordagem semasiológica, método também utilizado para a elaboração de terminologias de perspectiva linguística (partindo do termo para o conceito), a abordagem onomasiológica possui uma perspectiva extralinguística, partindo do conceito para o termo. A abordagem onomasiológica se caracteriza a partir do entendimento de uma concepção consensual da realidade sobre um objeto social, em que se identificam processos de seleção, realce e iluminação de alguns aspectos do ser e, por conseguinte, atenuação, ou mesmo ocultamento, de outros aspectos, nos quais esta “versão do referente” está consolidada nas definições (DAHLBERG, 1978a, 1978b, 1983; RIGGS, 1979, 1989a, b, 1996). Assim, o foco da análise se concentra no referente, no objeto (podendo ser abstrato ou concreto) e é a partir dele que as definições são elaboradas. Na verdade, a definição passa a ser um enunciado sobre o entendimento daquele referente em um dado domínio.

Consideramos assim discutir sob este enfoque as questões que envolvem a elaboração de modelos de domínios evidenciando o *modus operandi* para o levantamento do *corpus* terminológico.

Nesse sentido, apresentamos os papéis que as definições possuem neste processo: elas são ao mesmo tempo elementos que permitem a captura do conhecimento sobre o domínio, mas também são produtos de um acordo estabelecido entre o modelizador/ontologista e o especialista do domínio para a elaboração do modelo pretendido de representação do mundo modelado.

Este artigo está assim organizado: além desta introdução, apresentamos os tipos de definição em ontologias, considerando aqui as definições operativas como as mais apropriadas, ressaltando que, neste contexto, as definições são o resultado de uma prática de autoria, do especialista e do ontologista. Após, vamos discutir os procedimentos de captura do conhecimento sobre o domínio no âmbito da elaboração de modelos de domínios em ontologias. Investigamos os procedimentos semasiológicos e onomasiológicos, para finalizar com a apresentação do método onomasiológico de I. Dahlberg em sua teoria do conceito voltada para o referente. Logo depois, as considerações finais deste trabalho.

TIPOS DE DEFINIÇÃO EM ONTOLOGIAS

Podemos considerar que a função de uma definição, de forma geral, é principalmente ser capaz de explicitar, através do *definiens* (conjunto de enunciados sobre o que se quer definir) o seu respectivo *definiendum* (símbolo que se quer definir). Assim, a sua elaboração pressupõe, em ontologias, determinar condições necessárias e suficientes que permitam distinguir os objetos entre si em um dado mundo possível.

Nesse sentido, segundo Marconi (1997), o significado de um termo (o valor semântico de um sinal) surge quando falantes competentes – por exemplo, os especialistas em um determinado domínio – usam consistentemente uma série de letras ou símbolos para se comunicar sobre algo no mundo. Isso é o que é o referente do termo – por exemplo, objetos, processos ou atributos relevantes para a atividade em que os especialistas estão envolvidos. O significado é, então, algo como uma representação mental (ou conceito) que os especialistas têm em mente, quando usam um termo para se referir a um tipo de coisa no mundo onde está delimitado seu domínio de atuação. Nesse caso, a representação mental expressa por um termo é composta de certas características que são (pelo menos em parte) também apresentadas em uma definição. Logo, o termo, a representação mental e o enunciado definitório têm

todos, o mesmo referente. E é sobre ele, o referente, que as definições são elaboradas. Assim, uma definição é um enunciado encontrado como um pequeno texto no campo de definição de uma entrada de dicionário especializado e nos campos de definição ou comentário de uma ontologia (SEPPÄLÄ, 2015).

Considerando que o enunciado definitório, fruto da compreensão sobre o referente e da representação mental produzida pelo especialista, deve evidenciar o entendimento acordado em um dado domínio, a explicitação desse enunciado se dá de duas formas em ontologias: como definições textuais e como definições lógicas. Tal classificação é apresentada por Seppälä (2014), que afirma que as ontologias têm, por um lado, axiomas que fazem parte de uma definição lógica de termos e, por outro, definições em linguagem natural, as definições textuais. Em uma ontologia, uma definição textual é, idealmente, uma frase que designa o objeto e suas propriedades a partir de determinado comprometimento ontológico e que tem por função especificar o significado pretendido dos termos da ontologia a fim de evitar ambiguidades e erros de interpretação pelo homem. Esse tipo de definição da linguagem natural também é encontrado em dicionários terminológicos especializados. Dessa maneira, uma definição textual normalmente indica o tipo de classe a que as instâncias (o referente) do termo definido pertencem, e distingue essas instâncias de outras que se enquadram no mesmo tipo, listando uma ou mais das características dessas instâncias, sendo que em ontologia o primeiro elemento também deve ser o gênero a que a instância pertence. As definições lógicas possuem axiomas que restringem o significado pretendido de um termo afirmando as condições necessárias para a sua utilização. Eles assim funcionam de forma análoga às condições necessárias discutidas nas definições textuais e sua função de desambiguação é também análoga. Além disso, os axiomas possibilitam a formação das estruturas taxonômicas em ontologias e permitem que a máquina realize funções que permitirão estabelecer inferências sobre o domínio de forma automática.

As definições são o resultado de uma prática de autoria, do especialista e do modelizador/ontologista. Tais definições, lógicas ou textuais, necessitam, além de conter elementos que identificam o gênero próximo e a diferença específica, de outros que também possam identificar componentes, partes, função, finalidade, causa, etc. As definições que são dessa forma elaboradas foram denominadas por Hegenberg (1974) definições operacionais – aquelas em que os elementos da definição não identificam somente o gênero próximo e a diferença específica –, mas também elementos que procuram definir certas operações em que o conceito seria aplicado. Esta forma de definir faz parte de um movimento denominado operacionismo, que na década de 1960 do século passado vem se contrapor a procedimentos metodológicos de caráter mais conceitual e filosófico. Dentro dessa perspectiva, o operacionismo defende uma posição de que só é possível conhecer significados de um conceito a menos que possamos especificar as operações que foram realizadas, por nós ou por outrem, ao aplicar o conceito a qualquer situação concreta. Portanto, as definições operativas revelam a maneira de aplicação dos conceitos dentro de um dado contexto (CAMPOS, 2010).

No campo do trabalho terminológico, especificamente relacionado à elaboração de definições, diversas questões são apresentadas relacionadas à forma de construção de definições, ou seja, se estas estão de acordo com uma perspectiva linguística ou terminológica; se é possível definir algo a partir da existência ou não do referente no mundo (definições reais ou léxicas); se há diferenças entre os conceitos de descrição e definição; se a definição está relacionada a uma função comunicativa ou referencial da linguagem, entre outras tantas questões de fundamental importância relacionada ao ato de definir. No que tange ao objetivo do estudo que ora estamos desenvolvendo em ontologias, vamos nos concentrar na primeira questão que pretende colocar em discussão as perspectivas linguísticas e terminológicas para a captura do conhecimento sobre o domínio, e que

consequentemente está relacionada à dada forma semasiológica ou ontológica, respectivamente, de elaboração de definições.

ONTOLOGIAS E OS PROCEDIMENTOS DE CAPTURA DO DOMÍNIO: A ELABORAÇÃO DE MODELOS DE DOMÍNIOS

Em 1986, B.C. Vickery, em seu artigo “Knowledge Representation: a brief review” (VICKERY, 1986), já nos alertava sobre a importância da representação do conhecimento para sistemas computacionais e para as áreas de conhecimento para as quais esta questão era cara, como a ciência da informação, a inteligência artificial, a psicologia e a linguística. Nesse artigo, o enfoque não estava voltado para relacionar a representação do conhecimento às possibilidades tecnológicas da época: apresentava temáticas que se referiam à estrutura semântica de frases e enunciados, categorias e relações na análise de domínios e a representação do conhecimento por argumentos e axiomas; temáticas atualmente caras para aqueles que se dedicam ao estudo da representação e recuperação no ambiente da Web Semântica e de sistemas inteligentes. E, nesse mesmo trabalho, ele nos alerta que, já naquela época, a representação do conhecimento era relevante, além dos documentos e índices, e deveria ser discutida no âmbito da estrutura de registros e arquivos em bases de dados; da estrutura de dados nos programas de computador; da estrutura sintática e semântica da linguagem natural; da representação do conhecimento em inteligência artificial; dos modelos de memória humana – em todos esses campos nos quais é necessário decidir como o conhecimento pode ser representado de forma que as representações possam ser manipuladas.

Como se observa, a questão da representação do conhecimento vem preocupando o mundo da ciência da informação desde sua gênese. Atualmente, com o advento da Web Semântica, outras temáticas, na maioria das vezes não tão novas, vêm se incorporar a este mundo.

Modelos de domínios são obtidos a partir de processos que têm por objetivo uma representação do conhecimento de alta qualidade de fenômenos selecionados em algum domínio. Os modelos resultantes devem facilitar, dependendo do objetivo para o qual o modelo foi elaborado, a representação e a recuperação de informações/conhecimento; nesse caso, em sistemas de recuperação de informações. Para que um modelo possa ser uma descrição adequada da realidade do domínio, ele deve apresentar informações precisas e claras, não permitindo a ocorrência de ambiguidades sobre os aspectos que devem ser modelados.

Na elaboração desses modelos, procedimentos metodológicos são utilizados para a modelagem do domínio, ou seja, aqueles relacionados à modelagem conceitual e os relacionados à modelagem ontológica.

Modelagem conceitual consiste na identificação, análise e descrição dos elementos e restrições de um domínio do mundo real, normalmente utilizando uma linguagem de modelagem, a fim de que tais elementos sejam incorporados em um sistema de informação. Já modelagem ontológica pode ser definida como a criação de uma ontologia específica de domínio, através da captura das entidades relevantes do domínio e incorporação das mesmas em um conjunto de categorias que revelam sua natureza, por meio de uma linguagem de especificação de ontologia. Como pode-se perceber, existe uma forte semelhança entre esses dois tipos de modelagem, uma vez que ambas objetivam capturar e modelar elementos do mundo real. Porém, enquanto na modelagem conceitual busca-se estabelecer as relações entre os conceitos abstraídos de um domínio, na modelagem ontológica o objetivo é identificar os objetos e entender sua natureza por meio da descrição de suas propriedades. [Segundo Poli (2001), a primeira possui um caráter epistemológico, em oposição ao caráter ontológico da segunda]. Dessa forma, pode-se deduzir que as duas modelagens são possivelmente complementares, ou seja, a modelagem ontológica pode constituir-se numa base para modelagem conceitual, no sentido de prover ao projetista, de forma clara e sem ambiguidades, o conhecimento necessário sobre o domínio a ser modelado. (VILLELA, 2004, p. 243)

Consideramos, como Villela (2004), que ambas as abordagens se completam. Entretanto, nossa experiência tem evidenciado que, no âmbito dos modelos de domínio em

ontologias, a modelagem ontológica, que parte do extralinguístico, nos parece a mais apropriada, como defenderemos adiante.

Como Gruber (1993b) afirma as ontologias não pretendem representar todo o conhecimento, elas refletem determinada visão do mundo, em consenso dentro de uma comunidade. Ontologias informacionais podem ser consideradas, assim, como o estudo das características básicas de uma dada realidade. Visando explicitar a forma pela qual essas características básicas podem ser apropriadas pelo ontologista, é que apresentamos dois procedimentos que vêm sendo utilizados para a captura do conhecimento sobre um dado domínio, a perspectiva linguística e a extralinguística.

Alguns autores apresentam (CABRÉ, 2003; GOMES; CAMPOS; GUIMARÃES, 2010; LARA, 2010; SANTOS, 2010; HERMANS, 1989; ROCHE, 2006; TÁLAMO; LARA, 2009) a perspectiva linguística como um procedimento que parte do discurso verbal na forma de texto enquanto a perspectiva extralinguística parte do especialista enquanto detentor do conhecimento de um campo de especialidade. Temos, assim, um procedimento que parte do texto de especialidade e outro que parte do discurso do especialista para a elaboração de conceituações sobre o domínio e, no caso das ontologias de domínio, o modelo que o representa. Denomina-se a perspectiva linguística como aquela que se apoia em uma abordagem semasiológica, e a perspectiva extralinguística como aquela que se apoia em uma abordagem onomasiológica.

A abordagem semasiológica considera o termo como uma unidade lexical empregada no âmbito de um domínio de especialidade. A terminologia denominada moderna (CABRÉ, 2003) estuda os termos em contexto, ou seja, no interior do texto, sendo que uma palavra pode adquirir o estatuto de termo em função do seu papel no texto. Assim, a abordagem semasiológica parte do texto e da possibilidade de extração desses termos a partir de textos, assim como os seguidores dessa abordagem consideram o levantamento do corpus terminológico

para a elaboração de modelos como atividade que pressupõe a análise textual para a determinação desse corpus. No que tange às ontologias, a adoção dessa abordagem, na maioria das vezes, no que concerne à captura do conhecimento, reflete um procedimento de análise de textos eletrônicos, partindo da língua natural e do texto escrito em seus diversos formatos (SANTOS, 2010).

Por outro lado, a abordagem onomasiológica considera primeiramente que se chega ao termo através do extralinguístico, ou seja, a partir do entendimento do referente no âmbito de um domínio de especialidade, o termo seria o término de uma atividade de conotação sobre o referente. O termo denota o conceito. Nessa perspectiva, o conceito enquanto representação mental precede a forma linguística ou qualquer outro símbolo, sendo que as relações entre conceitos são estabelecidas em função de parâmetros extralinguísticos, a partir da identificação das propriedades do referente. No campo das ontologias informacionais, tal abordagem tem sido empregada na etapa de levantamento do conhecimento sobre o domínio, através de procedimentos de entrevistas com especialistas do domínio. Esse procedimento na elaboração de ontologias tem sido denominado captura do conhecimento sobre o domínio (CEUSTERS; SMITH, 2010; FERNANDES; GUIZZARDI; GUIZZARDI, 2011; GUIZZARD; MORALES-RAMIREZ; PERINI, 2014; TAIRAS; MERNIK; GRAY, 2008).

A construção das bases de conhecimento nas ontologias guarda alguma semelhança com sistema de recuperação de informação, entretanto ela tem suas especificidades. Assim, nesse sistema

o percurso da enunciação envolve posições relativas tanto ao enunciador (emissor) quanto ao enunciatário (receptor) que correspondem aos percursos onomasiológicos e semasiológicos, respectivamente. Nesse sentido, do conceito ao dito, tem-se a onomasiologia, e do dito à sua interpretação, a semasiologia. Logo, a orientação onomasiológica é da categoria do enunciador e a semasiológica é da categoria da recepção. (POTTIER, 1992 citado por LARA, 2010)

No caso das ontologias, em que a questão que se coloca é a descrição de forma consistente sobre um domínio que se deseja representar, a questão da recepção não se coloca, pois a comunicação se dá entre pares.

Para a elaboração de modelos de domínio em ontologias, consideramos defender a abordagem onomasiológica. Entretanto, vamos apresentar uma forma de procedimento de se chegar ao extralinguístico que não parte inicialmente do especialista, mas do estabelecimento de indagações sobre o referente no interior de um domínio. Tais indagações colocam o modelizador em uma posição de investigador sobre o domínio e não como um receptor do conhecimento do especialista ou mesmo de suas produções textuais, como tentaremos demonstrar.

Actualmente, o papel da terminologia na construção de ontologias não passa exclusivamente pelo texto. E essa tendência encontra-se, curiosamente, nos princípios metodológicos estabelecidos por Eugen Wüster [...]. Do nosso ponto de vista, embora uma perspectiva semasiológica não seja adequada para, sozinha, formalizar conhecimento num sistema computacional, pode contribuir com informação sobre os conceitos e as suas denominações, abarcando, por isso, o linguístico e o extralinguístico, uma vez que o significado dos termos contém também uma significação ligada ao acto comunicativo e cultural da sua produção, ligada, portanto, a uma parte conceptual e extralinguística. A terminologia que parte de uma abordagem linguística privilegia a verbalização do conhecimento. A terminologia que parte de uma abordagem extralinguística centra-se na conceptualização do conhecimento. (SANTOS, 2010, p. 75 e 89)

Para a realização de uma conceituação, vamos nos reportar a Dalhberg por meio de sua teoria do conceito voltada para o referente para apresentarmos o procedimento de como capturar o conhecimento sobre o domínio a partir de uma abordagem extralinguística.

A TEORIA DO CONCEITO E A CAPTURA DO CONHECIMENTO SOBRE O DOMÍNIO: A ABORDAGEM ONOMASIOLÓGICA

Dahlberg (1981) afirma que a compreensão sobre algo em dado domínio de especialidade pressupõe o conhecimento de três elementos: a) o seu referente, que são os objetos concretos e abstratos e que se encontram em um espaço extralinguístico; b) as características e propriedades desses referentes, que são proposições adequadas enunciadas no âmbito de um domínio sobre este referente; e c) a expressão verbal utilizada e que promove a comunicação entre um certo grupo de falantes. Isso se configura em seu triângulo conceitual. E esclarece: a fim de acumular conhecimento sobre o referente, deve-se coletar enunciados sobre o referente A, de modo a identificar suas características, em que dada característica nos revela a categoria à qual o referente pertence. As informações sobre B, suas características, ajudam a identificar o gênero próximo e seus demais relacionamentos como outros conceitos em dado domínio. O conhecimento de C determina a expressão verbal apropriada a ser usada.

A Teoria do Conceito (DAHLBERG, 1978a) fornece um método para analisarmos um item de referência em dado domínio, propiciando o que denominamos aqui captura de conhecimento sobre o domínio. Esse método se estabelece no âmbito de uma perspectiva onomasiológica, pois parte da análise de um elemento extralinguístico (o referente), a partir do qual são estabelecidas perguntas sobre este elemento, a fim de permitir certa sistematização desse referente em dado domínio. Nesse sentido, a predicação que ocorre sobre o referente não é feita de maneira aleatória; sua teoria nos indica um *modus operandi* de como devem ser realizadas as asserções sobre o referente em dado domínio, como veremos adiante. Aqui ainda é necessário ressaltar o aspecto sistemático de sua teoria, ou seja, possibilita a elaboração de definições que se interconectam formando um todo coeso que, neste caso, representa o conhecimento no âmbito de um domínio de especialidade. Por isso é que em seus trabalhos ela afirma que os conceitos se apresentam uns em relação aos outros, formando um sistema de conceitos.

Seu *modus operandi* parte, assim, da identificação das propriedades do referente. Logo, através da formulação dos enunciados sobre as propriedades dos referentes se obtêm as características dos respectivos conceitos. Para Dahlberg, o ato de definir é uma construção de aspectos selecionados do referente realizado pelo “olhar” de um grupo que aquiesce sobre o seu entendimento em um dado espaço de conhecimento. Portanto, esses elementos que formam o conceito – as características – são propriedades dos referentes (objetos concretos ou abstratos) que, no nível do conceito, passam a ser também características do conceito. As características são, pois, representações que se estabelecem a partir de um “acordo” sobre as propriedades dos referentes. Por isso, ‘conceito’ em Dahlberg representa uma unidade de conhecimento sobre o acordo firmado no interior de um domínio de conhecimento.

As características também são conceitos e são usadas para classificar e definir outros conceitos. Por exemplo, uma das características do conceito ‘Verniz à base de resina’ é ‘resina’ e ‘resina’ também é um conceito. Se não soubermos o que é ‘resina’, não poderemos entender corretamente o que seja um ‘verniz à base de resina’. Nesse sentido, o conceito nunca é estabelecido de forma isolada, mas sempre em relação com outros. Ele se estabelece através da comparação com outros conceitos, quando se identificam semelhanças e diferenças (por meio da análise e comparação das características), o que leva à reunião de conceitos, relacionados de maneira variada. Por exemplo, existem semelhanças entre os conceitos ‘fungicida’, ‘herbicida’ e ‘inseticida’. Eles têm em comum a característica ‘que mata’. O que os diferencia entre si é o objeto que sofre a ação da substância nomeada: o fungo, a erva daninha, o inseto, respectivamente. Se acrescentarmos o conceito ‘biocida’, veremos que ele também tem a característica ‘que mata’ além da característica ‘organismo vivo’ que, num grau de generalização maior, engloba os outros três, estabelecendo uma posição de superordenação (GOMES; CAMPOS, 2002).

A identificação das características dos conceitos é de fundamental importância para o entendimento do conceito em dado domínio, e assim consequentemente possibilita sua classificação ou, dito de outro modo, sua ordenação no interior

deste domínio. As características são elementos da definição e revelam as conexões existentes entre o conceito em questão e os demais através de suas relações. Dahlberg (1978b) classifica as características em essenciais e acidentais. As características essenciais são consideradas como os elementos necessários e suficientes, em um dado espaço de conhecimento, para determinar de que tipo um conceito é, ou seja, determinar sua identidade. Já as características acidentais são opcionais para um determinado tipo de conceito, ou seja, dito de outra maneira, não são necessárias ou constantes. Assim, por exemplo, para ser um coração, algumas características essenciais são: ser órgão, ser feito de músculo e circular sangue; e algumas características acidentais são: ser grande ou apresentar sopro. A característica essencial permite a ordenação do conceito no domínio, tendo por função estruturar modelos de domínios, como as taxonomias elaboradas para a estruturação do conhecimento de um domínio em ontologias. Por isso ela afirma que as características essenciais *definem*; as características acidentais *descrevem* (DAHLBERG, 1978b). A definição para ela é ordenadora, na medida em que possibilita a classificação do objeto no domínio. Por isso o primeiro elemento da definição deve revelar o gênero próximo.

A definição revela, pois, a ordenação dos conceitos a partir de um compromisso ontológico (acordo firmado) estabelecido em um dado domínio; ela é a consequência dessa ordenação (classificação).

Como dito anteriormente, nessa perspectiva, os conceitos se definem uns em relação aos outros. Nesse processo, ficam evidenciadas certas relações que fornecem as bases para sua ordenação. Sempre que houver entre dois conceitos características iguais, opostas, semelhantes ou outras, pode-se considerar que existem relações entre eles. As relações são classificadas em quantitativas e qualitativas (DAHBERG, 2014).

A espécie quantitativa compara dois conceitos do ponto de vista estritamente formal; aqui estão incluídas as seguintes relações: de identidade, de inclusão, de interseção e de disjunção. A espécie qualitativa

considera os aspectos formal e material, podendo ser classificadas como relação formal/categorial; relação material/paradigmática; relação sintagmática funcional. Essa rede de relações constitui o sistema de conceitos. A relação formal – categorial depende completamente da espécie de item de referência que se está analisando, no caso, fenômenos, objetos gerais, objetos materiais, quantidades, qualidades, comparações, operações, estados, processos, períodos, posições, lugares no espaço, por exemplo. Esta relação reúne os conceitos dentro de uma mesma categoria. Isso ocorre porque tais conceitos são todos de mesma natureza, ou seja, a partir da análise do conceito, as características essenciais levam à mesma categoria. A relação material-paradigmática permite organizar conceitos de mesma natureza, ou seja, aqueles que estão no interior de uma categoria. Pode ser hierárquica, partitiva e de oposição. A relação hierárquica baseia-se na relação lógica de implicação e é de dois tipos: relação de abstração (relação gênero-espécie) e relação lateral (relação dos conceitos no renque). A relação partitiva existe entre o todo e suas partes, e as partes também podem estar relacionadas entre si. A relação de oposição pode ser de três espécies: relação de oposição contraditória, relação de oposição contrária, relação positiva-indiferente-negativa. A relação funcional-sintagmática se dá entre conceitos de diferentes categorias. A partir de um conceito que denote um processo ou operação, leva a conceitos que suplementam essas ações, como na seguinte sequência: produção – produto – produtor – comprador, etc. (CAMPOS, p. 104, 2001).

Nos procedimentos de elaboração de modelos de domínios, as relações categoriais são fundamentais, pois permitem uma primeira estruturação do conceito no interior de um domínio.

Podemos ver que as categorias têm uma capacidade de estrutura: não apenas estruturam, de fato, todos os nossos elementos de conhecimento e unidades do conhecimento; elas fornecem, ao mesmo tempo, por este meio, o esqueleto, os ossos e tendões para estruturar todo o nosso conhecimento. Com seu uso consciencioso, então, o corpo do nosso conhecimento pode se manter unido, pode se mover, pode se manter flexível – e pode crescer organicamente. (DAHLBERG, 1978a, p. 34)

As categorias apresentadas por Dahlberg são metaconceitos que possibilitam que o modelizador chegue a qualquer domínio e consiga estabelecer uma primeira ordenação. Elas possibilitam a identificação dos referentes no interior de um domínio. No nível de uma representação, são classes de grande abrangência em um domínio, acima das quais não se pode colocar qualquer conceito. Em seu trabalho recente, Dahlberg (2014) apresenta a seguinte classificação para as categorias, comparando as categorias de Aristóteles e de Ranganthan: Entidades (princípios, objetos materiais, objetos imateriais) – em Aristóteles seriam as substâncias e, em Ranganthan, Personalidade; Qualidades (qualidades, quantidades, relação) – em Aristóteles se apresenta distinguida pelas categorias qualidade, quantidade e relação e, em Ranganthan, é representada pela categoria matéria; Atividades – que são as ações, processos e estados em Aristóteles e, em Ranganthan, denomina-se energia; por último, Dimensão – que em Aristóteles é considerada como as categorias lugar, tempo, condição e, em Ranganthan, é identificada pelas categorias espaço e tempo.

As categorias são utilizadas para a ordenação dos conceitos, mas, no nível do referente, do extralinguístico, elas também são utilizadas para se fazer perguntas sobre o domínio: em determinado domínio existem princípios, objetos materiais, imateriais, qualidades, processos, condição, entre outros?

A abordagem onomasiológica não inicia com o levantamento de *um corpus terminológico*, mas com a identificação dos referentes que podem popular as categorias – que funcionam aqui como metarrepresentações do domínio. Essa forma de adentrar no domínio nos leva, no caso das ontologias em que desejamos que o conhecimento possível possa ser explicitado, a poder realizar a captura do conhecimento auxiliado não somente pelo especialista, mas também pela consulta a glossários ou outros textos de especialidade sobre o domínio, para encontrar conceitos que respondam positivamente ou não a essas categorias. Esse é um procedimento de idas

e vindas, de análise e síntese. E, assim, quando se dá a identificação a qual categoria o referente pertence, é possível determinar as relações (lógicas, de parte – todo, funcionais) no interior dessas categorias, de forma a se construir um modelo de domínio para ontologias. A classificação desses conceitos no interior do domínio nos permitirá então determinar a sua enunciação por meio de uma definição que espelhe o acordo estabelecido.

Nesse sentido, mesmo que sejam utilizadas em um primeiro momento definições como um texto de especialidade, elas sofrerão alterações, para que seja possível evidenciar uma representação que expresse o compromisso ontológico assumido, no interior do mundo possível a ser representado. As definições são fruto desse acordo estabelecido no interior do domínio.

CONCLUSÕES

Neste estudo apresentamos discussões que envolvem as abordagens semasiológicas e onomasiológicas para a captura do conhecimento sobre domínios de especialidade. No caso de ontologias, o modelo de domínio elaborado deve ser expresso por meio de uma taxonomia que espelhe as conceituações e também de axiomas que descrevem as operações do referente no interior desse domínio. As definições em ontologias devem atender à leitura de humanos e máquinas e, assim, são a soma das descrições expressas na taxonomia e nos axiomas. Portanto, as definições em ontologias, além de serem apresentadas em forma de texto para serem lidas por humanos, devem também ser apresentadas por meio de axiomas que descrevem o objeto e suas relações no interior de um dado domínio. Consideramos que a abordagem onomasiológica, pautada em uma forma de captura de conhecimento, adentra o domínio através do entendimento dos objetos de conhecimento desse domínio, ou seja, que parte do extralinguístico apresenta um potencial metodológico bastante adequado à elaboração de modelos de domínio em ontologias, pois permite uma compreensão consensual de como os referentes se relacionam em um mundo fechado.

Consideramos, assim, que o papel das definições em uma ontologia é, portanto, precisar os atributos do referente e suas relações de forma consensual, assim assegurando uma herança através de um tipo de hierarquia, além da descrição de outros elementos que estão em relação no interior do domínio. A coerência nas definições e, por conseguinte, na classificação, exige uma perspectiva unificadora requerida pelo domínio a ser especificado, para que se possa atender a uma representação consistente e que atenda a determinado propósito.

AGRADECIMENTOS

Agradecimentos ao apoio da Agência de fomento CNPq/Brasil, pelo auxílio através de bolsa de produtividade pesquisa que possibilitou o desenvolvimento dos estudos.

REFERÊNCIAS

- BACHIMONT, B. *Qu'est-ce c'est une ontologie?*. *Technolangue.net*, 3 juil. 2006. Disponível em: <http://www.technolangue.net/imprimer.php?id_article=280>. Acesso em: 15 ago. 2017.
- CABRÉ, T. Theories of terminology their description, prescription and explanation. *Terminology*, v. 9, n. 2, p. 163-199, 2003.
- CAMPOS, M. L. A. *Linguagem documentária: teorias que fundamentam sua elaboração*. Niterói: UFF, 2001.
- _____. O papel das definições na pesquisa em ontologia. *Perspectivas em Ciência da Informação*, v. 15, n. 1, p. 220-238, 2010. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/v/a/8769>>. Acesso em: 15 ago. 2017
- _____. ROCHA, R. O papel das definições na formação de corpus terminológico para elaboração de ontologias de domínios: uma experiência a partir de extração automática e manual. In: CONGRESSO ISKO ESPANHA-PORTUGAL, 3., 2017, Lisboa. *Anais...* Lisboa: Faculdade de Letras da Universidade de Coimbra, 2017. No prelo.
- CEUSTERS, W.; SMITH, B. Foundations for a realist ontology of mental disease. *Journal of Biomedical Semantics*, v. 1, n. 10, p. 1-23, Apr. 2010.
- DAHLBERG, I. A referent-oriented analytical concept theory of interconcept. *International Classification*, v. 5, n. 3, p.142-150, 1978b.
- _____. Conceptual definitions for interconcept. *International Classification*, v. 1, n. 8, p.16-22, 1981.
- _____. *Die pilotstudie DB-Thesaurus*. Frankfurt: Deutsche Bibliothek, 1980.
- _____. *Ontical structures and universal classification*. Bangalore: Sarada Ranganathan Endowment for Library Science, 1978a.
- _____. Terminological definitions: characteristics and demands. In: PROBLÈMES de la définition et de la synonymie en terminologie. Québec: Girsterm, 1983. p. 13-51.
- _____. *Wissensorganisation: Entwicklung, Aufgabe, Anwendung, Zukunft*. W✓rzburg: Ergon, 2014. p. 35-51
- DAVIS, R.; SHROBE, H.; SZOLOVITS, P. What is knowledge representation?. *AI Magazine*, n. 14, v. 1, p.17-33, 1993.
- DING, Y.; FOO, S. Ontology research and development part 2: a review of ontology generation. *Journal of Information Science*, v. 5, n. 28, p. 375-388, 2002.
- FERNANDES, P. C. B.; GUIZZARDI, R. S. S.; GUIZZARDI, G. Using goal modeling to capture competency questions in ontology-based systems. *Journal of Information and Data Management (JIDM)*, v. 2, n. 3, p. 527-540, 2011.
- FERNANDEZ-LÓPEZ, M. Overview of methodologies for building ontologies. In: WORKSHOP ON ONTOLOGIES AND PROBLEM-SOLVING, METHODS: LESSONS LEARNED AND FUTURE TRENDS, 1999, Stockholm. *Proceedings...* Stockholm: Richard Benjamins, 2001. p. 26-34.
- _____. GÓMEZ-PÉREZ, A.; JURISTO, N. Methontology: from ontological art towards ontological engineering. In: SPRING SYMPOSIUM ON ONTOLOGICAL ENGINEERING, 1997, Palo Alto, CA. *Proceedings...* Menlo Park, CA: The AAAI Press, 1997. p. 33-40.
- GANGEMI, A.; PISANELLI, D. M.; STEVE, G. An overview of the ONIONS Project: applying ontologies to the integration of medical Terminologies. *Data Knowledge Engineering*, v. 2, n. 31, p. 183-220, 1999.
- GOMES, H.; CAMPOS, M. L. A. Aspectos teóricos aplicados às linguagens documentárias: módulo 2. 2002. Material didático para curso à distância no Sesc.
- _____. GUIMARÃES, L. S. Organização da informação e terminologia: a abordagem onomasiológica. *DataGramZero*, v. 11, n. 5, out. 2010. Disponível em: <<http://www.brapci.ufpr.br/brapci/v/a/9489>>. Acesso em: 15 ago. 2017
- GRUBER, T. A translation approach to portable ontology specifications. *Knowledge Acquisition*, London, v. 5, n. 2, p. 199-220, 1993b.
- _____. Towards principles for the design of ontologies used for knowledge sharing. *International Journal Human Computer Studies*, v. 43, n. 5/6, p. 907-928, 1993a.

- GRUNINGER, M.; FOX, M. S. Methodology for the design and evaluation of ontologies. In: WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWLEDGE SHARING, AT THE INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 14., 1995, Montreal. *Proceedings...* Montreal: IJCAI, 1995.
- GUARINO, N.; WELTY, C. Evaluating ontological decisions with OntoClean. *Communications of the ACM*, v. 45, n. 2, p. 61-65, Feb. 2002.
- GUIZZARDI, R. S. S.; MORALES-RAMIREZ, I.; PERINI, A. A goal-oriented analysis to guide the development of a user feedback ontology. In: WORKSHOP CO-LOCATED WITH CHAISE. CEUR WORKSHOP, 7., 2014, Thessaloniki. *Proceedings...* Thessaloniki: CEUR-WS, 2014. p. 16-17. v. 1157.
- HEGENBERG, L. *Definitions: termos teóricos e significado*. São Paulo: Cultrix, 1974.
- HERMANS, A. La définition des termes scientifiques. *Meta*, v. 34, n. 3, p. 529-532, 1989.
- LARA, M. L. G. Terminologia e Ciência da Informação. In: ALVES, I. M. (Org.). *Os estudos lexicais em diferentes perspectivas*. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas, 2010. p. 8-16. v. 2.
- MARCONI, D. *Lexical competence: language, speech and communication series, a Bradford book*. Cambridge: The MIT Press, 1997.**
- METAXIOTIS, K.; PSARRAS, J. E.; ASKOUNIS, D. Building ontologies for production scheduling systems: towards a unified methodology. *Information Management & Computer Security*, v. 9, n. 1, p. 44-51, 2001.
- MIZOGUCHI, R. What ontological engineering can do for solving real-world problems. *Knowledge and Systems Engineering Advances in Intelligent Systems and Computing*, v. 244, p. 3-3, 2014.
- NEWELL, A. The Knowledge level. *Artificial Intelligence*, v. 18, p. 87-127, 1982. Disponível em: <<http://repository.cmu.edu/cgi/viewcontent.cgi?article=2616&context=compsci>>. Acesso em: 15 ago. 2017.
- PASLARU-BONTAS, E. *A contextual approach to ontology reuse: methodology, methods and tools for the semantic web*. 2007. Dissertação (Mestrado em Matemática)- Freien Universitat Berlin, Berlin, 2007.
- POLI, R. *Framing ontology*. 2001. Disponível em: <[http://www.formalontology.it/Framing first.htm](http://www.formalontology.it/Framing%20first.htm)>. Acesso em: 15 ago. 2017.
- POTTIER, B. *Sémantique générale*. Paris: PUF, 1992.
- RIGGS, F. W. Information Science and social science: the need for onomantics. *International Forum for Information and Documentation*, v. 1, n. 14, p.12-21, 1989a.
- _____. Onomantics and terminology. *Knowledge Organization*, n. 23, p. 25-3, 1996.
- _____. Terminology and lexicography: their complementarity. *Terminology and lexicography: their complementarity*, v. 2, n. 2, p. 90-110, 1989b.
- _____. Terminology for the Social Sciences. In: INTERNATIONAL SYMPOSIUM THEORETICAL AND METHODOLOGICAL PROBLEMS OR TERMINOLOGY, 1979, Moscow. *Proceedings...* Moscow, 1979. p. 591-606.
- ROCHE, C. Lexical and conceptual structures in ontology. In: INTERNATIONAL CONFERENCE ON INDUSTRIAL, ENGINEERING AND OTHER APPLICATIONS OF APPLIED INTELLIGENT SYSTEMS, 2006, Annecy. *Proceedings...* Heidelberg: Springer Berlin / Heidelberg, 2006. p. 1034-1041.
- SANTOS, C. S. A. *Terminologia e ontologias: metodologias para representação do conhecimento*. 2010. 245 f. Tese (Doutorado em Linguística) – Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, Departamento de Línguas e Culturas, Universidade de Aveiro, Lisboa, 2010. Disponível em: <<https://ria.ua.pt/bitstream/10773/2876/1/2010000702.pdf>>. Acesso em: 15 ago. 2017.
- SEPPÄLÄ, S. An ontological framework for modeling the contents of definitions. *Terminology*, v. 21, n. 1, p. 23-50, June 2015.
- _____. SCHREIBER, Y.; RUTTENBERG, A. Textual and logical definitions in ontologies. In: INTERNATIONAL WORKSHOP ON DEFINITIONS IN ONTOLOGIES, 2., 2014, Houston. *Proceedings...* Buffalo: University at Buffalo, 2014. Disponível em: <<http://ceur-ws.org/Vol-1309/paper3.pdf>>. Acesso em 15 ago. 2017.
- SURE, Y.; STAAB, A.; STUDER, R. Methodology for development and employment of ontology based knowledge management applications. *SIGMOD Record*, v. 31, n. 4, p.18-23, 2002.
- TAIRAS, R.; MERNIK, M.; GRAY, J. Using ontologies in the domain analysis of domain-specific languages. In: INTERNATIONAL WORKSHOP ON TRANSFORMING AND WEAVING ONTOLOGIES IN MODEL DRIVEN ENGINEERING, 1., 2008, Toulouse. *Proceedings...* Toulouse: Springer, 2008. Disponível em: <<http://ceur-ws.org/Vol-395/paper02.pdf>>. Acesso em: 15 ago. 2017.
- TÁLAMO, M. F. T. G. A. M.; LARA, M. L. G. Interfaces entre linguística, terminologia e documentação. *Brazilian Journal of Information Science*, v. 3, n. 2, p. 54-70, 2009. Disponível em: <<http://www.brapci.ufpr.br/brapci/v/a/9185>>. Acesso em: 15 ago. 2017.

USCHOULD, M.; KING, M. Towards a methodology for building ontologies. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWLEDGE, 14., 1995, Montreal. *Proceedings...* Montréal: DBLP, 1995. Disponível em: <<http://www.aiai.ed.ac.uk/project/oplan/documents/1995/95-ont-ijcai95-ont-method.pdf>>. Acesso em: 15 ago. 2017.

VICKERY, B. C. Knowledge representation: a brief review. *Journal of documentation*, v. 42, n. 3, p. 145-59, Sept. 1986.

_____. Ontologies. *Journal of Information Science*, v. 23, n. 4, p. 227-286, 1997.

VILLELA, M. L. B.; OLIVEIRA, A. P.; BRAGA, J. L. Modelagem ontológica no apoio à modelagem conceitual. In: SIMPÓSIO BRASILEIRO DE ENGENHARIA DE SOFTWARE, 18., 2004, Brasília, DF. *Anais...* Brasília, DF: SBES, 2004. p. 241-256.

Desambiguação de relações em tesouros e o seu reúso em ontologias

Benildes Coura Moreira dos Santos Maculan

Doutora em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG – Brasil. Professora Adjunta da Escola de Ciência da Informação (ECI) da UFMG – Belo Horizonte, MG – Brasil.

Lattes: <http://lattes.cnpq.br/5336218259257800>

E-mail: benildes@gmail.com

Elisângela Cristina Aganette

Doutora em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG – Brasil. Professora Adjunta da Escola de Ciência da Informação (ECI) da UFMG – Belo Horizonte, MG – Brasil.

Lattes: <http://lattes.cnpq.br/5336218259257800>

E-mail: elisangelaaganette@hotmail.com

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

RESUMO

Um tesouro é composto de conceitos organizados por uma rede paradigmática (relações essenciais e estáveis entre termos e conceitos), e uma rede sintagmática (relações específicas em um contexto de uso). Tesouros e ontologias são sistemas de organização do conhecimento (SOCs) capazes de representar a base de conhecimento de um domínio, a partir de regras e critérios de estruturação. A base terminológica de tesouros pode ser utilizada na construção de ontologias. Porém, nem sempre as relações estabelecidas na estrutura dos tesouros, por terem uma semântica ainda considerada fraca, podem ser reutilizadas nas ontologias. Isso ocorre porque as ontologias exigem a explicitação das relações atribuídas entre objetos, propriedades, classes e indivíduos, e também a declaração de axiomas. Em 2015, trabalho apresentou o refinamento semântico de relações na estrutura de um tesouro de agropecuária. Tendo como parâmetros (1) o resultado do trabalho mencionado, (2) a literatura sobre a conversão de tesouros em ontologias e (3) as recomendações para a construção de tesouros e ontologias, este artigo explora a questão sobre o reúso do conhecimento estruturado na estrutura de tesouros na construção de ontologias. Os resultados apontam que os relacionamentos semanticamente refinados na estrutura do tesouro podem ser mais bem aproveitados na criação de ontologias, mas que não há uma correlação perfeita. Este estudo contribui com a ciência da informação, uma vez que traz insumos para maior entendimento sobre as semelhanças e diferenças estruturais entre tesouros e ontologias para o reúso de conhecimento.

Palavras-chave: Tesouro. Ontologias. Reúso de conhecimento.

An approach to deal with ambiguity of relationships in thesauri for reuse in ontologies

ABSTRACT

A thesaurus is composed of concepts organized by a paradigm network (essential and stable relations between terms and concepts), and a syntagmatic network (specific relations in a context of use). Thesauri and ontologies are knowledge organization systems (SOCs) that are capable of representing the knowledge base of a domain, based on rules and structuring criteria. The terminological basis of thesauri can be used in the construction of ontologies. However, not always the relations established in the structure of thesauri, because they have a semantic still considered weak, can be reused in the ontology. This is because ontologies require explicit relations between objects, properties, classes and individuals, as well as the declaration of axioms. In 2015, a work presents the semantic refinement of relationships in the structure of a thesaurus of the Agricultural area. (1) the results of the aforementioned work, (2) the literature on the conversion of thesauri into ontologies and (3) recommendations for the construction of thesauri and ontologies, this article explores the question about the reuse of structured knowledge in the structure of thesauri in the construction of ontologies. The results point out that the semantically refined relationships in the thesaurus structure can be better used in the creation of ontologies, but that there is no perfect correlation. In Information Science, this study contributes because it brings inputs to a greater understanding about the similarities and structural differences between thesauri and ontologies for reuse of knowledge.

Keywords: *Thesaurus. Ontologies. Reuse of knowledge.*

Desambiguación de relaciones en tesauros y su reutilización en ontologías

Resumen: Un tesauro se compone de conceptos organizados por una red paradigmática (relaciones esenciales y estables entre términos y conceptos), y una red sintagmática (relaciones específicas en un contexto de uso). Tesauros y ontologías son sistemas de organización del conocimiento (SOC) que son capaces de representar la base de conocimiento de un dominio, a partir de reglas y criterios de estructuración. Se puede utilizar la base terminológica de tesauros en la construcción de ontologías. Pero no siempre las relaciones establecidas en la estructura de los tesauros, por tener una semántica aún considerada débil, pueden ser reutilizadas en las ontologías. Esto ocurre porque las ontologías exigen la explicitación de las relaciones asignadas entre objetos, propiedades, clases e individuos, y también la declaración de axiomas. En 2015, trabajo presentó el refinamiento semántico de relaciones en la estructura de un tesauro del área de la Agropecuaria. Teniendo como parámetros (1) el resultado del trabajo, (2) la literatura sobre la conversión tesauro en ontologías y (3) las recomendaciones para la construcción de tesauros y ontologías, este artículo explora la cuestión de la reutilización de conocimiento estructurado en la estructura de tesauros en la construcción de ontologías. Los resultados apuntan que las relaciones semánticamente refinadas en la estructura del tesauro pueden ser más bien aprovechadas en la creación de ontologías, sin embargo, que no hay una correlación perfecta. Este estudio contribuye con la Ciencia de la Información, una vez que trae insumos para una mejor comprensión de las semejanzas y diferencias estructurales entre tesauros y ontologías para reúso de conocimiento.

Palabras-clave: *Tesauro. Ontologías. Reúso de conocimiento.*

INTRODUÇÃO

Os tesouros, um tipo de sistema de organização do conhecimento (SOC), são desenvolvidos para representar o conhecimento de um domínio, a partir de um conjunto de termos descritores, preferidos e não preferidos, inter-relacionados em um sistema conceitual (estrutura do tesouro), fazendo o controle da terminologia em níveis diferenciados de controle e padronização. No âmbito da ciência da informação (CI), as ontologias são também um tipo de SOC, desenvolvidas como um modelo conceitual e base de conhecimento de um domínio, para serem interpretáveis pela máquina. A sua criação, em geral, é realizada utilizando-se linguagem formal de representação de dados capazes de proporcionar inferências a partir das relações atribuídas entre seus diferentes elementos.

Na literatura sobre tesouros, é usual a recomendação para que a terminologia de tesouros seja utilizada na construção de outros tesouros de mesmo domínio, com propósitos diferentes, e a mesma recomendação é encontrada para a criação de ontologias. Porém, não há indicação para reuso das relações dos tesouros. Em pesquisa realizada por Maculan (2015), a autora observa que na literatura internacional há substancial reuso de conhecimento estruturado de tesouros na criação de ontologias, inclusive das relações, fato que não se repete no Brasil. Para Boccato, Ramalho e Fujita (2008, p. 207), essa reutilização é um “fato [que se] justifica [...], pois eles [os tesouros] podem servir como substratos teóricos para a construção de ontologias”. Considera-se que o conhecimento estruturado nos tesouros está pautado nas bases teóricas da biblioteconomia, CI e terminologia, sendo, a princípio, uma fonte confiável para ser reutilizada na construção de outros instrumentos. Contudo, o problema que se apresenta é se essa fonte, sobretudo no que concerne às relações semânticas estabelecidas entre termos e conceitos, efetivamente contribui para a criação de ontologias. Nesse sentido, este artigo parte do pressuposto de que, no Brasil, há pouca discussão sobre o reuso da estrutura de tesouros para o desenvolvimento de ontologias, em especial das relações semânticas. Para

analisar essa questão, utilizam-se como parâmetros a literatura sobre o tema, as normas de construção e os resultados do refinamento semântico de relações em tesouros de Maculan (2015).

O artigo está organizado da seguinte maneira: a metodologia descreve os procedimentos adotados; a seção sobre ambiguidade aponta os problemas que afetam a elaboração de proposições em tesouros e ontologias; a seção dos tesouros apresenta as relações de sua estrutura conceitual; a seção das ontologias proporciona as bases de sua construção; o estudo de caso oferece a proposta de refinamento de relações em um tesouro; a seção de reuso de conhecimento estruturado dos tesouros na criação de ontologias apresenta os resultados e as análises; por fim, têm-se as considerações finais, seguidas da lista de referências utilizadas no artigo.

METODOLOGIA

Este artigo apresenta resultados de um estudo exploratório-descritivo que visou ao maior entendimento sobre a temática do reuso de conhecimento estruturado de tesouros para a construção de ontologias, em especial no que se refere às relações estabelecidas. Este estudo se justifica, uma vez que é tema ainda pouco explorado na literatura nacional. Primeiramente, buscou-se na literatura melhor visão do problema e, em seguida, delimitou-se este estudo, com os procedimentos: (1) estabelecer a proposta de estudo; (2) formular o problema estudado; (3) mapear a literatura sobre o tema; (4) estabelecer os parâmetros para a análise; (5) apresentar a análise dos resultados. Nesse sentido, estabeleceu-se a proposta de estudo, a partir do pressuposto da carência de estudos nacionais sobre o reuso de conhecimento estruturado de tesouros na construção de ontologias. Em seguida, realizou-se um mapeamento na literatura sobre o tema, ao se constatar a escassez de discussão sobre essa temática em âmbito nacional. Segundo Maculan (2015), cerca de 15 trabalhos trataram a respeito da conversão de tesouros em ontologias na literatura internacional até julho de 2013, e, no Brasil, há somente Campos et al. (2008), que trabalharam com a conversão do tesouro do folclore brasileiro em uma ontologia de domínio

e discutiui o reuso das relações de tesouros na criação de ontologias. A busca bibliográfica realizada neste estudo confirmou o pressuposto aqui estipulado.

Os parâmetros para as análises tiveram por base a literatura sobre conversão de tesouros em ontologias, recomendações para o desenvolvimento de tesouros e ontologias, em especial, da norma ISO 25964, Parte 1 e 2 (2011; 2013), e os resultados de Maculan (2015). A autora trabalhou com o modelo de Soergel et al. (2004) para a reengenharia de tesouros que determina a explicitação dos relacionamentos na estrutura dos tesouros, tornando-os com maior expressividade (refinamento) semântica. Com base nisso, este estudo buscou responder às questões: (1) a explicitação das relações de equivalência nos tesouros pode contribuir para a criação de equivalências na estrutura de ontologias?; (2) a explicitação das relações hierárquicas (gênero-espécie, todo-parte e instância) pode contribuir para a criação de hierarquias na estrutura de ontologias?; (3) a explicitação das relações associativas nos tesouros pode contribuir para a criação de relações não hierárquicas na estrutura de ontologias?; (4) que função pode ter a estrutura dos papéis *Qualia* na construção de ontologias?

AMBIGUIDADE, DESAMBIGUAÇÃO E PROPOSIÇÕES

A ambiguidade é o emprego de palavras, termos ou expressões que permitem mais de uma interpretação. No contexto linguístico, a ambiguidade pode ser estrutural (ou gramatical) e lexical (ou polissêmica). Na primeira, os problemas são relativos à construção de sentenças, tais como a posição de complementos, adjuntos ou redução de orações (GOMES, 2009). Na segunda, ocorre a identidade de duas lexias que são “formas perfeitamente iguais que se distinguem semanticamente (um significante para dois significados, no plano do conteúdo)” (ZAVAGLIA, 2003a, p. 83).

A ambiguidade estrutural pode ocorrer na elaboração de textos para as definições de termos que representam conceitos em tesouros e ontologias, pois os enunciados não podem carregar diferentes

possibilidades discursivas. Nos tesouros, grande parcela da comunicação da estrutura do domínio está expressa nas definições e nas notas de escopo, que estabelecem uma finalidade e intenção única (clara, precisa, coerente e unívoca) para cada conceito. Nesse sentido, elaborar esses enunciados exige refletir a intenção de comunicação pretendida no escopo do instrumento criado.

A ambiguidade lexical pode acontecer de diferentes maneiras, porém, neste artigo, serão tratados aqueles termos que mais interessam na construção de tesouros: sinonímia, polissemia e homonímia. A sinonímia ocorre quando há um ou mais termos (palavras ou expressões) com sentido igual ou semelhante. Nos tesouros, esse fenômeno é controlado ao se estabelecer uma relação de equivalência entre um descritor preferido e outro(s) descritores não preferido(s). Essas relações informam que somente um descritor é autorizado para uso, indicando: origem linguística diferente: poliglota e multilíngue; nomes populares e nomes científicos: pinguins e enfenisciformes; substantivos comuns e nomes comerciais: fotocópia e xerox; variantes para conceitos novos: deletar e remover; termos reduzidos e termos antiquados: rádio e radiocomunicação; grafias variantes: Rumênia e Romênia; variações linguísticas: termos originais de culturas diferentes – mandioca e aipim; abreviaturas e nomes por extenso: PVC e polivinil clorido; forma fatorada e não fatorada de um termo: carvão + minério e minério de carvão.

A ambiguidade da polissemia é um fenômeno natural das línguas em geral e ocorre pela associação de diferentes sentidos a um único termo (palavra, expressão ou imagem), e pode ser considerada contrária à sinonímia. Ela pode ocorrer a partir de quatro formas básicas: (1) mudança de aplicação: mesma palavra difere conforme o contexto – indexação (economia), indexação (biblioteconomia); (2) especialização: mesma palavra pode ter diferentes sentidos especializados, mas apenas um deles será aplicado – análise de assunto, análise conceitual; (3) linguagem figurada: uma palavra usada em sentidos figurados, sem

perder o significado original – uso de metáfora e metonímia; (4) influência estrangeira: empréstimo entre idiomas diferentes.

Por fim, a ambiguidade da homonímia pode acontecer de (1) forma gráfica, quando termos (palavras ou expressões) possuem fonias iguais e sentidos diferentes, tal como em, por exemplo, acento (sinal gráfico) e assento (cadeira); e de (2) forma fônica, quando possuem identidade igual (grafia) e sentidos diferentes em quatro circunstâncias (ZAVAGLIA, 2003b): (a) homonímia semântica, mesma classe gramatical, com sentidos diferentes – cabo (militar) e cabo (da enxada); (b) homonímia categorial, quando possuem sentido e classes gramaticais distintas – conto (verbo) e conto (substantivo); (c) homonímia etimológica: étimo (morfema base para a formação de uma palavra, por derivação ou composição) e sentidos diferentes – manga (fruto: do idioma malaiala) e manga (do vestuário: do latim *manica*); (d) homonímia heterófona: quando têm sentido e fônica distintas – acerto (substantivo) e acerto (verbo).

Na organização conceitual da estrutura de tesouros e de ontologias, os termos estão correlacionados por tipos diferentes de relações. Segundo Zavaglia (2003a, p. 84), “o levantamento e a identificação das relações léxico-semânticas” são igualmente necessários no processamento da linguagem natural (PNL), porque são “recursos linguísticos para implementação computacional” (idem), sendo importante tratar o fenômeno da homonímia para a interpretação da máquina. Assim, as redes de significação podem ser estabelecidas por proposições: uma sentença declarativa sobre algo, que pode ser declarado e representado por termos, palavras ou símbolos, que permite determinar se o seu conteúdo possui um valor lógico verdadeiro ou falso (GOMES, 2009).

Na filosofia, a proposição é uma entidade abstrata que exprime, em uma frase declarativa, aquilo que tem valor (validade) de verdadeiro ou falso (MURCHO, 2003). De acordo com o autor, o conjunto de proposições (premissas; primitivas) utilizadas para provar uma coisa é denominado

“argumento”. Almeida (2014) afirma que fundamentos filosóficos vêm sendo utilizados para dar condição ontológica de proposição para o conteúdo de um documento. O autor argumenta que, para isso ocorrer, uma análise deveria seguir “o seguinte princípio: ‘a informação transportada pela sentença S é uma proposição apropriadamente associada a S’ (FOX, 1983, p. 84)” (ALMEIDA, 2014, p. 250). Nesse sentido, a informação indexada seria o conteúdo proposicional do documento.

De modo geral, entretanto, sendo a linguagem o meio do homem se expressar sobre as verdades, a ambiguidade está sempre presente e, com ela, a inconsistência. Assim, a construção de tesouros e de ontologias busca por sua eliminação. Nesses instrumentos, as relações semânticas são representadas por proposições verbais (conexão entre termos e conceitos), determinando fatos, ações, lugares, períodos temporais, modos, causas, efeitos, destinos, procedências, circunstâncias, qualidades do objeto, entre outros, que explicitam o sentido da ligação. Portanto, conhecer o comportamento dos verbos é importante na construção de tesouros e ontologias.

TESAUROS E SUAS RELAÇÕES SEMÂNTICAS

Um tesouro é um SOC constituído por um conjunto de termos descritores, preferidos e não preferidos, que formam um sistema de conceitos inter-relacionados, sendo desenvolvidos com base em conceitos (ISO 25964-1, 2011). Tem como função fazer o controle do vocabulário de um dado domínio de especialidade, em diferentes níveis de controle, sendo utilizado no processo de indexação, na etapa de tradução da linguagem natural utilizada nos documentos para a linguagem adotada pelo sistema (UNESCO, 1971). Pode também ter a função de recuperação da informação, pois, quando acoplado na saída de dados, permite a conversão da expressão de busca, em linguagem natural, para a linguagem controlada que foi utilizada na entrada de dados no sistema (CAVALCANTI, 1978). Com isso, é ainda possível ampliar (termos mais gerais)

ou restringir (termos mais específicos) a busca, tornando a recuperação mais relevante à demanda do usuário.

Um tesouro é constituído por quatro elementos básicos: (1) um léxico; (2) uma estrutura gramatical, que representa o padrão de relações existentes entre conceitos e termos; (3) uma rede paradigmática (relações essenciais e estáveis); (4) uma rede sintagmática (relações válidas no contexto de uso) (SVENONIUS, 2000). A construção dessa estrutura leva em consideração as bases da Teoria do Conceito e da Terminologia para o levantamento do vocabulário (conceitos e termos que os designam) e para a elaboração da definição, com o registro das relações semânticas. Também considera a Teoria da Classificação Facetada (TCF), que respalda a análise analítico-sintética dos assuntos e a formação de classes gerais, subclasses e facetas, e regras de validação das escolhas realizadas.

Dentre as normas e diretrizes para a construção de tesouros, destaca-se a norma internacional mais atual, a ISO 25964-1 (2011), com as linhas gerais para o estabelecimento de relações entre termos e conceitos, assim como as formas de desambiguação que são recorrentemente discutidas na literatura. O sistema conceitual de um tesouro abarca relações de equivalência, hierárquicas e associativas, conectadas por relações binárias, que podem ter propriedades tais como reflexivas e irreflexivas, simétricas e assimétricas ou transitivas e intransitivas.

As relações de equivalência ocorrem no nível do campo semântico do termo, pois, também na linguagem de especialidade, diferentes termos podem representar conceitos idênticos, semelhantes ou com sentidos sobrepostos. Dodebei (2002, p. 91) afirma que a relação de equivalência ocorre quando “um conceito pode ser representado por vários símbolos distintos ou quando se quer reduzir, por questões pragmáticas, os níveis de implicação conceitual”, sendo uma relação recíproca. De forma resumida, as equivalências abarcam os tipos: (1) sinonímia (total e parcial); (2) correspondência em outro idioma; (3) variações ortográficas; (4) variações linguísticas; (5) erro de soletração; (6) forma completa do termo; (7)

abreviatura; (8) siglas; (9) forma singular e plural e (10) nome popular e nome científico (MACULAN, 2015; ISO 25964-1, 2011).

Quanto às relações hierárquicas, ocorrem no nível conceitual e são de três tipos: (1) gênero-espécie: incluem relacionamentos reflexivos, transitivos, assimétricos e taxonômicos (semelhanças entre táxons), que conectam uma classe (gênero: grupo de seres ou objetos não individualizados) e seus membros (espécie: ser individualizado ou objeto específico). Nessas relações, incidem os fenômenos de hiperonímia (hiperônimo; superordenação), e de hiponímia (hipônimo; subordinação), que indicam uma inclusão semântica de maior especificidade (FERREIRA, 2009); (2) todo-parte: incluem relacionamentos transitivos e assimétricos, nos quais o todo é um termo mais abrangente (superordenado) que abarca outro conceito, de forma lógica e inerente, que é uma parte em posição subordinada, que, em geral, pode incidir em objetos (segmentação) ou na divisão de uma área do conhecimento (DAHLBERG, 1979). A norma ISO 25964-1 (2011) determina um número limitado de situações para a representação dessa relação: (a) sistemas e organismos do corpo; (b) localização geográfica; (c) disciplinas e campos de discurso; (d) estruturas sociais. A diretriz estabelece que outros tipos de relação todo-parte, ainda que possíveis, não devem ser incluídos nos tesouros, sobretudo quando a tal parte puder ser, também, parte de outro todo; e (3) de instância: incluem classes básicas (mais gerais) de elementos, tais como objetos, eventos, propriedades ou fenômenos, que representam um conjunto desses elementos, que, em geral, é representado por um substantivo comum, e as instâncias, por nomes próprios. Por exemplo, “Regiões Montanhosas” (superordenado: substantivo) e suas instâncias, “Alpes” e “Himalaias” (subordinado: nomes próprios) (ISO 25964-1, 2011).

Também as poli-hierarquias podem ocorrer em tesouros em circunstâncias tais como em relações: (a) de gênero-espécie: instrumento musical, órgão, que é subordinado dos superordenados instrumentos de teclado e de vento; (b) de todo-parte: bioquímica, que é subordinado aos superordenados biologia e química; (c) mistas: crânio, que é subordinado

aos superordenados osso (gênero-espécie) e cabeça (todo-parte).

As relações associativas, ou não hierárquicas, acontecem no nível conceitual e são as ligações entre conceitos de uma mesma categoria ou não, que implica um elo com propriedade simétrica. Soergel (1974) observa que, ainda que não seja a regra, as relações associativas podem ser assimétricas, unidirecionais, pois nem sempre há a necessidade de se indicar reciprocidade. Em especial, isso ocorre em tesouros multilíngues, pois em comunidades com diferentes culturas e/ou idiomas que compartilham um mesmo tesouro, nem sempre um conceito é adequado para todas as comunidades assistidas (ISO 25964-1, 2011). Svenonius (2000) afirma que, se na definição de um conceito está implícito outro conceito não ligado por relações hierárquicas, então haverá uma relação associativa entre eles; por exemplo, se para definir o conceito “ave” emprega-se o conceito “ornitologia”, é apropriado estabelecer uma relação associativa entre eles. Também é recomendado criar relações associativas entre conceitos com significados sobrepostos dentro do escopo coberto por um tesouro, visando orientar o usuário; por exemplo, entre os conceitos “navio” e “barco” que podem representar conceitos diferentes em dado contexto. Ressalta-se

que, se os dois conceitos estiverem em coordenação em uma mesma classe, a relação associativa não deverá ser estabelecida.

O quadro 1 apresenta situações nas quais a relação associativa deve ser criada, porém esse conjunto de situações não necessariamente limita-se a esses exemplos, sendo preciso mapear o domínio para determinar o conjunto necessário e suficiente ao escopo (ISO 25964-1, 2011).

Nos tesouros, as diferentes relações (equivalência, hierárquicas e associativas), em geral, são identificadas na sua estrutura pelos símbolos apresentados no quadro 2.

Como pode ser observado, apesar de todo o trabalho intelectual para o mapeamento e estabelecimento das diferentes relações (equivalência, hierárquicas e associativas) que ocorrem entre termos e conceitos, os símbolos utilizados na sua representação ainda carecem de um refinamento mais específico, ainda que haja a recomendação para que esse aprimoramento possa ser realizado. Um trabalho sobre refinamento de relações em tesouros é apresentado na seção do estudo de caso.

Quadro 1– Situações para estabelecimento de relações associativas em tesouros

SITUAÇÃO	EXEMPLO
1) Uma disciplina ou campo de estudo e os objetivos ou fenômenos estudados	Estética e Beleza
2) Uma operação ou processo e seu agente ou instrumento	Controle de temperatura e Termostato
3) Uma ação e o resultado da ação	Tecelagem e Tecido
4) Uma ação e o seu recipiente ou alvo	Colheita e Produtos agrícolas
5) Objetos e materiais e suas propriedades	Venenos e Toxicidade
6) Artefatos e suas partes (se não se qualifica na relação hierárquica todo/parte)	Instrumento ótico e Lentes
7) Conceitos ligados por uma dependência causal	Duelo e Morte
8) Um objeto ou processo e seu contra-agente	Plantas e Herbicidas
9) Um conceito e sua unidade de medida	Corrente elétrica e Amperes
10) Um conceito representado por um termo composto e o conceito representado pelo substantivo foco	Barcos para amarrar e Barcos
11) Um organismo ou substância derivada ou produzida de outra	Latão e Cobre

Fonte: Elaborado pelas autoras, com base na norma ISO 25964-1 (2011).

Quadro 2 – Descrição dos símbolos utilizados em tesouros

SÍMBOLO DAS RELAÇÕES	DESCRIÇÃO
Equivalência (sem especificação) USE; UF / USE; UP	USE, apontando o descritor preferencial UF (<i>used for</i> ; UP: usado para), apontando o descritor não preferencial
Equivalência (com especificação) USE; UF / USE; UP FT; AB / TC; AB *Outros	USE, apontando o descritor preferencial; e UF (<i>used for</i> ; UP: usado para), apontando o descritor não preferencial FT (<i>full term</i> ; TC: termo completo), apontando o descritor do nome completo; e AB (<i>abbreviation</i> ; abreviatura), apontando o descritor do nome abreviado ou sigla *Recomendação para subdividir as relações de equivalência para refinamento da semântica das ligações estabelecidas entre conceitos
Hierárquica (sem especificação) BT; TG NT; TE	BT (<i>broader term</i> ; TG: termo geral), apontando o descritor da classe mais geral NT (<i>narrower term</i> ; TE: termo específico), apontando o descritor mais específico
Hierárquica (com especificação) BTG; NTG / TGG; TEG BTP; NTG / TGP; TEP BTI; NTI / TGI; TEI	BTG (<i>broader term (generic)</i>); TGG: termo geral de gênero); BTP (<i>broader term (partitive)</i>); TGP: termo geral partitivo); BTI (<i>broader term (instantial)</i>); TGI: termo geral de instância), apontando o descritor da classe mais geral NTG (<i>narrower term (generic)</i>); TEG: termo específico de gênero); NTP (<i>narrower term (partitive)</i>); TEP: termo específico partitivo); NTI (<i>narrower term (instantial)</i>); TEI: termo específico de instância), apontando o descritor mais específico
Associativa RT; TR; TA	RT (<i>related term</i> ; TR: termo relacionado; TA: termo associado), apontando para o descritor sugerido para ser levado em consideração
Associativa (com especificação) RT; TR; TA *CAUSE/EFFECT; CAUSA/ EFEITO	RT (<i>related term</i> ; TR: termo relacionado; TA: termo associado), apontando para o descritor sugerido para ser levado em consideração *Recomendação para subdividir as relações específicas para refinamento da semântica das ligações estabelecidas entre conceitos

Fonte: Elaborado pelas autoras, com base na norma ISO 25964-1 (2011).

ONTOLOGIAS E AS RELAÇÕES SEMÂNTICAS

No papel de um SOC e no âmbito da CI, as ontologias são criadas a partir de representações de conhecimento para serem interpretáveis pela máquina como artefatos capazes de proporcionar inferências a partir das relações atribuídas entre seus diferentes elementos. É um modelo conceitual e base de conhecimento, podendo representar relacionamentos complexos entre objetos, por meio de uma modelagem formal de dados. Diferentes autores têm investigado sobre ontologias, tanto em relação a questões teóricas quanto a aplicações (GRUBER, 1992; GUARINO, 1995; 1998; VICKERY, 1997; SOERGEL, 1997; BORST, 1997; NOY; HAFNER, 1997; WAND; STOREY; WEBER, 1999; HOVY, 2002; GILCHRIST, 2003; ALMEIDA, 2006; 2014). Na CI, Vickery (1997) define ontologia como

um esquema (ou sequência formal de eventos), que, de alguma forma, pode envolver o uso de categorias semânticas, de conceitos significantes em um domínio particular (que pode ser amplo, como o universo de conhecimento), juntamente com uma definição ou uma nota de escopo para cada conceito, assim como mecanismos para a exibição das relações entre conceitos (VICKERY, 1997, p. 284).

O autor aponta, na definição, uma analogia entre os tesouros e as ontologias, embora deixe claro que há diferenças, em especial, nos propósitos e funções previstas para cada instrumento. De modo geral, uma ontologia é formada por: (a) classes ou conceitos: uma taxonomia; (b) relações: representadas pela interação entre termos e classes; (c) atributos e propriedades (*slots*): descritas pelas características dos conceitos e instâncias; (d) instâncias ou indivíduos: elementos que pertencem a uma classe e que possuem propriedades; (e) axiomas: representam sentenças verdadeiras; (f) regras para permitir inferências (NOY; MCGUINNESS, 2001).

Desde a concepção filosófica, o objeto material da ontologia é a formalidade (a forma) do “Ser”, que determina o “que é” (*quid*), e, na CI, a busca pela formalidade continua sendo priorizada (BORCHERT, 2006), mas a partir de níveis diferenciados (ALMEIDA; BAX 2003): (a) altamente informais: em linguagem natural; (b) semi-informais: em linguagem natural, de forma restrita; (c) semiformais: em linguagem artificial e definidas de maneira formal; (d) rigorosamente formais: com semântica formal, teoremas e provas. Apesar desses distintos níveis de formalidade, uma ontologia controla a ambiguidade, sinonímia, hierarquias, associações, com relacionamentos complexos, regras e axiomas.

Almeida e Bax (2003) destacam que não há consenso sobre a classificação das tipologias de ontologias. Dentre os tipos discutidos na CI, destacam-se (GUARINO, 1998; ALMEIDA; BAX, 2003): (1) ontologias de representação; (2) ontologias genéricas; (3) ontologias genéricas de domínio (*core ontologies*); (4) ontologias de domínio; (5) ontologia de aplicação; (6) ontologias de tarefas. A construção de ontologias é desafiante devido ao elevado consumo de tempo requerido em seu desenvolvimento e em sua implantação, em especial quando exige a integração de diferentes aplicações (SCHIESSL; BRÄSCHER, 2011). Os autores asseguram que ela é uma aliada em ambiente Web, pois “possibilita o trabalho cooperativo entre computadores e pessoas” (idem, p. 304). Entretanto, as conclusões de um estudo realizado por Silva, Souza e Almeida (2008, p. 74) demonstraram problemas com a “falta de um padrão para construção de ontologias e na falta de explicações sistemáticas de como, onde e sob quais limites podem ser utilizadas as abordagens teóricas dentro do processo de elaboração”. Esse fato tem tornado o desenvolvimento de ontologias, nas discussões científicas promovidas na área da CI, suscetível a algumas críticas (AGANETTE, 2015).

Quanto às relações semânticas nas ontologias, elas são estabelecidas por meio de uma organização taxonômica de conceitos, com hierarquias de generalidade e de especialidade que compõem o

conceito de heranças entre as distintas entidades (CARVALHEIRA, 2007). Pode-se afirmar que as relações estabelecidas em sua estrutura acontecem tal qual nos tesouros, já especificadas na seção anterior, porém, com maior complexidade de representação e condicionadas ao tipo e ao propósito do instrumento. Nas ontologias aplicadas de orientação aristotélicas, por exemplo, na *Basic Formal Ontology* (BFO), o instrumento trata somente de universais e particulares, com um conjunto limitado de relações ontológicas, que estão especificadas na *Relation Ontology* (RO) (ALMEIDA; MENDONÇA; AGANETTE, 2013), que visa permitir a interoperabilidade entre diferentes aplicações na biomedicina.

De modo geral, as linguagens para construção de ontologias proporcionam primitivas (regras de formação) para a determinação de classes, subclasses e decomposições: disjuntiva, quando “os conceitos são subclasses de um conceito comum e onde nenhum membro de uma subclasse é membro de outra classe”; exaustiva de subclasse, quando “todos os membros da classe ou conceito comum são membros de alguma de suas subclasses”; e de negação (não é uma subclasse de), quando “um conceito não é uma especialização de um outro conceito (CARVALHEIRA, 2007, p. 15-16). Assim, as relações entre classes e indivíduos podem ser especificadas a partir de combinações lógicas de interseção, união e complemento.

Dentre as linguagens que vêm sendo utilizadas para a implementação de ontologias, destacam-se: *Web Ontology Language* (OWL; Lite, DL, Full), *Knowledge Interchange Format* (KIF), *Resource Description Framework* (RDF e RDFS), *Ontology Inference Layer* (OIL), *DARPA Agent Markup Language + OIL* (DAML+OIL). Em geral, a linguagem utilizada é a OWL (desenvolvida para Web); é considerada de maior expressividade, capaz de declarar fatos, estruturar classes (hierarquias de subsunções; tipologia e instanciações) e propriedades (hierarquias de subpropriedades) em RDF e RDFS, além de possibilitar declarações de equivalência para classes e propriedades, que podem ser determinadas iguais ou desiguais, a depender da intencionalidade delas.

Por meio da OWL, é possível estabelecer as seguintes declarações taxonômicas (classe: generalidade e subclasse: especialidade), de instanciação (instâncias), de propriedade (tipo de dado; objeto; hierarquia de propriedade) e de relações.

A ISO 25964-2 (2013) é a primeira norma internacional que traz regras específicas para a construção de ontologias e estabelecimento de relações. As normas se restringem às ontologias formais de domínio, adotando a terminologia da linguagem OWL, mas determinam que, em geral, as ontologias devem ser expressas em lógica de primeira ordem, e que podem ser criadas com distintos níveis de especificidade, propósito e para diversos domínios.

Para o estabelecimento de classes, a norma recomenda que sejam feitas declarações explícitas de restrições de propriedade e critérios para que um indivíduo possa ser membro de uma classe. Ademais, pode-se definir as relações de uma classe com outras classes, usando conectores lógicos (proposições de relação) ou qualquer outro tipo de restrição necessária.

Por exemplo, “uma classe pode ser definida como sendo uma interseção entre duas classes (denotando todos os indivíduos que pertencem às duas classes) ou como complemento de outra classe (denotando os indivíduos que não pertencem àquela classe)” (ISO 25964-2, 2013, p. 73).

As hierarquias de classe são criadas por estabelecimento de subclasses que vão obedecer aos axiomas declarados para cada classe, pois as subclasses (filho) irão herdar (princípio da herança) as características e propriedades (atributos; relações entre membros da classe e com outras classes) da classe (pai). Todas as relações entre as classes e subclasses são transitivas e as propriedades podem possuir restrições declarativas. Os indivíduos (instâncias) são os objetos de um domínio em particular, sobre os quais são feitas declarações, tais como: é uma pessoa, é um livro, tem determinado sintoma, é um processo ou evento, entre outros. Ainda que as ontologias, em si, não necessitem do uso de linguagem natural, todas as classes, subclasses, propriedades e indivíduos são identificados por termos para que possam ser interpretáveis por humanos (ISO 25964-2, 2013).

Quadro 3 – Declarações utilizando linguagem OWL

Item	Declarações
País; Pessoa	Classe (generalidade)
Estudante	Subclasse de Pessoa (especialidade)
Brasil e Portugal	Instância da classe País
Brasil e Portugal	Disjunção (decomposição disjuntiva)
Nacionalidade	Propriedade de Pessoa (domínio) e País (imagem ou <i>range</i>)
Nacionalidade	Propriedade inversa igual a <tem_cidadão>
Idade	Propriedade funcional
Idade	Propriedade de tipo de dado onde Pessoa é o domínio e o range é um número inteiro
João	Instância da classe Brasileiro e que tem na propriedade Idade um valor 35
Brasileiro	Conjunto de membros da classe Pessoa
Multinacional	Conjunto de membros da classe Pessoa, que possui pelo menos dois valores da propriedade Nacionalidade

Fonte: Elaborado pelas autoras com base em Carvalheira (2007, p. 20).

REFINAMENTO DE RELAÇÕES EM TESAuros: UM ESTUDO DE CASO

Este estudo de caso apresenta resultados da pesquisa de Maculan (2015), que aplicou a proposta de reengenharia de tesauros (SOERGEL et al., 2004; LAUSER et al., 2006), cujos procedimentos visam ao refinamento semântico das relações na estrutura conceitual de um tesouro a partir de três etapas: (1) definição da estrutura do tesouro; (2) coleta de terminologia; (3) aplicação e edição da reengenharia no tesouro. Os procedimentos (1) e (2) seguiram as diretrizes de construção de tesauros da norma ISO 25964-1 (2011), que estão descritas em Maculan (2015). Neste artigo, para atender ao escopo, será apresentado o procedimento (3) do refinamento das relações.

A reformulação foi realizada no tesouro Thesagro, do domínio da Agropecuária, no recorte temático da Intensificação Agropecuária (Intagro), a partir da concepção teórica de Boserup (1965). Originalmente, a estrutura conceitual do Thesagro possui as relações de equivalência, hierárquicas e associativas, e a representação delas é feita pelos símbolos na língua inglesa: *USE* e *Used For* (UF), *Broader Term* (BT) e *Narrower Term* (NT), e *Related Term* (RT). O refinamento semântico das relações levou em consideração os fundamentos da Teoria do Conceito, Terminologia e Teoria da Classificação Facetada, contando com a validação de especialista do domínio a cada estágio do trabalho, seguindo três fases:

(1) Estabelecimento das classes básicas e conjunto de conceitos: o domínio temático foi modelado de forma híbrida, com os métodos (1) dedutivo: análise macro para obter uma visão geral do domínio temático, que resultou em um conjunto de nove classes básicas que representam o campo semântico – agricultura extensiva, agricultura intensiva, material e métodos, ambiente, agronomia, território e paisagem, socioeconômica, espaço e tempo, instituições; (2) indutivo: obter diversas visões micro (particulares) do domínio temático, a partir de dados coletados (*corpus*) na literatura do domínio Agropecuário, nacional e internacional.

Dessa análise, resultaram cerca de 600 termos para serem estruturados dentro das nove classes básicas, tendo sido feito um recorte amostral composto de 30 conceitos representativos das classes básicas, e, assim, do domínio temático.

(2) Estabelecimento da estrutura conceitual: foram modelados os grupos (*clusters*) semânticos da amostra, com o estabelecimento de relações semânticas (equivalência, hierárquica e associativa), realizado a partir das definições (delimitação de significado) para cada conceito. As definições foram elaboradas tendo como insumos os excertos definitórios recolhidos na literatura (glossários, dicionários, bases de dados da agropecuária, textos técnicos e científicos). Como resultado, foram obtidos cerca de 600 relacionamentos.

(3) Refinamento dos relacionamentos: a explicitação das relações foi realizada utilizando-se 44 diferentes tipos, incluindo os seus inversos (por exemplo, <influencia> e a sua inversa <é_influenciado_por>), sendo criadas 13 novas relações e 31 relações provenientes do *Agrontology*¹, do tesouro Agrovoc (*Multilingual Agricultural Thesaurus*).

Também para auxiliar na construção do sistema conceitual do tesouro, foi aplicada a estrutura dos papéis *Qualia*, desenvolvida por Pustejovsky (1995), para a identificação da semântica dos verbos, pois toda relação é explicitada a partir de uma expressão verbal (proposição). A estrutura *Qualia* possui os papéis: (a) Formal; (b) Constitutivo; (c) Tético; (d) Agentivo. Essa identificação deu sentido aos verbos das 44 relações e facilitou a determinação das equivalências, dos gênero-espécies (hiperonímia/hiponímia) e dos todo-partes (holonímia/meronímia).

A identificação dos papéis *Qualia* auxiliou a associação de propriedades às relações, facilitando o refinamento semântico das expressões verbais utilizadas nas relações, permitindo apreender os seus significados específicos, sobretudo das relações hierárquicas e associativas.

¹Disponível em: <<http://aims.fao.org/sites/default/files/uploads/file/aos/agrontology/index.htm>>.

Quadro 4 – Conjunto relações com refinamento semântico

NOME DA RELAÇÃO	PAPEL QUALIA	TIPO RELAÇÃO
1. afeta (TR)	AGENTIVO	ASSOCIATIVA
2. causa (TR)	AGENTIVO	ASSOCIATIVA
3. éFeitoAPartirDe (TR)	AGENTIVO	ASSOCIATIVA
4. influencia (TR)	AGENTIVO	ASSOCIATIVA
5. seDesenvolveEm (TR)	AGENTIVO	ASSOCIATIVA
6. éInsumoPara (TR)	CONSTITUTIVO	ASSOCIATIVA
7. inclui (TR)	CONSTITUTIVO	ASSOCIATIVA
8. fazUsoDe (TR)	TÉLICO	ASSOCIATIVA
9. temPrática (TR)	TÉLICO	ASSOCIATIVA
10. temSinonímiaAproximada (UP EQ-APR)	FORMAL	EQUIVALÊNCIA
11. temSinônimo (UP EQ-SIN)	FORMAL	EQUIVALÊNCIA
12. temTraduçãoEN	FORMAL	EQUIVALÊNCIA
13. éProdutoDe (TGG)	AGENTIVO	GENÉRICA
14. termoEspecificoGênero (TEG)	FORMAL	GENÉRICA
15. termoGenéricoGênero (TGG)	FORMAL	GENÉRICA
16. termoEspecificoInstância (TEI)	CONSTITUTIVO	INSTÂNCIA
17. éComponenteDe (TGP)	CONSTITUTIVO	PARTITIVA
18. éParteDe (TGP)	CONSTITUTIVO	PARTITIVA
19. temComponente (TEP)	CONSTITUTIVO	PARTITIVA
20. temParte (TEP)	CONSTITUTIVO	PARTITIVA

Fonte: Maculan (2015, p. 222-223).

O REÚSO DE RELAÇÕES DE TESAUROS NAS ONTOLOGIAS: ANÁLISES

Primeiramente, destaca-se uma diferença básica entre tesouros e ontologias. Um tesouro é uma linguagem documentária e, mais especificamente, uma linguagem de indexação, que tem como primordial função ser instrumento de tradução da linguagem natural utilizada nos documentos para o vocabulário controlado adotado por um sistema de recuperação da informação (SRI). Já as ontologias são criadas como um modelo conceitual de mundo, alinhadas a sistemas formais de representação, a partir de diferentes acepções. Nesse sentido, haverá ontologias de alto nível, nível médio e de domínio, por exemplo. Segundo Almeida, Mendonça e Aganette (2013), seria possível uma ontologia ser utilizada como uma linguagem documentária, no sentido de restringir a linguagem natural, “caso esteja referindo-se ao uso de linguagens de representação adotadas para especificar ontologias [...] implementada[s] em

um sistema de recuperação da informação baseado na Web” (idem, p. 13). Porém, esse não é o papel típico de uma ontologia.

No mapeamento de um domínio, tanto para criação de tesouros quanto de ontologias, as classes de maior abstração são oriundas da linguagem natural. Como os tesouros almejam uma restrição do vocabulário utilizado em um domínio em particular, em que, pelo menos, as garantias da literatura e do usuário são levadas em consideração, o nível de representação ainda ficará bem próximo da linguagem natural. Por sua vez, nas ontologias, o estabelecimento das classes mais gerais (categorias) se inicia no nível da linguagem natural e é elevado ao nível da semântica formal (ALMEIDA; MENDONÇA; AGANETTE, 2013), no qual o sentido de referencialidade é fundamental, pois são declaradas asserções sobre objetos (indivíduos, fatos, eventos, propriedades, entre outros), que são externos à própria língua.

Assim, o “significado é entendido como uma relação entre a linguagem, por um lado, e, por outro, aquilo sobre o qual a linguagem fala” (MÖLLER; VIOTTI, 2003, p. 138). Dessa maneira, as ontologias exigem relações explícitas, em um contexto de linguagem não natural, como a lógica, para que sejam passíveis de interpretação pelas máquinas e não apenas dispostos à abstração pelo ser humano (AGANETTE, 2015).

Na perspectiva das considerações anteriores, este artigo se atém às discussões relativas às possibilidades de reúso das relações semânticas (equivalência, hierárquicas e associativas), estabelecidas na estrutura de um tesouro, para a construção de ontologias.

Como já mencionado, as relações de equivalência acontecem, na estrutura dos tesouros, entre um descritor preferido e um não preferido, pois manter o elo “permitirá que a busca por qualquer um deles recupere informações referentes a todos os outros que estão conectados entre si” (MACULAN, 2015, p. 146). Assim, ela é operacionalizada no nível da significância (campo semântico) do termo, que é o conjunto dos significados que o termo possui dentro de um contexto. Nas ontologias, Schulz et al. (2009) afirmam que a relação de equivalência acontece no nível da referencialidade, pois trata-se de uma representação descritiva formal, no nível referencial, e apresentam uma relação de equivalência entre dois conceitos e a explicam

Hepatite \equiv Doença_Inflamatória \sqcap \exists tem_localização.
Fígado

O operador de equivalência \equiv nesta fórmula nos diz que: (I) cada ocorrência específica de hepatite é uma ocorrência de doença inflamatória localizada em algum fígado, e também (ii) que todas as ocorrências de doença inflamatória localizadas em algum fígado são ocorrências de hepatite. Assim, em qualquer situação, o termo à esquerda pode ser substituído pela expressão à direita, sem qualquer perda de significado (SCHULZ et al., 2009, p. 37-38).

Como pode ser observado, essa relação não ocorre da mesma maneira que acontece nos tesouros. Na declaração formal de equivalência, não há qualquer conhecimento implícito e ela será sempre verdadeira em toda e qualquer situação.

De maneira semelhante, ocorre equivalência entre classes, propriedades e indivíduos

EquivalentClasses <ontA:Person ontB:Human>;
EquivalentObjectProperties <ontA:hasChild
ontB:child>; EquivalentDataProperties <ontA:hasAge
ontB:age> (ISOTANI; BITTENCOURT, 2014, s.p.).

Esse exemplo foi declarado usando linguagem OWL e os autores esclarecem que esse tipo de equivalência é feito, em geral, para a integração e o reúso de ontologias (ontA e ontB).

Nos tesouros, usualmente, as relações hierárquicas são representadas por *Broader Term* (BT) e *Narrower Term* (NT), incluindo as relações gênero-espécie, todo-parte e de instâncias, com semântica ainda considerada restrita. Porém, no trabalho de Maculan (2015), cada um dos tipos de relação foi explicitado, aumentando a expressividade semântica das relações, como descrito na seção do estudo de caso. Com isso, todas as relações hierárquicas, refinadas, poderiam ser reutilizadas na modelagem de uma ontologia no mesmo domínio, ainda que seus elementos e partes devam ser representados com maior complexidade. Cada entidade (classe, subclasse, propriedade, indivíduo) na ontologia é tratada de maneira isolada, situação que não ocorre na construção de tesouros em geral, mas aconteceu no trabalho de Maculan (2015). Cuidado também deve ser tomado para a eliminação de poli-hierarquias nas ontologias, pois não são permitidas, atendendo ao princípio da não herança dupla (KLESS et al., em 2012; SANAA et al., 2013). Ademais, ressalta-se o já mencionado sobre os rótulos das classes de maior abstração, assim como das suas subclasses, que devem ser levadas para o nível da semântica formal e não ser em linguagem natural.

As relações associativas, em geral, são representadas nos tesouros por *Related Term* (RT), podendo incluir uma série de diferentes tipos de associações entre conceitos, o que é considerado com semântica ainda mais restrita do que nas relações hierárquicas. Do ponto de vista da recuperação de informações, esse elemento do tesouro permite a expansão de buscas pelo usuário. Entretanto, como já mencionado, a recuperação de documentos não

é um papel típico das ontologias, ainda que haja, segundo Schulz et al. (2009), alguns trabalhos nesse sentido (MÖLLER et al., 2004; GIUNCHIGLIA; DUTTA; MALTESE, 2013). Como no trabalho de Maculan (2015) as relações associativas foram explicitadas, considera-se que elas podem ser insumos para o mapeamento de relações para a criação de ontologias, sobretudo as de domínio.

Schulz et al. (2009) alertam, entretanto, para o fato de que, nas ontologias formais, não seria possível, utilizando o mesmo exemplo da entidade hepatite ($\langle \text{Hepatite} \equiv \text{Doença_Inflamatória} \sqcap \exists \text{tem_localização.Fígado} \rangle$), apresentado anteriormente, afirmar que, algumas vezes, a hepatite vem acompanhada de um sintoma, a febre ($\langle \text{Inflamação} \sqcap \exists \text{tem_localização.Fígado} \sqcap \exists \text{normalmente_tem_sintoma.Febre} \rangle$). Para os autores, a máquina não conseguiria interpretar, “por exemplo, que a hepatite tem o sintoma febre na maioria dos casos (mas não em todos)” (idem, p. 38), pois, quando se atribui uma afirmativa para uma entidade, há de se admitir que o sintoma da febre irá ocorrer em todos os casos. Não podendo ser assim, as ontologias não podem utilizar esse tipo de associação, coisa que é possível nos tesouros. Portanto, ao reutilizar conhecimento estruturado dos tesouros na criação de ontologias, esse tipo de situação sempre deverá ser analisado, mesmo que as relações estejam explicitadas, conforme o trabalho de Maculan (2015). Assim, nas ontologias, é preciso verificar a existência de interpretações ambíguas, por meio de declarações definitórias e axiomas expressos em linguagem formal.

Outro ponto que se considera interessante na construção de hierarquias e relações associativas em ontologias, proposto por Maculan (2015), é a determinação da estrutura dos quatro papéis *Qualia*: Constitutivo, Formal, Téliico e Agentivo (PUSTEJOVSKY, 1995; ZAVAGLIA, 2003a; 2003b) para as diferentes entidades. Na construção de ontologias, os papéis *Qualia* podem ser utilizados para tratar as extensões do significado do termo (ou conceito) e das proposições verbais (relações semânticas), caracterizando-os (léxicos e verbos)

em um nível semântico, pois é interessante para a organização e representação de conhecimento e, também, para a interpretação desse significado (lexical e verbal) por sistemas computacionais.

Segundo Zavaglia (2003b, p. 2), a estrutura *Qualia* “é capaz de suprir o vocabulário básico para expressar aspectos diferentes do significado lexical (*wordmeaning*)”, apresentando o exemplo para a unidade lexical “pudim”:

Quadro 5 – Significados de uma unidade léxica a partir de papéis *Qualia*

Pudim	
<i>Qualia</i> =	CONST = ingredientes
	FORMAL = substância
	TELIC = comer
	AGENT = fazer

Fonte: Zavaglia (2003b, p. 3).

Pelo exemplo, notam-se as dimensões de significado que a unidade lexical “pudim” possui, expressas pelos quatro papéis *Qualia*, que se referem às relações envolvidas no contexto em que a unidade lexical é utilizada. A autora aponta que esse tipo de especificação de papéis foi utilizado para a desambiguação semântica entre unidades lexicais homônimas (ZAVAGLIA, 2003a), pois a caracterização das informações nos quatro tipos de papéis *Qualia* permite interpretar, também pela máquina, em qual contexto a unidade lexical está inserida.

Zavaglia (2003a; 2003b) também investigou sobre os tipos de relações semânticas a partir da determinação dos seus papéis dentro da estrutura *Qualia*, conforme exemplos no quadro 6.

É possível notar que a atribuição de papéis *Qualia* às proposições verbais (quadro 6) permite determinar o valor semântico das diferentes expressões verbais, auxiliando na desambiguação semântica das relações. Zavaglia (2003a) assegura que determinar os papéis *Qualia* às relações semânticas dá maior expressividade a elas, sendo possível estabelecer um padrão, principalmente, para relações de hiperonímia (todo) e hiponímia (partes), identificado pelo papel *Qualia* Constitutivo.

Quadro 6 – Relações Semânticas conforme a estrutura *Qualia*

PAPEL DA ESTRUTURA QUALIA	TIPOS DE RELAÇÕES SEMÂNTICAS
FORMAL	<é_um>; <é_um_sinônimo>; <é_um_antônimo>
CONSTITUTIVO	<é_um_membro_de>; <contém>; <quantifica>; <vive_em>; <atividade_constitutiva>; <está_em>; <tem_como_cor>; <tem_como_membro>; <feito_de>; <produzido_por>; <é_parte_de>; <propriedade_de>; <medido_por>
TÉLICO	<é_uma_atividade_de>; <objeto_da_atividade>; <é_a_habilidade_de>; <usado_para>; <usado_por>; <destinado_a>; <usado_contra>
AGENTIVO	<experiência_agentiva>; <resultado_de>; <origem>; <derivado_de>; <produzido_por>

Fonte: Zavaglia (2003b, p.3; 2003a, p.92-93).

CONSIDERAÇÕES FINAIS

Este estudo teve como proposta explorar a questão do reuso, na construção de ontologias, de conhecimento estruturado existente nos tesouros, em especial, das relações semânticas estabelecidas entre seus termos e conceitos. As análises das possíveis contribuições foram realizadas tendo como parâmetros a literatura sobre o tema, a norma ISO 25964, Parte 1 e 2 (2011; 2013) e os resultados apresentados por Maculan (2015). As apreciações realizadas neste artigo não são cabais, mas reflexões iniciais das autoras, e visam fornecer insumos para maior entendimento sobre as convergências e as divergências entre tesouros e ontologias, assunto de interesse nos estudos em representação do conhecimento na ciência da informação.

A partir das análises realizadas, considerou-se que as relações estabelecidas na estrutura de tesouros, quando explicitadas como no trabalho de Maculan (2015), em certo ponto podem contribuir com a determinação de relações de equivalência na construção de ontologias, desde que tudo que estiver relacionado com as entidades em equivalência puder ser, necessariamente, verdadeiro, em qualquer ocasião. Assim, percebeu-se que há similaridades na atribuição de relacionamentos no refinamento semântico descrito no trabalho de Maculan (2015) e na atribuição de relacionamentos na ontologia, porém, não uma correlação perfeita.

Ressalta-se que, como a ontologia estabelece hierarquias a partir da lógica descritiva, é possível determinar particularidades sobre uma propriedade

geral dos indivíduos e fazer generalizações que valem para um indivíduo. Isso é possível porque, na ontologia, o que ocorre são sentenças declarativas (verdadeira ou falsa) para o domínio, que sejam argumentos dedutivamente válidos (se as premissas são verdadeiras, a conclusão também é verdadeira). Porém, esse tipo de inferência, a partir da representação estabelecida na estrutura dos tesouros, não é possível que ocorra. Nos tesouros, as hierarquias podem ser construídas de maneira mais informal, podendo incluir, em uma mesma classe, por exemplo, diferentes indicadores de facetas (rótulo nodal), que não têm a mesma função de descritores preferidos, mas que compõem a parte sistemática do tesouro, indicando a base lógica do tipo de subdivisão que está sendo representado. Esse tipo de elemento pode dar pistas para a criação de ontologias, contudo, as ontologias não podem ser empregadas de forma indiscriminada.

Considerando as premissas de Guarino (1998), por exemplo, um dos componentes de uma ontologia são os conceitos (classes), representados por entidades concretas, reais ou fictícias, descrevendo funções, tarefas ou processos de raciocínio (GÓMEZ-PÉREZ, 1999). Para o autor, nessa abordagem (não realista) o compromisso primordial nem sempre está com a realidade, mas com a aplicação em si (necessidades e contexto). Assim, a contribuição da estrutura do tesouro para reuso na criação de ontologias também será dependente da abordagem adotada pelo modelador.

Verificou-se, ainda, que as diferenças nos níveis de granularidade representacional (subdivisão de entidades; relações hierárquicas) do conhecimento no tesouro, para ser reutilizadas na criação de ontologias, irão depender do tipo de aplicação, nível de formalidade e propósito, podendo exigir maior detalhamento e especificação. Nas ontologias, a granularidade não é um atributo de indivíduos, mas do próprio sistema, e se refere ao número de objetos representados nele.

Outro ponto que se considerou importante é a atribuição de papéis a partir da estrutura dos papéis *Qualia*, tanto na construção de tesouros quanto no desenvolvimento de ontologias. Se o tesouro for criado com refinamento das relações semânticas em sua estrutura, assim como com a atribuição de papéis semânticos, tal como ocorreu no trabalho de Zavaglia (2003a; 2003b) e de Maculan (2015), é esperado que as ontologias possam fazer, de maneira mais proveitosa, o reúso das proposições utilizadas nas relações semânticas que foram estabelecidas.

REFERÊNCIAS

AGANETTE, E.C. *Representação do conhecimento biomédico: uma investigação sobre a teoria da terminologia e a teoria da ontologia aplicada no domínio do sangue humano*. 2015.

238 f. Tese (Doutorado em Ciência da Informação)- Programa de Pós-Graduação em Ciência da Informação, Escola de Ciência da Informação, Belo Horizonte, ECI/UFMG, 2015.

ALMEIDA, M.B. *Um modelo baseado em ontologias para representação da memória organizacional*. 2006. 316 f. Tese (Doutorado em Ciência da Informação)- Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

_____. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. *Perspectivas em Ciência da Informação*, v. 19, n. 3, p. 242-258, jul./set. 2014.

_____. BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, v. 32, n. 3, p. 7-20, set./dez. 2003.

_____. MENDONÇA, F. M.; AGANETTE, E. C. Interfaces entre ontologias e conceitos seminais da Ciência da Informação: em busca de avanços na organização do conhecimento. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 29 de outubro a 1 de novembro de 2013, Florianópolis. *Anais...* Florianópolis: ENANCIB, 2013.

BOCCATO, V. R. C.; RAMALHO, R. A. S.; FUJITA, M. S. L. A contribuição dos tesouros na construção de ontologias como instrumento de organização e recuperação da informação em ambientes digitais. *Ibersid*, v. 2, p. 199-209, 2008.

BORCHERT, D. M. (Ed.). *Encyclopedia of Philosophy*. 2nd ed. New York: Thomson/Gale, Macmillan Reference, 2006. Disponível em: <<http://www.lambsound.com/Downloads/philosophy/Volume%2001.pdf>>. Acesso em: 22 jul. 2017.

BORST, W. N. *Construction of engineering ontologies for knowledge sharing and reuse*. 1997. 227 f. Tese (Doutorado)- Institute for Telematica and Information Technology, University of Twente, Enschede, The Netherlands, 1997.

BOSERUP, E. *The conditions of agricultural growth: the economics of agrarian change under population pressure*. London: Allen e Unwin, 1965.

CAMPOS, M. L. M. et al. O uso de tesouro como base terminológica para a elaboração de ontologias de domínio: uma experiência com o domínio do Folclore e Cultura Popular. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. *Anais...* São Paulo: USP, 2008.

CARVALHEIRA, L. C. C. *Método semi-automático de construção de ontologias parciais de domínio com base em textos*. 2007. 143 f. Dissertação (Mestrado em Engenharia de Sistemas Digitais)- Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia de Computação e Sistemas Digitais, São Paulo, USP, 2007.

- CAVALCANTI, C.R. *Indexação e tesouro: metodologia e técnicas*. Brasília: ABDE, 1978.
- DAHLBERG, I. Teoria da classificação, ontem e hoje. In: CONFERÊNCIA BRASILEIRA DE CLASSIFICAÇÃO BIBLIOGRÁFICA, 1972, Rio de Janeiro. *Anais...* Brasília: IBICT/ABDE, 1979. P. 352-370, V. 1.
- _____. Teoria do conceito. *Ciência da Informação*, v. 7, n. 2, p. 101-07, 1978.
- DODEBEI, V. L. D. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto, 2002.
- FERREIRA, A. B. H. *Novo dicionário Aurélio da Língua Portuguesa*. 4. ed. Curitiba: Positivo, 2009.
- FOX, C. J. *Information and misinformation: an investigation of the notions of information, misinformation, informing, and misinforming*. Westport: Greenwood, 1983.
- GILCHRIST, A. Thesauri, taxonomies and ontologies: an etymological note. *Journal of Documentation*, v, 59, n. 1, p. 7-18, 2003.
- GIUNCHIGLIA, F.; DUTTA, B.; MALTESE, V. *From knowledge organization to knowledge representation*. Trento, Italy: Università di Trento, 2013.
- GOMES, R. M. *Desambiguação de sentido de palavras dirigida por técnicas de agrupamento sob o enfoque da mineração de textos*. 2009. 118 f. Dissertação (Mestrado em Engenharia Elétrica)- Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2009.
- GÓMEZ-PÉREZ, A. G. Ontological engineering: a state of the art: expert update: knowledge based systems and applied Artificial Intelligence. *British Computer Society*, v. 3-2, p. 33-43, 1999.
- GRUBER, T. R. *What is an ontology?*. 1992. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 22 jul. 2013.
- GUARINO, N. Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human and Computer Studies*, v. 43, n. 5/6, p. 625-640, 1995.
- GUARINO, N. Formal ontology and information systems. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGIES IN INFORMATION SYSTEMS, 1., 1998, Trento, Italy, *Proceedings...* Trento, Italy: FOIS, IOS Press, 1998. P. 1-13.
- HOVY, E. Comparing sets of semantic relations in ontologies. In: GREEN, R.; BEAN, C. A.; MYAENG, S. H. (Ed.). *The semantics of relationships: an interdisciplinary perspective*. Norwell, MA, USA: Springer, 2002. Cap. 6, p. 91-110. V. 3.
- INTERNATIONAL STANDARD ORGANIZATION - ISO. *ISO 25964-1: thesauri and interoperability with other vocabularies. Part 1: thesauri for information retrieval*. Genève, 2011.
- _____. *ISO 25964-2: thesauri and interoperability with other vocabularies: part 2: interoperability with other vocabularies*. Genève, 2013.
- ISOTANI, S.; BITTENCOURT, I. I. *Dados abertos conectados*. São Paulo: Centro de Estudos sobre Tecnologia Web, 2014.
- KLESS, D. et al. A Method of re-engineering a thesaurus into an ontology. In: FORMAL ONTOLOGY IN INFORMATION SYSTEMS; INTERNATIONAL CONFERENCE – FOIS, 7., 2012, Amsterdam. *Proceedings...* Amsterdam: IOS Press, 2012. P. 133-146.
- LAUSER, B. et al. From Agrovoc to the agricultural ontology service: concept server an OWL model for creating ontologies in the agricultural domain. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2006, Colima, Mexico. *Proceedings...* México: DCMI, 2006.
- MACULAN, B. C. M. S. *Estudo e aplicação de metodologia para reengenharia de tesouro: remodelagem do Thesagro*. 2015. 345 f. Tese (Doutorado em Ciência da Informação)- Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, ECI/UFMG, 2015.
- MÖLLER, A. L. P.; VIOTTI, E. C. *Introdução à linguística II: princípios de análise*. São Paulo: Contexto, 2003.
- MÖLLER, H. et al. Textpresso: an ontology-based information retrieval and extraction system for biological literature. *Plos Biology*, v. 2, n. 11, p. 1984-1998, 2004.
- MURCHO, D. *O lugar da lógica na Filosofia*. Lisboa: Plátano Edições Técnicas, 2003.
- NOY, N. F.; HAFNER, C. D. The state of the art in ontology design: a survey and comparative review. *AI Magazine*, v. 36, n. 3, p. 53-74, 1997.
- _____. MCGUINNESS, D. L. *Ontology development 101: a guide to creating your first ontology*. Stanford, CA: Stanford University, 2000.
- ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA - UNESCO. *Guidelines for the establishment and development of monolingual thesauri for information retrieval*. Paris, 1971.
- PUSTEJOVSKY, J. *The generative lexicon*. Cambridge, MA: MIT Press, 1995.
- SANAA, M. et al. A methodological approach for converting thesaurus to domain ontology: application to tourism. *International Journal of Engineering and Innovative Technology (IJEIT)*, v. 3, n. 5, Nov. 2013.
- SCHIESSL, M.; BRÄSCHER, M. Do texto às ontologias: uma perspectiva para a ciência da informação. *Ciência da Informação*, v. 40, n. 2, p. 301-311, maio/ago. 2011.
- SCHULZ, S. et al. Strengths and limitations of formal ontologies in the biomedical domain. *Reciis: Electronic Journal of Communication, Information and Innovation in Health*, v. 3, n. 1, p. 31-45, 2009.

- SILVA, D. L.; SOUZA, R. R.; ALMEIDA, M. B. Ontologias e vocabulários controlados: comparação de metodologias para construção. *Ciência da Informação*, v. 37, n. 3, p. 60-75, set./dez. 2008.
- SOERGEL, D. *Indexing languages and thesauri: construction and maintenance*. Los Angeles, CA: Wiley, 1974.
- _____. *Functions of a thesaurus / classification / ontological knowledge base*. Maryland: College of Library and Information Services, University of Maryland, 1997.
- _____. et al. Reengineering thesauri for new applications: the Agrovoc example. *Journal of Digital Information*, v. 4, n. 4, 2004.
- SVENONIUS, E. *The intellectual foundations of information organization*. Cambridge: The MIT Press, 2000.
- VICKERY, B. C. Ontologies. *Journal of Information Science*, v. 23, n. 4, p. 277-286, 1997.
- WAND, Y.; STOREY, V. C.; WEBER, R. An ontological analysis of the relationship construct in conceptual modeling. *ACM Transactions*, v. 24, n. 4, p. 494-528, Dec. 1999.
- ZAVAGLIA, C. A homonímia no português: tratamento semântico segundo a estrutura *Qualia* de pustejovsky com vistas a implementações computacionais. *Alfa*, v. 47, n. 2, p. 77-99, 2003a.
- _____. Base de conhecimento léxico-ontológico para o Português do Brasil: uma proposta de modelo. In: WORKSHOP EM TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA, 1., 12 de outubro de 2003, São Carlos – SP. *Anais...* São Carlos/SP: USP, 2003b. P. 1-9.

Part II
Applications and uses of ontologies in information science

Parte I
Contribuições teóricas e metodológicas da pesquisa em ontologias

Parte I
Contribuciones teóricas y metodológicas de la investigación en ontologías

Construção de ontologia na prática: um estudo de caso aplicado ao domínio obstétrico

Fernanda Farinelli

Doutora em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, Minas Gerais - Brasil.

<http://lattes.cnpq.br/1907817850408525>

E-mail: fernanda.farinelli@gmail.com

Peter L. Elkin

Doutor em Medicina, New York Medical College (NYMC) - Valhalla, Nova York, Estados Unidos.

Professor e Coordenador do Departamento de Informática Biomédica da Universidade do Estado de Nova York (SUNY) Buffalo, Nova York, Estados Unidos.

<http://medicine.buffalo.edu/faculty/profile.html?ubit=elkinp>

<https://www.linkedin.com/in/peter-l-elkin-md-377640/>

E-mail: elkinp@buffalo.edu

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

RESUMO

O volume de informação e a variedade de fontes de informação geram desafios para a integração informacional. A necessidade de integração de informação entre sistemas de informação distintos alavancou pesquisas para identificar alternativas capazes de proporcionar interoperabilidade semântica entre sistemas, ou seja, a especificação da informação sem gerar ambiguidades. Como contribuição da ciência da informação, as ontologias servem como alternativa de padronização semântica das informações. Entretanto, o processo de construção de ontologias ainda gera muita dúvida entre pesquisadores. Muitos autores descrevem metodologias para a construção de ontologias, mas observa-se uma lacuna entre os métodos descritos e a sua aplicação na prática. Busca-se demonstrar na prática a construção de uma ontologia que adotou duas consolidadas metodologias: o realismo ontológico e a metodologia NeOn. Em relação aos métodos e procedimentos técnicos realizados, esta pesquisa é um estudo de caso que investiga a prática da construção de uma ontologia biomédica no domínio obstétrico, com o objetivo de descrever e explicar o processo de construção da ontologia praticado. Espera-se contribuir com o avanço da pesquisa em construção de ontologias no campo da ciência da informação, dada sua aplicação na solução problemas de organização e recuperação de informações em ambientes informacionais de diversos campos científicos.

Palavras-chave: Construção de ontologia. Metodologia para construção de ontologia. Ontologia biomédica. Ontologia formal. Ontologia de domínio.

Ontology building in practice: a case study applied to the obstetric domain

ABSTRACT

The volume of information and the variety of sources of information poses challenges for informational integration. The need to integrate information between distinct information systems leveraged research to identify alternatives capable of providing semantic interoperability between systems, that is, the specification of the information without creating ambiguities. As a contribution of Information Science, ontologies serve as an alternative for the semantic standardization of information. However, the process of constructing ontologies still raises a lot of doubt among researchers. Many authors describe methodologies for the construction of ontologies, but there is a gap between the methods described and their application in practice. It seeks to demonstrate in practice the construction of an ontology that has adopted two consolidated methodologies: the ontological realism and the NeOn methodology. In relation to the technical methods and procedures performed, this research is a case study that investigates the practice of the construction of a biomedical ontology in the obstetrical domain, with the purpose of describing and explaining the process of ontology construction practiced. It is hoped to contribute with the advancement of the research in construction of ontologies in the field of Information Science, given its application in the solution problems of organization and information retrieval in informational environments of diverse scientific fields.

Keywords: Ontology building. Methodology for ontology building. Biomedical ontology. Formal ontology. Domain ontology.

Construcción de la ontología en la práctica: un estudio de caso aplicado al dominio obstétrico

RESUMEN

El volumen de información y la variedad de fuentes de información generan desafíos para la integración informacional. La necesidad de integración de información entre sistemas de información distintos apalancó investigaciones para identificar alternativas capaces de proporcionar interoperabilidad semántica entre sistemas, es decir, la especificación de la información sin generar ambigüedades. Como contribución de la Ciencia de la Información, las ontologías sirven como alternativa de estandarización semántica de las informaciones. Sin embargo, el proceso de construcción de ontologías aún genera mucha duda entre investigadores. Muchos autores describen metodologías para la construcción de ontologías, pero se observa una laguna entre los métodos descritos y su aplicación en la práctica. Se busca demostrar en la práctica la construcción de una ontología que adoptó dos consolidadas metodologías: el realismo ontológico y la metodología NeOn. En cuanto a los métodos y procedimientos técnicos realizados, esta investigación es un estudio de caso que investiga la práctica de la construcción de una ontología biomédica en el dominio obstétrico, con el objetivo de describir y explicar el proceso de construcción de la ontología practicada. Se espera contribuir con el avance de la investigación en construcción de ontologías en el campo de la Ciencia de la Información, dada su aplicación en la solución de problemas de organización y recuperación de informaciones en ambientes informacionales de diversos campos científicos.

Palabras clave: *Construcción de ontología. Metodología para la construcción de ontología. Ontología biomédica. Ontología formal. Ontología de dominio.*

INTRODUÇÃO

As organizações, tanto no âmbito público quanto privado, anseiam por fornecer serviços de qualidade a indivíduos e entidades com quem se relacionam, e para isso, dependem do conhecimento envolvido nas suas práticas e atividades. Observa-se que a maneira como as organizações adquirem, compartilhem, criam, validam e publicam seus conhecimentos refletem diretamente sua competitividade e sua governança. Portanto, dispor de informações corretas no menor tempo possível torna-se um grande desafio em uma sociedade que exige das organizações decisões rápidas e adaptação às variações econômicas e questões legais.

Atualmente, o maior desafio das organizações é a necessidade de integração de informações com seus parceiros internos e externos. O desafio é ainda maior quando verificamos uma lacuna de interoperabilidade semântica entre os sistemas de informação adotados pelas organizações, principalmente em função da falta de padrões ou da não adoção de padrões quando existentes. Iniciativas em torno da obtenção de soluções capazes de promover a interoperabilidade entre os sistemas de informação é uma tendência mundial para as organizações.

Para superar a falta de interoperabilidade semântica, várias iniciativas adotaram soluções baseadas em uma ontologia formal devido à sua capacidade de representar o conhecimento sem ambiguidade. Vários autores reconhecem que as ontologias de domínio podem resolver problemas de interoperabilidade semântica entre a heterogeneidade terminológica (BITTNER; DONNELLY; WINTER, 2005; GANGEMI et al., 2002; SMITH, 2003b).

Em geral, o campo da ciência da informação lida com solução de problemas em torno da necessidade de informação e sua aplicabilidade à sociedade. Uma relevante contribuição da ciência da informação à sociedade são as teorias relacionadas à representação e organização do conhecimento. Assim, embora a pesquisa em ontologia se origine na filosofia, rapidamente, a ontologia assumiu seu lugar como

um tema de pesquisa na ciência da informação devido à sua capacidade de representação do conhecimento (ALMEIDA, 2013; GRUBER, 2008; SMITH, 2003B; SØERGEL, 1999; VICKERY, 1997).

Entretanto, no campo ciência da informação, apesar do reconhecimento das ontologias como ferramenta de representação do conhecimento, pouco se discute sobre como construir uma ontologia na prática para que se represente o conhecimento sem ambiguidade. Este estudo é uma iniciativa que busca enriquecer a discussão do papel social da ciência da informação, ajudando outros campos científico em referência a soluções relacionadas à representação e organização do conhecimento por meio das ontologias. Finalmente, o objetivo geral deste estudo é demonstrar na prática o processo de construção de uma ontologia de domínio que trata o conhecimento acerca dos registros de saúde eletrônicos gerados nos cuidados obstétricos e neonatais.

METODOLOGIA

Para atingir os objetivos de pesquisa científica, o pesquisador deve usar um conjunto de procedimentos intelectuais e técnicos conhecidos como métodos científicos para encontrar respostas para problemas de pesquisa (GIL, 2008; MARCONI; LAKATOS, 2003, 2007; YIN, 2009). Esta pesquisa busca compreender o processo de construção de ontologias na prática, e assim, é uma pesquisa aplicada qualitativa. Além disso, trata-se de uma pesquisa exploratória na qual o pesquisador busca obter maior familiaridade com o assunto da pesquisa. Em relação aos métodos e procedimentos técnicos realizados, este é um estudo de caso que investiga a prática da construção de uma ontologia biomédica no domínio obstétrico e neonatal, com o objetivo de descrever e explicar o processo de construção da ontologia praticado.

PROCESSO DE DESENVOLVIMENTO DA ONTONEO

OntONeo (*the Ontology of Obstetric and Neonatal domain*) é uma ontologia de domínio ainda em desenvolvimento, que abrange o conhecimento envolvido nos cuidados obstétricos e *NeOnatais*. O conhecimento sobre o domínio obstétrico e neonatal engloba o conhecimento referente ao cuidado da mulher nas fases pré-gravidez, pré-natal, parto e puerpério (pós-natal), além do conhecimento necessário ao cuidado ao recém-nascido. O escopo da OntONeo inclui termos de várias disciplinas médicas, incluindo pelo menos embriologia, anatomia, fisiologia, ginecologia, obstetrícia, pediatria e neonatologia. A ontologia OntONeo abrange termos existentes em outras ontologias, tais como exames de imagem, exames laboratoriais, vacinas, imunizações, distúrbios, doenças, cirurgias, hospitalizações, diagnósticos, tratamentos, dados demográficos, e assim por diante.

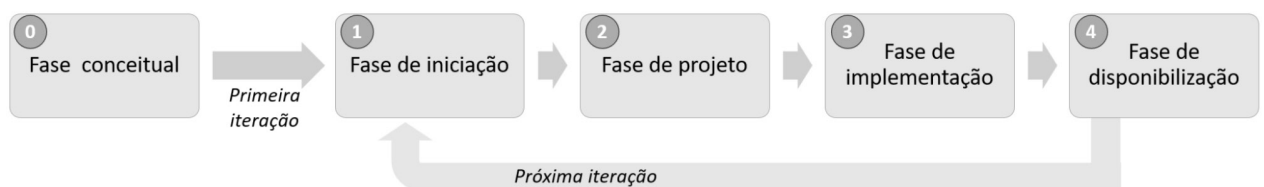
Para o desenvolvimento da OntONeo, optou-se por combinar duas consolidadas metodologias já existentes, a metodologia do realismo ontológico (ARP; SMITH; SPEAR, 2015, p. 13; SMITH; CEUSTERS, 2010) e a metodologia NeOn (SUÁREZ-FIGUEROA, 2010). OntONeo é fundamentada nos princípios da metodologia do realismo ontológico buscando principalmente promover a coerência semântica, tanto para humanos quanto para computadores. Tais princípios são distribuídos ao longo do processo usado para construção da OntONeo.

Além disso, a metodologia NeOn contribui para este trabalho ao fornecer um conjunto de melhores práticas da engenharia de ontologia. Finalmente, o desenvolvimento da OntONeo também se orientou pelos princípios de outra iniciativa bem consolidada, a OBO Foundry. Esta iniciativa definiu um conjunto de princípios guia para o desenvolvimento de ontologia interoperáveis, a fim de minimizar ambiguidades de termos e definições, além de promover a integração de ontologias biomédicas (SMITH et al., 2007).

Do ponto de vista prático, a metodologia utilizada para desenvolver a OntONeo combina as cinco etapas da metodologia do realismo ontológico com os cenários e atividades da metodologia NeOn. O desenvolvimento da OntONeo segue um ciclo de vida de desenvolvimento, iterativo-incremental, como sugerido pelas duas metodologias de referência. Neste contexto, a construção da OntONeo segue cinco fases, como mostrado na figura 1.

Observe que a fase conceitual (número zero) acontece apenas uma vez, enquanto as fases numeradas de 1 a 4 ocorreram uma vez em cada iteração realizada. As fases 1 a 4 constituem a parte iterativa-incremental do ciclo de vida de desenvolvimento da OntONeo. As próximas subseções descrevem cada uma das cinco fases em detalhes.

Figura 1 – Fases de desenvolvimento da ontologia OntONeo



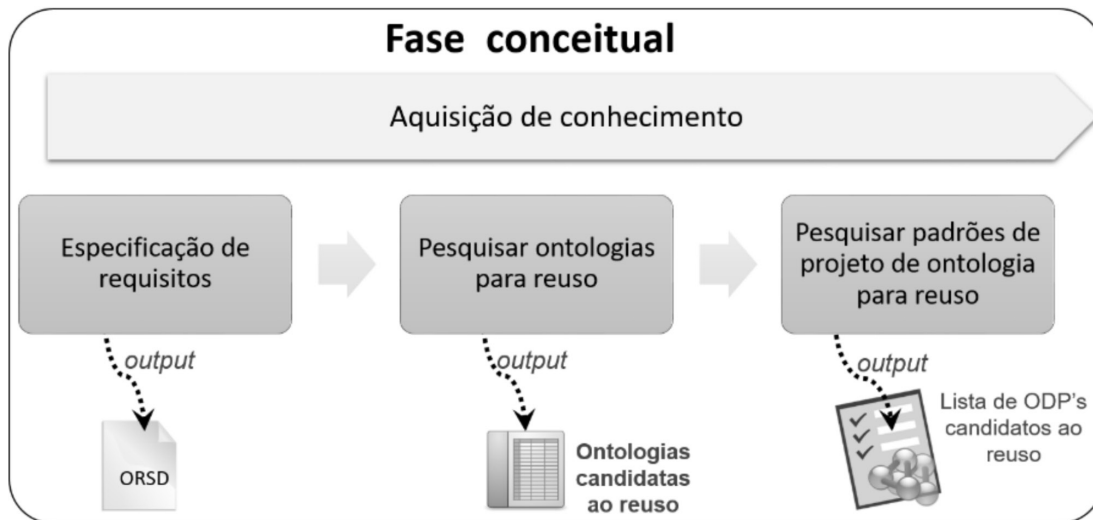
Fonte: Do próprio autor.

FASE CONCEITUAL

A *fase conceitual* visa identificar o propósito da ontologia, o conjunto de requisitos que deve satisfazer e as questões de competência utilizadas para validar a ontologia. Esta fase incorporou dois princípios da metodologia do realismo ontológico, a saber, “demarcar o escopo e o domínio do conhecimento da ontologia” e “reunir informações do domínio ontológico” (ARP; SMITH; SPEAR, 2015, pp. 51-53). Além disso, de acordo com a NeOn, qualquer desenvolvimento de ontologia começa com a atividade de especificação de requisitos de ontologia, e conta com o suporte da atividade de aquisição de conhecimento. Com base nessas duas referências de metodologia, esta fase teve as seguintes atividades: aquisição de conhecimento, especificação de requisitos, pesquisar ontologias para reuso e pesquisar padrões de projeto de ontologia para reuso, organizadas conforme figura 2.

A atividade de aquisição de conhecimento consiste em coletar conhecimento do domínio ontológico de várias fontes de informação, como livros didáticos, documentos, relatórios do domínio alvo, além de consulta aos especialistas do domínio. Assim, para adquirir conhecimento do domínio obstétrico, o pesquisador empregou métodos diretos, por exemplo, entrevistas, observações, análise de documentos e revisão da literatura. Essas entrevistas e observações visaram identificar diretamente os profissionais de saúde e seu ambiente de trabalho qualquer conhecimento relevante sobre o domínio obstétrico e neonatal para subsidiar o desenvolvimento da OntONEo. Na verdade, apesar de ser descrita nesta fase, a atividade de aquisição de conhecimento é realizada durante todas as fases. Esta atividade foi essencial para compreender o conhecimento sobre o domínio obstétrico e neonatal e para garantir melhores decisões sobre a definição e conceituação das entidades.

Figura 2 – Atividades da fase conceitual



Fonte: Do próprio autor.

A atividade especificação de requisitos concentra-se no estabelecimento do conjunto de requisitos que a ontologia deve atender. O objetivo desta atividade é produzir o documento de especificação de requisitos de ontologia, de acordo com o modelo e as recomendações da metodologia *NeOn* disponíveis em (SUÁREZ-FIGUEROA, 2010) e (SUÁREZ-FIGUEROA; GÓMEZ-PÉREZ; VILLAZÓN-TERRAZAS, 2009). O documento registra as informações necessárias para a construção da nova ontologia. Ao final desta atividade é esperado o documento de especificação de requisitos da ontologia, incluindo seu propósito e escopo, a linguagem de implementação que será usada na formalização, o grupo-alvo de usuário e os usos pretendidos para a ontologia, bem como o conjunto de requisitos funcionais e não funcionais que ela deve atingir.

A atividade chamada pesquisar ontologias para reuso (*ontology search*) é a ação em que os ontologistas devem procurar as ontologias já existentes que atendam aos requisitos ou parte dos requisitos da ontologia em desenvolvimento (SUÁREZ-FIGUEROA, 2010). Existem várias ferramentas para pesquisa de ontologia, tais como: Ontobee (XIANG et al., 2011, 2013); Bioportal (NOY et al., 2008; RUBIN; SHAH; NOY, 2008); Ontology Lookup Service (CÔTÉ et al., 2006); Swoogle (DING et al., 2004); Watson RDF (D'AQUIN et al., 2007); e vocab.cc (STADTMÖLLER; HARTH; GROBELNIK, 2013). No caso da OntONeo, por ser uma ontologia do domínio biomédico que segue os princípios OBO Foundry, a atividade de pesquisa de ontologia para reuso priorizou a ferramenta Ontobee, já que sua pesquisa se concentra em ontologias da biblioteca OBO Foundry. Quando a pesquisa na Ontobee não retornava nenhum resultado, o ontologista realizou a pesquisa no BioPortal e Ontology Lookup Service devido à cobertura desses motores de busca para ontologias e terminologias biomédicas. A pesquisa na Ontobee ocorreu de duas maneiras: pesquisa por ontologia de acordo com o domínio de cobertura de ontologia e busca por termos-chave da OntONeo listados no documento de requisitos.

No caso da atividade chamada Pesquisar padrões de projeto de ontologia para reuso (*Design Pattern Search*), o ontologista verifica na lista de padrões de projeto disponível aqueles que podem ser úteis para a ontologia de desenvolvimento da ontologia. O catálogo oficial de padrões de projeto de ontologia (*ontology design patterns*) está disponível no site <http://ontologydesignpatterns.org/>.

A fase conceitual ajuda a identificar o escopo geral e os itens que a ontologia em desenvolvimento deve abranger. No entanto, apenas na próxima fase o ontologista detalhará esses itens. Como o desenvolvimento do OntONeo é iterativo e incremental, todos os itens são divididos, refinados e entregues por partes. As próximas subseções descrevem as quatro fases do ciclo de vida de desenvolvimento incremental iterativo da OntONeo.

FASE DE INICIAÇÃO

O objetivo da fase de iniciação é estabelecer o escopo da iteração e refinar os requisitos que a iteração vai abranger. Além disso, dado o planejamento geral do projeto de ontologia criado na fase conceitual, esta fase define o planejamento de toda iteração. As atividades da fase de inicialização são as seguintes: aquisição de conhecimento (descrita na fase anterior), documentação da ontologia, planejamento de iteração, elicitação de requisitos e seleção de ontologias para reuso, organizadas conforme a figura 3.

A atividade de documentação da ontologia é uma atividade de suporte que gera qualquer documentação valiosa para entender a própria ontologia e as decisões tomadas durante o desenvolvimento da ontologia. Como atividade de aquisição de conhecimento, a atividade de documentação de ontologia ocorre durante todo o desenvolvimento de ontologia. Não existe um padrão de documentação a seguir nesta atividade. A documentação criada depende de cada iteração. Idealmente, a recomendação é criar um único repositório para armazenar a documentação, evitando perdas e replicação de informações relevantes.

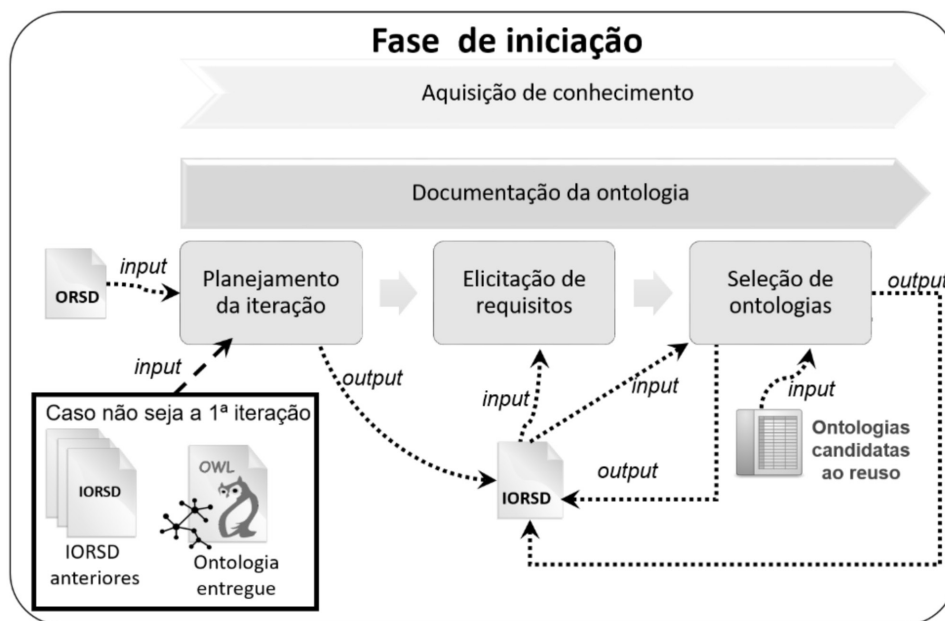
No caso da OntONEo, foram adotados três padrões de documentação: i) uso de anotações ontológicas; ii) página do projeto na internet; e iii) o repositório de desenvolvimento do projeto. O item i é detalhado na fase de implementação, e os itens ii e iii estão detalhados na fase de projeto.

O início de cada iteração de desenvolvimento da ontologia é a atividade de planejamento de iteração, no qual a equipe de ontologistas prioriza os itens que serão trabalhados durante toda a iteração. A metodologia do realismo ontológico sugere que o desenvolvimento da ontologia inicie pelas entidades mais comuns e suas relações, e depois avance para as entidades mais complexas. A fim de orientar a execução completa da iteração e ajudar a comunidade a entender cada versão de ontologia, as decisões sobre o escopo da iteração são documentadas em um documento de especificação de requisitos da iteração, que segue o mesmo modelo de documento para a especificação de requisitos da ontologia, porém, com o escopo reduzido para a iteração. Cada iteração possui seu próprio documento de especificação de requisitos.

A atividade de elicitação de requisitos consiste em extrair conhecimento das fontes de conhecimento disponíveis. O objetivo desta atividade é reunir informações relevantes para a compreensão do domínio do conhecimento da ontologia. Além das técnicas citadas na atividade de aquisição de conhecimento, ainda podem ser empregadas técnicas conhecidas, como por exemplo, *brainstorming*, JAD (*Joint Application Design*), entrevistas não estruturadas com especialistas em domínio, análise informal de documentos, análise de histórias e narrativas de usuários, pesquisa bibliográfica e documental. Toda informação obtida nesta atividade é guardada no documento de especificação de requisitos da iteração.

Durante a atividade de seleção de ontologia, o ontologista deve determinar as ontologias candidatas de reutilização mais adequadas ao escopo da iteração ontológica. Neste caso, uma nova pesquisa por ontologias pode ser necessária, já que houve maior detalhamento sobre o escopo de conhecimento que será representado. As ferramentas de pesquisa de ontologia Ontobee, BioPortal e Ontology Lookup Service também podem ser usadas aqui para facilitar a seleção de ontologias candidatas ao reuso.

Figura 3 – Atividades da fase de iniciação



Fonte: Do próprio autor.

Lembre-se de verificar se a iteração em planejamento é a primeira iteração ou não. Quando não for a primeira, o ontologista deve estar ciente do que já foi desenvolvido nas iterações anteriores, para decidir se será necessária alguma manutenção nos termos já representados.

FASE DE PROJETO

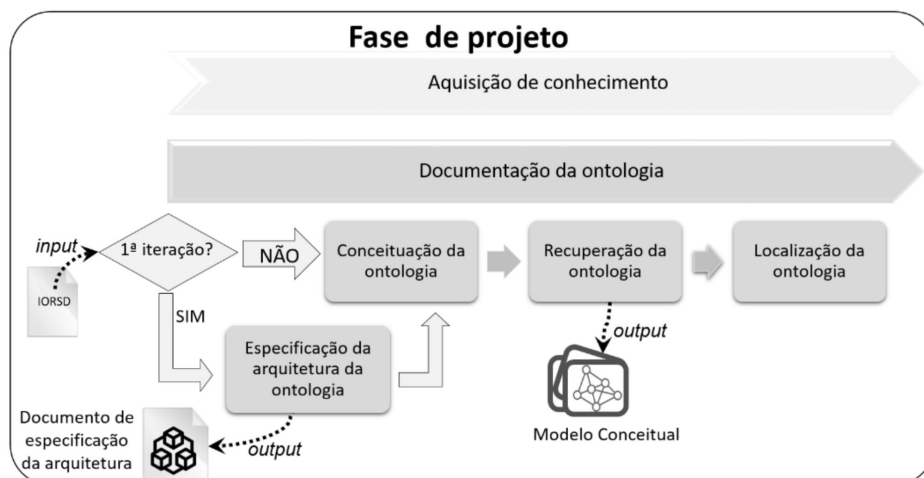
A *fase de projeto* consiste em definir as especificações para a implementação dos requisitos de ontologia. O foco desta fase é obter, ao final, a definição da arquitetura da ontologia, do ambiente de desenvolvimento e do modelo conceitual relacionado aos requisitos obtidos na fase anterior. Nesta fase, em todas as iterações são realizadas as atividades de conceituação da ontologia, recuperação da ontologia e localização da ontologia. A atividade de especificação da arquitetura da ontologia é realizada apenas na primeira iteração, pois todas as iterações subsequentes seguem a mesma definição. São organizadas conforme a figura 4.

A especificação da arquitetura da ontologia é uma atividade que visa definir os elementos arquitetônicos para a ontologia em desenvolvimento. Em geral, os elementos arquitetônicos constituem partes dos requisitos não funcionais da ontologia.

Alguns exemplos dos elementos da arquitetura ontológica são: ontologia de alto nível base para a ontologia em desenvolvimento, o *namespace* e a *url* base da ontologia, página da internet para o projeto da ontologia, o repositório de desenvolvimento, idioma padrão e alternativos para elementos da ontologia, linguagem de codificação/implementação da ontologia; tipo de licenciamento, etc. Foi definido um modelo de documento de especificação de arquitetura para registrar toda decisão acerca da arquitetura da ontologia, conforme apresentado no apêndice 1 deste artigo.

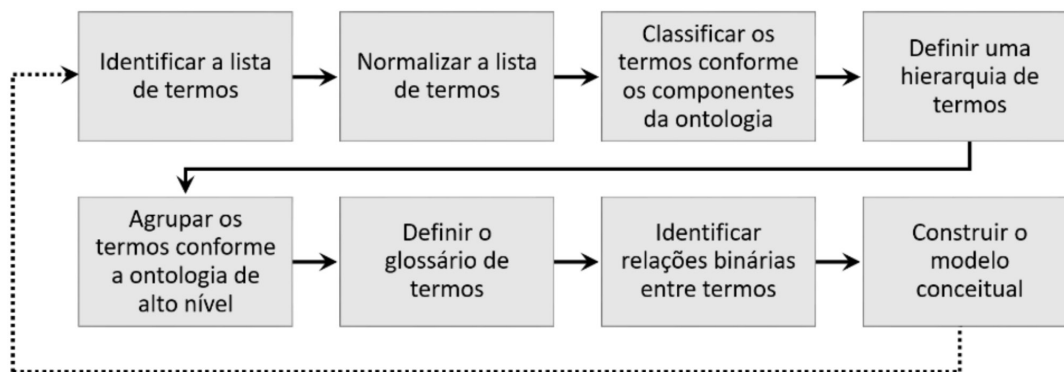
O objetivo da atividade de conceituação da ontologia é organizar e estruturar a informação coletada durante as atividades de aquisição de conhecimento, especificação e elicitação de requisitos. Durante esta atividade, são abordados vários passos da metodologia de realismo ontológico (ARP; SMITH; SPEAR, 2015), conforme figura 5. Sugere-se executar tais passos conforme a sequência definida na figura 5, no entanto, cada ontologista pode definir a sequência mais conveniente para ele.

Figura 4 – Atividades da fase de projeto



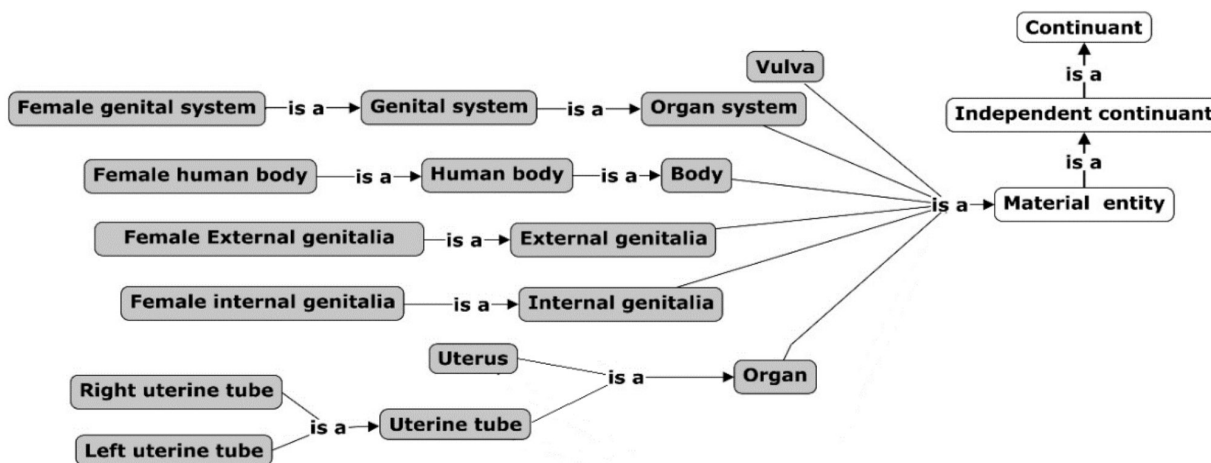
Fonte: Do próprio autor.

Figura 5 – Passos realizados durante a conceituação da ontologia



Fonte: Do próprio autor.

Figura 6 – Exemplo de modelo conceitual criado para a ontologia OntONEo



Fonte: Do próprio autor.

A *Basic Formal Ontology* (BFO) versão 2 foi a ontologia de nível superior escolhida para fundamentar os termos da OntONEo. Assim, para agrupar os termos da ontologia em desenvolvimento conforme as definições da BFO utilizou-se o *Basic Formal Ontology 2.0: Specification and User's Guide* como referência (SMITH et al., 2015). Para construção gráfica do modelo conceitual que reflete a lista de termos da iteração, utilizou-se a ferramenta CMAP Collaborative Ontology Environment (GARCIA et al., 2007; HAYES et al., 2005; HAYES; SAAVEDRA; REICHHERRZER, 2003). A figura 6 apresenta um exemplo de modelo conceitual criado durante o desenvolvimento da OntONEo.

A atividade de recuperação da ontologia visa apoiar a reutilização da ontologia preparando as ontologias potencialmente reutilizáveis para sua importação durante a fase de implementação. Para garantir a consistência da ontologia, OntONEo privilegia a reutilização de ontologias preexistentes que utilizam a mesma ontologia de alto nível, a BFO. A recuperação de ontologias preexistentes acontece em duas abordagens diferentes: uma é a recuperação de toda a ontologia e a outra é recuperar apenas alguns termos de uma ontologia.

Na primeira abordagem, foram recuperadas em sua totalidade as ontologias BFO e Relational Ontology (RO) do portal OBO Foundry¹, baixando os arquivos OWL² (Ontology Web Language) disponíveis. Foram adicionadas a esses arquivos anotações a título de documentação, sendo elas: data de recuperação da ontologia, versão recuperada, responsável pela recuperação, idioma padrão da ontologia quando essa informação não está documentada. Além disso, algumas informações não relevantes para o desenvolvimento da OntONeo, por exemplo informações de contribuintes e comentários, foram removidas.

Na segunda abordagem, foram selecionados apenas termos relevantes de outras ontologias com base nas diretrizes do MIREOT (*Minimal Information to Reference External Ontology Terms*). MIREOT é uma iniciativa que argumenta em favor de importar apenas elementos necessários à reutilização, evitando a sobrecarga da importação de toda a ontologia (COURTOT et al., 2011). Para isso, foi utilizada a ferramenta Ontofox, uma ferramenta baseada na Web capaz de gerar um arquivo de saída OWL com elementos de ontologia selecionados com base nas configurações de entrada dos usuários (XIANG et al., 2010). Assim como na primeira abordagem, após a geração do arquivo de saída da Ontofox, anotações de documentação são incluídas. O arquivo de saída da Ontofox pode ser importado para a ontologia em desenvolvimento, como no caso da OntONeo, ou podem ser incorporadas apenas partes desse arquivo na ontologia de destino. Lembre-se que em um desenvolvimento iterativo-incremental, a cada iteração pode ser necessário incluir novos elementos das ontologias em reuso.

A atividade de localização da ontologia consiste em converter uma ontologia do idioma e cultura original para outra que o ontologista precisa. A execução desta atividade depende dos requisitos de ontologia e do conjunto de ontologias selecionadas como candidatas para reutilização. Basicamente

devem-se manter os rótulos e as definições tanto da ontologia em desenvolvimento como das ontologias reutilizáveis nos idiomas requeridos. No caso da OntONeo, os rótulos e as definições das ontologias em reuso são mantidos em inglês e português.

FASE DE IMPLEMENTAÇÃO

A fase de implementação visa transformar a conceituação produzida em um modelo formal representado por um código legível por máquina. Durante esta fase ocorre a formalização do modelo conceitual, criado na atividade de conceituação, usando a linguagem de codificação definida na atividade de especificação do requisito. Nesta fase, também as ontologias preparadas na atividade de recuperação de ontologia são importadas para a ontologia em desenvolvimento. As atividades realizadas durante a fase de implementação são: importação de ontologia de alto nível, integração de ontologia, anotação da ontologia, formalização da ontologia, e avaliação da ontologia, organizadas conforme a figura 7.

Conforme definido durante a especificação da arquitetura da ontologia, o editor de ontologia usado durante a fase de implementação é o desktop Protégé (versões 4.3 e 5.2).

O Protégé é uma ferramenta de desenvolvimento de ontologia de código aberto amplamente utilizada que ajuda os usuários a definir ontologias com uma interface de usuário gráfica amigável (NOY et al., 2003).

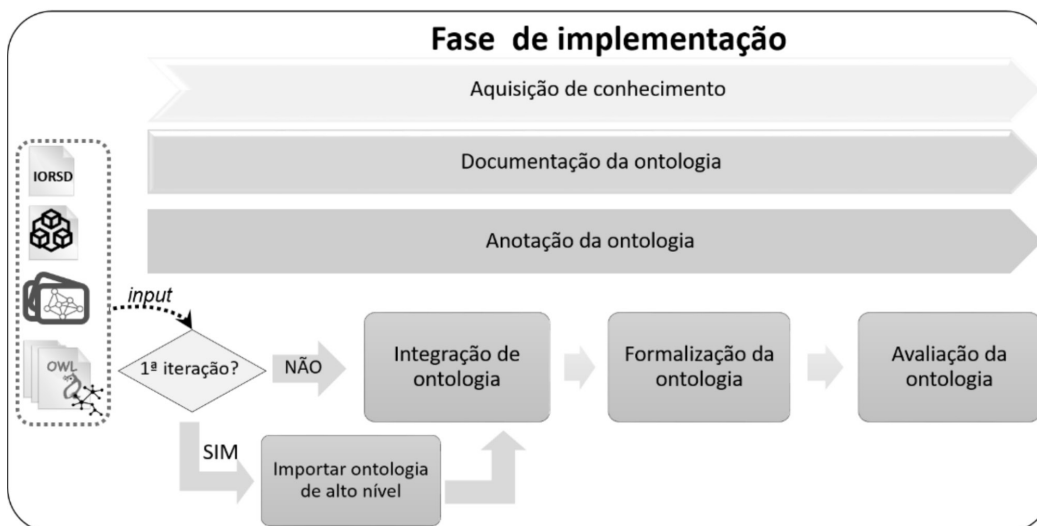
Na primeira iteração de desenvolvimento, inicia-se pela atividade importação de ontologia de alto nível, pois se deve importar para a ontologia em desenvolvimento a ontologia fundacional; no caso da OntONeo, nesta atividade importavam-se as ontologias BFO e RO, cujos arquivos OWL foram preparados durante a fase de projeto.

A atividade de integração de ontologia visa incluir as ontologias que serão reutilizadas na ontologia em construção. Assim, importam-se os arquivos OWL das ontologias reutilizáveis preparadas durante a fase de projeto.

¹ O portal OBO Foundry está disponível no seguinte endereço: <http://www.obofoundry.org/>.

² Documentação disponível em: <http://www.w3.org/TR/owl-features>

Figura 7 – Atividades da fase de implementação



Fonte: Do próprio autor.

Note que, a partir da segunda iteração, o ontologista deve estar ciente da abordagem adotada para a integração ontológica. No caso da integração de uma nova ontologia para reutilização, o ontologista deve importar o arquivo OWL criado na fase anterior como um todo. No entanto, no caso de ontologia já integrada, a nova iteração pode alterar o arquivo original, incluindo ou excluindo elementos.

Na atividade anotação da ontologia são incluídas anotações à ontologia em desenvolvimento ou a algum elemento específico de ontologia. O termo anotação (*annotation*) refere-se a um dado usado para documentar ou adicionar informações a alguns dados. Anotações são uma espécie de metadados da ontologia, ou seja, dados sobre dados. Logo, durante toda a fase de implementação e de acordo com o escopo específico do uso da anotação, são adicionadas anotações à ontologia documentando informações sobre seu desenvolvimento e enriquecendo-a com detalhes relevantes. O quadro 1 do apêndice 2 apresenta as principais anotações consideradas no desenvolvimento da OntONeo.

A formalização da ontologia é a atividade em que o ontologista representa o modelo conceitual usando uma linguagem formal, como por exemplo, a OWL. Nesta atividade, cada termo existente na

lista preliminar refinada durante a fase de iniciação que compõe o modelo conceitual é codificado usando a linguagem de implementação definida no documento de especificação da arquitetura. Esses termos podem ser representados através de classes, relações, funções, instâncias e assim por diante.

A atividade de avaliação da ontologia é o processo de checagem da qualidade técnica da ontologia desenvolvida. É verificada a consistência das definições por meio das questões de competência. Utiliza-se um *plug-in* de consulta SPARQL do próprio Protégé para submeter consultas que refletem as questões de competência, e o resultado da consulta é analisado para verificar se é compatível com o resultado esperado. Analisa-se também a compatibilidade da ontologia construída com os requisitos identificados. Adicionalmente, pode-se submeter à ontologia em construção à ferramenta OOPS! (*Ontology Pitfall Scanner!*) disponível no endereço <http://oops.linkeddata.es/> (POVEDA-VILLALÓN; SUÁREZ-FIGUEROA, 2012). OOPS! é uma ferramenta *on-line* para avaliação ontológica que verifica uma ontologia contra a lista de armadilhas no desenvolvimento de ontologias apresentada por Poveda-Villalón, Suárez-Figueroa e Gómez-Pérez. (2010).

FASE DE ENTREGA

A *fase de entrega* consiste em publicar a nova versão da ontologia. Durante a fase de entrega de cada iteração do desenvolvimento da OntONEo, a nova versão da ontologia é publicada e torna-se disponível para a comunidade. A OntONEo é publicada no Portal OBO Foundry no endereço <http://www.obofoundry.org/ontology/ontoneo.html>, e no BioPortal no endereço <https://biportal.bioontology.org/ontologies/ONTONEO>.

ANÁLISE E DISCUSSÃO DOS RESULTADOS

As ontologias têm cada vez mais se tornado objeto de pesquisa no campo da ciência da informação, como argumentado por Almeida (2013). Muito se discute sobre a contribuição das ontologias como ferramenta adequada para representação e organização do conhecimento (ALMEIDA, 2013; SØERGEL, 1999; VICKERY, 1997). No entanto, pouco se discute sobre construir uma ontologia a fim de representar o conhecimento sem ambiguidade. Assim, este foi o principal objetivo deste estudo. Demonstrar os passos para a construção de uma ontologia. A seção anterior apresentou as fases e atividades realizadas em cada fase. Não só na teoria, mas cada atividade foi realizada na prática para a construção de uma ontologia do domínio obstétrico e neonatal.

A ontologia OntONEo é uma ontologia de domínio que cobre o conhecimento do domínio obstétrico e neonatal fornecendo uma representação consensual deste domínio de conhecimento. Atualmente OntONEo se encontra na versão número 1.5, o que equivale ao produto entregue na sexta iteração de desenvolvimento. OntONEo faz parte da biblioteca de ontologias biomédicas OBO Foundry, demonstrando seu compromisso com os 13 princípios estabelecidos pelo OBO Foundry Consortium. OntONEo é uma iniciativa em seus estágios iniciais de desenvolvimento e a versão atual (disponível em <https://ontoneo.com/>) ainda possui limitações.

Para estabelecer a metodologia para o desenvolvimento, foi realizada uma revisão sistemática da literatura, que identificou a existência de várias metodologias. As metodologias e métodos mais citados de meados dos anos 90 até meados de 2016 são: Enterprise Ontology (USCHOLD; KING, 1995; USCHOLD; GRUNINGER, 1996); Toronto Virtual Enterprise (GRUNINGER; FOX, 1995); Methontology (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JURISTO, 1997); 101 Method (NOY; MCGUINNESS, 2001); DILIGENT (PINTO; STAAB; TEMPICH, 2004); On-To-Knowledge Methodology (SURE; STAAB; STUDER, 2004); NeOn Methodology (SUÁREZ-FIGUEROA, 2010); SABiO - Systematic Approach for Building Ontologies (FALBO, 2014); UPON - Up for ONtology (DE NICOLA; MISSIKOFF; NAVIGLI, 2005), e a metodologia do realismo ontológico (ARP; SMITH; SPEAR, 2015; SMITH; CEUSTERS, 2010). O foco deste artigo não é discutir as metodologias ou apresentar o detalhe de cada uma delas. Na verdade, alguns pesquisadores já o fizeram. Eles identificam, avaliam e comparam diferentes metodologias, como os trabalhos de Mendonça (2015) e Silva (2014).

Em geral, conforme as metodologias citadas anteriormente, o processo de desenvolvimento de ontologia envolve as seguintes atividades: especificação de requisitos de ontologia; aquisição de conhecimento; conceituação e documentação; formalização e implementação. Entretanto, observou-se deficiente consenso sobre qual metodologia é mais apropriada. Além disso, observou-se que a maior parte das metodologias era mais voltada para a forma de desenvolvimento, como por exemplo, a NeOn, enquanto a metodologia do realismo ontológico se preocupa com a semântica da ontologia. Portanto, para o desenvolvimento da OntONEo optou-se por aproveitar o que se tem de melhor em consolidadas metodologias existentes.

Por conseguinte, a metodologia de desenvolvimento da ontologia OntONeo segue primordialmente as recomendações da metodologia do realismo ontológico, o que é uma das garantias de representação do conhecimento sem ambiguidade. Isso porque tal metodologia cumpre-se com três princípios principais: i) a principal fonte de conhecimento é a realidade; ii) o domínio do conhecimento a ser representado é parte da realidade; iii) o domínio especializado conhece a realidade em questão, na medida em que a ciência é a melhor aproximação que se pode ter sobre a realidade. Em suma, o realismo ontológico visa representar a realidade de determinado domínio científico, e não o que se imagina sobre a realidade.

Para o autor do artigo, com base em sua experiência no desenvolvimento de sistemas de informação, a metodologia do realismo ontológico não tem preocupação com o processo ou a forma de desenvolvimento, e sim foca no “*como fazer*” para representar determinado conhecimento. Embora exista certa organização entre os passos sugeridos por essa metodologia, precisamos de processos e ferramentas para apoiar o desenvolvimento deles. Assim, para suprir a lacuna sobre “o que fazer” e “quando fazer” foram recuperadas contribuições da metodologia NeOn.

A metodologia NeOn é uma metodologia para a construção de ontologia que apoia o desenvolvimento colaborativo de ontologias, bem como a reutilização e evolução dinâmica de ontologias em rede em ambientes distribuídos. Optou-se por tal metodologia foi ela foi concebida com base nas melhores práticas de outras três metodologias: Methontology, On-To-Knowledge e Diligent. Ela fornece orientação para todos os aspectos-chave do processo de engenharia ontológica: desenvolvimento colaborativo, reutilização de recursos ontológicos e não ontológicos, evolução e manutenção de ontologias em rede ou com desenvolvimento distribuído. Uma vantagem de usar a metodologia NeOn é seu caráter incremental, que é muito adequado para desenvolvimento de ontologias após a premissa de

falibilismo ao qual o realismo ontológico se refere. O falibilismo reflete a disposição evolutiva da ciência e faz parte do quadro do realismo ontológico (MUNN; SMITH, 2008).

Toda ontologia deve começar seu desenvolvimento pela *fase conceitual*, que é responsável pela definição de escopo da ontologia a ser desenvolvida, pela identificação dos requisitos que ela deve cobrir e pela identificação tanto de ontologias quanto de padrões de projetos existentes candidatos ao reuso. Para isso, uma das atividades primordiais é a aquisição de conhecimento. O conhecimento pode ser adquirido com especialistas do domínio e fontes de conhecimento consolidadas, como livros, padrões internacionais e modelos de referência. A aquisição de conhecimento fornece subsídios para a atividade de especificação de requisitos, que por sua vez subsidia as atividades de pesquisar ontologias para reuso e pesquisar padrões de projeto de ontologia para reuso. Tudo é registrado no documento de especificação de requisitos da ontologia.

A próxima etapa, a *fase de iniciação* é responsável por definir o planejamento de todas as outras fases. Durante a iniciação, determinam-se os requisitos que serão tratados na iteração, identificam-se as principais fontes de conhecimento para se compreender os requisitos, e finalmente detalha-se o conhecimento sobre cada requisito. Na elicitação dos requisitos, sugere-se a identificação de questões de competência e sua expectativa resposta, que poderão ser usadas depois para validação da ontologia. Ao definir essas questões e adquirir brevemente o conhecimento dos especialistas no domínio para possíveis respostas a essas questões, o ontologista consegue aprofundar seu conhecimento sobre o domínio, levando à melhor compreensão do propósito da ontologia e à identificação de novos usos previstos. Adicionalmente, é esperado extrair das fontes de conhecimento disponíveis uma lista de termos relevantes para a compreensão do domínio do conhecimento da ontologia, candidatos a se tornar um elemento da ontologia em desenvolvimento. A lista de termos orienta a seleção de ontologia para reuso e todas as demais fases de desenvolvimento.

A *fase de projeto* determina diversas informações relativas à implementação da ontologia, desde questões arquiteturais ao modelo conceitual relacionado aos requisitos e termos obtidos na fase anterior. São capturadas as principais conceituações do domínio de competência de ontologia, e organizadas graficamente como entidades e relações. Nesta fase, ainda são recuperadas as ontologias que serão reutilizadas pela ontologia em desenvolvimento. Os artefatos de saída desta etapa constituem insumos para a *fase de implementação*, na qual a ontologia é realmente codificada usando a linguagem de codificação definida no documento de arquitetura. Esta é a fase mais importante no processo de desenvolvimento da ontologia, na qual o produto ontológico é efetivamente criado incluindo os recursos ontológicos em reuso. Por fim, existe a *fase de entrega*, em que acontece a publicação da ontologia implementada. As fases de desenvolvimento da ontologia OntONeo foram descritas na seção anterior.

Independentemente do modelo de ciclo de desenvolvimento escolhido, ele sempre inicia pela fase zero, seguindo para a fase um em diante. É importante ressaltar que, embora a figura 1 sugira um fluxo de trabalho em termos iterativo-incremental para as fases de um a quatro, esta metodologia não prescreve nenhum modelo de ciclo de vida específico. Assim, modelos de ciclo de vida tanto cascata quanto iterativo-incremental podem ser adotados. Caso seja definido seguir o desenvolvimento por um ciclo de vida em cascata, o desenvolvimento se dá de forma sequencial, ou seja, só se avança para a fase seguinte quando uma fase estiver completa. No entanto, recomendamos fortemente o desenvolvimento iterativo-incremental, primeiro porque o desenvolvimento da ontologia pode ser em que várias partes que poderiam ser trabalhadas em paralelo e integradas quando completas. Segundo porque podem ser realizadas entregas parciais da ontologia para a comunidade já começar a utilização. E finalmente se pode começar a trabalhar com partes mais gerais da ontologia, e à medida que o desenvolvimento avança, o conhecimento sobre o domínio evolui e consolida. É muito difícil ter uma visão de todo o escopo do domínio científico que a ontologia vai cobrir já no início do projeto. Finalmente,

um aspecto importante do desenvolvimento iterativo-incremental de ontologias é a possibilidade de avaliar a qualidade da estrutura e arquitetura da ontologia construída já no início, nas primeiras iterações, podendo-se agir para corrigir algum problema tão logo ele apareça, além de evitar que novos problemas de mesma natureza ocorram.

Uma vantagem da abordagem usada no desenvolvimento da OntONeo é sua capacidade de se acoplar a diferentes tamanhos de projeto de ontologia. A abordagem se adaptada facilmente a uma série de variáveis, como por exemplo, o tamanho da ontologia, o domínio de interesse, a complexidade da ontologia em desenvolvimento, a experiência do time de desenvolvimento e o tamanho do time de desenvolvimento.

Um diferencial deste trabalho está na associação das diversas ferramentas desenvolvidas para apoiar os desenvolvedores de ontologia com as atividades de desenvolvimento que podem ser usadas. Por exemplo, o uso da ferramenta Ontobee para pesquisa de ontologia para reuso, e o uso da ferramenta Ontofox para recuperar os elementos de uma ontologia para serem reutilizados.

Este trabalho é parte da pesquisa de doutorado da autora, que traz maiores detalhes sobre a metodologia usada, além de exemplos de como cada atividade foi realizada e resultados obtidos em cada fase.

CONSIDERAÇÕES FINAIS

Este artigo apresentou uma proposta de processo de construção de uma ontologia de domínio testado na prática com o desenvolvimento de uma ontologia de domínio chamada OntONeo, que trata o conhecimento acerca dos registros de saúde eletrônicos gerados nos cuidados obstétricos e neonatais.

Para isso, no desenvolvimento da OntONeo, optou-se por combinar duas consolidadas metodologias já existentes, a metodologia do realismo ontológico (ARP; SMITH; SPEAR, 2015, p. 13; SMITH; CEUSTERS, 2010) e a metodologia *NeOn* (SUÁREZ-FIGUEROA, 2010).

O realismo ontológico, além de fornecer motivos teóricos, oferece um conjunto de recomendações para realizar as atividades do desenvolvimento ontológico. Neste contexto, um domínio real, em vez de um modelo de dados, deve ser usado para representar a ciência estabelecida nesse domínio, proporcionando melhor aproximação da realidade. Em relação a aspectos práticos, a NeOn fornece uma visão pragmática orientada para os principais aspectos do processo de engenharia de ontologia através de uma lista extensa e racional de processos e atividades de desenvolvimento compiladas de diversas outras metodologias.

Com este estudo, espera-se contribuir com o avanço da pesquisa em construção de ontologias no campo da ciência da informação, dada sua aplicação na solução problemas de organização e recuperação de informações em ambientes informacionais de diversos campos científicos.

AGRADECIMENTOS

Esta pesquisa é parte da tese de doutorado da autora para obtenção do título de Doutora em Ciência da Informação pela Universidade Federal de Minas Gerais. Durante o desenvolvimento da tese, a autora recebeu apoio em formato de bolsa de estudos de dedicação exclusiva das instituições de fomento CNPq e Fapemig. Adicionalmente, a autora recebeu apoio de bolsa de estudos para doutorado sanduíche pelo processo número BEX 10767/14-2 financiado pela Capes.

Agradece-se também à Dra. Zilma N. Reis, da Universidade Federal de Minas Gerais, à Dra. Faye Justicia-Linde, e ao Dr. Peter L. Elkin, ambos da Universidade do Estado de Nova York, em Buffalo, pelas orientações em questões médicas; ao Dr. Maurício Barcellos de Almeida, da Universidade Federal de Minas Gerais, por sua orientação e parceria durante todo o desenvolvimento da tese que deu origem a este artigo; e ao Dr. Barry Smith, da Universidade do Estado de Nova York, em Buffalo, por sua orientação em questões ontológicas e terminológicas.

REFERÊNCIAS

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

ARP, R.; SMITH, B.; SPEAR, A. D. *Building ontologies with basic formal ontology*. Cambridge, Massachusetts: The Mit Press, 2015.

BITTNER, T.; DONNELLY, M.; WINTER, S. Ontology and semantic interoperability: large-scale 3D data integration: challenges and opportunities. 2005. P. 139-160.

CÔTÉ, R. G. et al. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC bioinformatics*, v. 7, n. 1, p. 97, 2006.

COURTOT, M. et al. MIREOT: The minimum information to reference an external ontology term. *Applied Ontology*, v. 6, n. 1, p. 23-33, 2011.

D'AQUIN, M. et al. Watson: a gateway for the semantic web. In: EUROPEAN SEMANTIC WEB CONFERENCE, 4., 2007, Austria. *Electronic proceedings...* Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.119.32>>. Acesso em: 14 ago. 2017.

DE NICOLA, A.; MISSIKOFF, M.; NAVIGLI, R. A proposal for a unified process for ontology building: UPON. In: INTERNATIONAL CONFERENCE ON DATABASE AND EXPERT SYSTEMS APPLICATIONS, 16., 2005, Copenhagen, Denmark. *Proceedings ...* Berlin: Springer, 2005. P. 655-664.

DING, Li et al. Swoogle: A semantic web search and metadata engine. In: ACM CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 13., 2004, Washington, DC, USA. *Proceedings ...* New York: ACM, 2004. P. 652-659.

FALBO, R. D. A. SABiO: Systematic Approach for Building Ontologies. In: JOINT WORKSHOP ONTO.COM/ODISE ON ONTOLOGIES IN CONCEPTUAL MODELING AND INFORMATION SYSTEMS ENGINEERING, 1., 2014, Rio de Janeiro. *Proceedings ...* 2014.

FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. Methontology: from ontological art towards ontological engineering. In: SPRING SYMPOSIUM ON ONTOLOGICAL ENGINEERING, 1997, Palo Alto, CA. *Proceedings...* Menlo Park, CA: The AAAI Press, 1997. P. 33-40.

GANGEMI, A. Ontology design patterns for semantic web content. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 2005, Galway, Ireland. *Proceedings...* Berlin: Springer, 2005. P. 262-276.

_____ et al. A formal ontological framework for semantic interoperability in the fishery domain. In: EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE ECAI'02 - WORKSHOP ON ONTOLOGIES AND SEMANTIC INTEROPERABILITY, 2002, Lyon, France. *Proceedings ...* 2002.

- GARCIA, A. et al. CMAPS supporting the development of OWL ontologies. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 7., 2008, Karlsruhe, Germany. *Proceedings ...* 2008.
- GIL, A. C. *Métodos e técnicas de pesquisa social*. 6. ed. São Paulo: Atlas, 2008. 200 p.
- GRUBER, T. R. Ontology. In: LIU, L.; ÖZSU, M. T. (Ed.). *Encyclopedia of database systems*. Londres: Springer-Verlag, 2008. P. 1963-1965.
- GRUNINGER, M.; FOX, M. S. Methodology for the design and evaluation of ontologies. In: WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWLEDGE SHARING, AT THE INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 14., 1995, Montreal. *Proceedings...* Montreal: IJCAI, 1995.
- HAYES, P.; SAAVEDRA, R.; REICHHERZER, T. A collaborative development environment for ontologies (CODE). In: SEMANTIC INTEGRATION WORKSHOP 2003, Sanibel Island, Florida, USA. *Proceedings ...* CEUR Workshop, v. 7. p.139.
- _____ et al. COE: Tools for collaborative ontology development and reuse. In: KNOWLEDGE CAPTURE CONFERENCE (KCAP. 2005), 2005, Banff, Canada. *Electronic proceedings...* Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.592.2720>>. Acesso em: 14 ago. 2017.
- MARCONI, M. D. A.; LAKATOS, E. M. *Fundamentos de metodologia científica*. 5. ed. São Paulo: Atlas, 2003. 311 p.
- _____. *Metodologia do trabalho científico*. 7. ed. São Paulo: Atlas 2007. 228 p.
- 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. 320 f. Tese (Doutorado em Ciência da Informação)– Escola de Ciência da Informação da Universidade Federal de Minas Gerais, Brasil, 2015. Disponível em: <<http://hdl.handle.net/1843/BUBD-A35H3K>>. Acesso em: 14 ago. 2017.
- MUNN, K.; SMITH, B. *Applied ontology: an introduction: metaphysical research*. Londres: Ontos Verlag, 2008.
- NOY, N. F.; MCGUINNESS, D. L. *Ontology development 101: a guide to creating your first ontology: Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880* 2001. Disponível em: <https://protege.stanford.edu/publications/ontology_development/ontology101.pdf>. Acesso em: 14 ago. 2017.
- _____ et al. Protege-2000: an open-source ontology-development and knowledge-acquisition environment. 2003. In: AMIA SYMPOSIUM, 2003, Washington, DC, USA. *Proceedings ...* 2003.
- _____ et al. BioPortal: a web repository for biomedical ontologies and data resources. In: INTERNATIONAL SEMANTIC WEB CONFERENCE (ISWC2008), 7., 2008, Karlsruhe, Germany. *Proceedings ...* 2008.
- PINTO, H. S.; STAAB, S.; TEMPICH, C. DILIGENT: towards a fine-grained methodology for distributed, loosely-controlled and evolving Engineering of ontologies. In: EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, 16., 2004, Valencia, Spain. *Proceedings ...* Amsterdam: IOS Press, 2004. P. 393-397.
- POVEDA-VILLALÓN, M.; SUÁREZ-FIGUEROA, M. C. OOPS!: Ontology Pitfalls Scanner!, 2012. *International Journal on Semantic Web and Information Systems*, v. 10, n. 2, p. 7-34, 2014.
- _____; GÓMEZ-PÉREZ, A. Common pitfalls in ontology development. In: CONFERENCE OF THE SPANISH ASSOCIATION FOR ARTIFICIAL INTELLIGENCE, 2009, Seville, Spain. *Proceedings ...* Berlin: Springer, 2009. P. 91-100.
- _____. Common pitfalls in ontology development. In: _____. *Current topics in artificial intelligence*. Londres: Springer, 2010a. P. 91-100.
- _____. A double classification of common pitfalls in ontologies. In: ONTOQUAL 2010: WORKSHOP ON ONTOLOGY QUALITY AT THE INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT (EKAW 2010), 17., 2010, Lisbon, Portugal. *Proceedings...* Portugal: [s.n.], 2010b.
- RUBIN, D. L.; SHAH, N. H.; NOY, N. F. Biomedical ontologies: a functional perspective. *Briefings in bioinformatics*, v. 9, n. 1, p. 75-90, 2008.
- SILVA, D. L. D. *Ontologias para representação de documentos multimídia: análise e modelagem*. 2014. 442 f. Tese (Doutorado em Ciência da Informação)– Escola de Ciência da Informação da Universidade Federal de Minas Gerais, Brasil, 2014. Disponível em: <<http://hdl.handle.net/1843/BUOS-9NCGYM>>. Acesso em: 14 ago. 2017.
- SMITH, B. Ontology. In: FLORIDI, L. (Ed.). *The Blackwell guide to the philosophy of computing and information*. Oxford: Blackwell, 2003a. Cap. 11. P.155-166.
- _____. *Ontology and information systems*. 2003b. Disponível em: <[http://ontology.buffalo.edu/ontology\(PIC\).pdf](http://ontology.buffalo.edu/ontology(PIC).pdf)>. Acesso em: 14 ago. 2017.
- _____ et al. *Basic formal ontology 2.0: specification and user's guide*. 2015.
- _____ et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, v. 25, n. 11, p. 1251-1255, 2007.
- _____; CEUSTERS, W. Ontological realism: a methodology for coordinated evolution of scientific ontologies. *Applied ontology*, v. 5, n. 3-4, p. 139-188, 2010.

SØERGEL, D. The rise of ontologies or the reinvention of classification. *Journal of the Association for Information Science and Technology*, v. 50, n. 12, p. 1119-20, 1999.

STADTMÖLLER, S.; HARTH, A.; GROBELNIK, M. Accessing information about linked data vocabularies with vocab. cc. In: LI, Juanzi et al (Ed.). *Semantic web and web science*. Londres: Springer, 2013. P. 391-396.

SUÁREZ-FIGUEROA, M. C. *NeOn methodology for building ontology networks: specification, scheduling and reuse*. 2010. 268 f. Tese (Doutorado em Informática)- Departamento de Inteligência Artificial da Universidade Politécnica de Madrid, Madri, Espanha, 2010. Disponível em: < <http://oa.upm.es/3879/>>. Acesso em: 14 ago. 2017.

_____; GÓMEZ-PÉREZ, A.; VILLAZÓN-TERRAZAS, B. How to write and use the ontology requirements specification document. In: MEERSMAN, R. et al (Ed.). *On the move to meaningful internet systems: OTM 2009*. Londres: Springer Berlin Heidelberg, 2009. Cap. 16. P. 966-982.

SURE, Y.; STAAB, S.; STUDER, R. On-to-knowledge methodology (OTKM). In: STAAB, S.; STUDER, R. (Ed.). *Handbook on ontologies*. Berlin: Springer Berlin Heidelberg, 2004. Cap. 6. P. 117-132.

USCHOLD, M.; GRUNINGER, M. Ontologies: principles, methods and applications. *The knowledge engineering review*, v. 11, n. 02, p. 93-136, 1996.

USCHOULD, M.; KING, M. Towards a methodology for building ontologies. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWLEDGE, 14., 1995, Montreal. *Electronic proceedings...* Montréal: DBLP, 1995. Disponível em: <<http://www.iai.ed.ac.uk/project/oplan/documents/1995/95-ont-ijcai95-ont-method.pdf>>. Acesso em: 14 ago. 2017.

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

XIANG, Z. et al. OntoFox: web-based support for ontology reuse. *BMC Research Notes*, v. 3, n. 1, p. 175, 2010.

_____. et al. Ontobee: a linked data server and browser for ontology terms, ontology. In: INTERNATIONAL CONFERENCE ON BIOMEDICAL ONTOLOGY (ICBO2011), 2., 2011, Buffalo, NY, USA. *Proceedings...* 2011.

_____. Ontobee: a linked data server that publishes RDF and HTML data simultaneously. *Semantic Web Journal*, v. 261, 2013. Disponível em: <http://www.semantic-web-journal.net/system/files/swj261_0.pdf>. Acesso em: 14 ago. 2017.

YIN, R. K. *Case study research: design and methods*. 4th ed. Los Angeles, CA, US: Sage Publications, 2009.

APÊNDICE 1

MODELO DE DOCUMENTO DE ESPECIFICAÇÃO DE ARQUITETURA DA ONTOLOGIA.

Obstetric and Neonatal Ontology Architecture Specification Document		
1	Top-level ontology to extend	Basic Formal Ontology version 2.0
2	Ontology namespace	http://purl.obolibrary.org/obo/ontoneo/
	PREFIX or IDSPACE	Ontoneo
	Ontology IRI	http://purl.obolibrary.org/obo/ontoneo/ontoneo.owl
	Default purl	http://purl.obolibrary.org/obo/ontoneo/
	Local identifier	<a href="http://purl.obolibrary.org/obo/ontoneo/<PREFIX>_<999999999>">http://purl.obolibrary.org/obo/ontoneo/<PREFIX>_<999999999> Ex: http://purl.obolibrary.org/obo/ontoneo/ONTONEO_00000001
3	Ontology project website	https://ontoneo.com/
4	Development repository	https://github.com/ontoneo-project/Ontoneo
5	Development tools	Protégé version 5.2 Ontobee ontology search engine Ontofox system to get pieces of ontologies
6	Coding language	OWL 2 Web Ontology Language
7	Default language to ontology elements	English
8	Alternative language to ontology elements	Portuguese
9	License	Creative Commons CC-BY license version 4.0
10	Minimal ontology documentation	
<p>Each new element must be documented with: term creator, date of term creation, label both in English and Portuguese, definition both in English and Portuguese. Each reuse element must be documented with: source ontology.</p>		
11	Additional information	
<p>The ontology should follow the principles of the OBO Foundry. The ontology should reuse other ontologies already accepted by the OBO Foundry whenever possible. The ontology should prioritize the use open source tools. The ontology should use the modularization strategy in order to promote reuse flexibility of thematic modules.</p>		

APÊNDICE 2**ANOTAÇÕES DOCUMENTADAS NA ONTONEO**

Quadro 1 – Principais anotações usadas na ontologia OntONEo

Fonte		Anotação	Escopo de uso
PREFIX	namespace		
dc	http://purl.org/dc/elements/1.1/	creator	Ontology
		contributor	Ontology
		license	Ontology
		date	
		source	
foaf	http://xmlns.com/foaf/0.1/	homepage	Ontology
		mbox	Ontology
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#	Description	
rdfs	http://www.w3.org/2000/01/rdf-schema#	label	Ontology
		range	Properties
		domain	Properties
		comment	Ontology and entities
		isDefinedBy	Any component
owl	http://www.w3.org/2002/07/owl#	versionIRI	Ontology
		versionInfo	Ontology
		priorVersion	Ontology
obo	http://purl.obolibrary.org/obo/	IAO_0000111 editor preferred term	Entity
		IAO_0000114 has curation status	Entity and object property
		IAO_0000115 definition	Entity
		IAO_0000117 term editor	Entity
		IAO_0000118 alternative term	Entity
		IAO_0000119 definition source	Entity
		IAO_0000412 imported from	Imported components
obolInOwl	http://www.geNeOntology.org/formats/obolInOwl#	id	Entity
		created_by	
		creation_date	
		hasDbXref	Entity
protege	http://protege.stanford.edu/plugins/owl/protege#	defaultLanguage	Ontology
ontoneo	http://purl.obolibrary.org/obo/ontoneo/	ONTONEO_1000000 alternative language	Ontology
		ONTONEO_1000002 translator to Portuguese	Entity and relation

Fonte: Do próprio autor.

Semantic middleware for industrial sensors

Fernando Silva Parreiras

Doutor em Ciência da Computação pela Universität Koblenz-Landau, mestre em Ciência da Informação pela UFMG e graduado em Ciência da Computação pela FUMEC. Professor Adjunto na Universidade FUMEC, Belo Horizonte, MG, Brasil

Vitor Afonso Pinto

Doutorando e Mestre em Sistemas de Informação e Gestão do Conhecimento pela Universidade FUMEC. Pós-graduado em Gestão Estratégica de Tecnologia da Informação pela Universidade Gama Filho e graduado em Sistemas de Computação pela Universidade Federal Fluminense.

Marco Antônio Calijorne Soares

Mestrando em Sistemas de Informação e Gestão do Conhecimento pela Universidade FUMEC. Graduação em Ciência da Computação pela PUC Minas.

Daniel Henrique Mourão Falci

Mestrando em Sistemas de Informação e Gestão do Conhecimento pela Universidade FUMEC. Graduado em Análise e Desenvolvimento de Sistemas pela Universidade Estácio de Sá (Unesa) - Belo Horizonte, MG - Brasil. Arquiteto de Software na Visual Sistemas Eletrônicos LTDA. - Belo Horizonte, MG - Brasil.

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

ABSTRACT

For many years, plant engineers have used data collected from industrial sensors for supporting the diagnosis of failures. Recently, data scientists are using these data to make predictions on industrial processes. However, the meaning and the relationships of each specific sensor is unknown to people outside the engineering context. Conventional approaches to create a semantic layer for industrial sensors require a rigid “term alignment” followed by a lot of manual efforts. Hence, the problem is frequently set aside by industries. However, this condition limits the usage of advanced analytics tools in industries, preventing the capture of potential benefits. Since there are naming conventions and some other rules defined by engineers, this study takes these standards into account and analyze the metadata of sensors intending to automate the creation of a semantic middleware able to indicate the meaning of each sensor and its relationships with other sensors, equipments, areas, plants and other entities. This study intends to answer the following research question: Which approach could automate the creation of a semantic middleware for industrial sensors? In order to address the objectives of this study, we performed an empirical research using sensor metadata from three different plants from a mining company. As a result, we present MINDSense, a method that creates an ontology capable of describing the meaning of industrial sensors and its relationships. We conclude that this method contributes to leverage advanced analytics in industries and to increase the potential of new studies on top of industrial sensors data.

Keywords: Semantic Middleware. Ontology. Industrial sensors. Data Science.

Middleware semântico para sensores industriais

RESUMO

Por muitos anos, engenheiros de plantas tem usado dados coletados de sensores industriais para suportar o diagnóstico de falhas. Recentemente, cientistas de dados estão usando esses dados para fazer previsões em processos industriais. Contudo, o significado de cada sensor específico e suas relações é desconhecido para pessoas que não estão no contexto da engenharia. As abordagens convencionais para criar uma camada semântica para sensores industriais requerem rígido “alinhamento de termos” seguido de muitos esforços manuais. Por isso, o problema é frequentemente deixado de lado pelas indústrias. No entanto, esta condição limita o uso de ferramentas de analytics em indústrias, evitando a captura de benefícios potenciais. Uma vez que existem convenções de nomenclatura e outras regras definidas pelos engenheiros, este estudo leva esses padrões em consideração e analisa os metadados de sensores visando automatizar a criação de um middleware semântico capaz de indicar o significado de cada sensor e suas relações com outros sensores, equipamentos, áreas, plantas e outras entidades. Este estudo pretende responder à seguinte questão de pesquisa: Qual abordagem poderia automatizar a criação de um middleware semântico para sensores industriais? Para atender aos objetivos deste estudo, realizamos uma pesquisa empírica usando metadados de sensores de três plantas diferentes de uma empresa de mineração. Como resultado, apresenta-se o MINDSense, método que cria uma ontologia capaz de descrever o significado dos sensores industriais e seus relacionamentos. Conclui-se que este método contribui para alavancar análises avançadas nas indústrias e aumentar o potencial de novos estudos sobre dados de sensores industriais.

Palavras-chave: Camada semântica. Ontologia. Sensores industriais. Ciência de dados

Middleware semánticos para sensores industriales

RESUMEN

Durante años, los ingenieros utilizaron datos recolectados de sensores industriales para soportar el diagnóstico de fallas. Recientemente, los científicos de datos están utilizando estos datos para hacer previsiones en procesos industriales. El significado de cada sensor específico y sus relaciones son desconocidas para las personas que no están en el contexto de la ingeniería. Los enfoques convencionales para crear una semántica para sensores industriales requieren un rígido “alineamiento de términos” seguido de esfuerzos manuales. Por eso, el problema es dejado de lado por las industrias. Esta condición limita el uso de herramientas de análisis en las industrias, evitando la captura de beneficios potenciales. Una vez que existen convenciones de nomenclatura y otras reglas definidas por los ingenieros, este estudio lleva estos estándares en consideración y analiza los metadatos de sensores para automatizar la creación de un middleware semántico capaz de indicar el significado de cada sensor y sus relaciones con otros sensores, equipos, áreas, plantas y otras entidades. Este estudio pretende responder a la siguiente pregunta: ¿Qué enfoque podría automatizar la creación de un middleware semántico para sensores industriales? Para atender a los objetivos de este estudio, realizamos una investigación empírica usando metadatos de sensores de tres plantas diferentes de una empresa minera. Como resultado, presentamos el MINDSense, un método que crea una ontología capaz de describir el significado de los sensores industriales y sus relaciones. Concluimos que este método contribuye a aprovechar análisis avanzados en las industrias y aumentar el potencial de nuevos estudios sobre datos de sensores industriales.

Palabras clave: Middleware semántico. Ontología. Sensores industriales. Ciencia de datos

INTRODUCTION

Companies are aware that the timely analysis and monitoring of business processes are essential to identify non-compliant situations and react immediately to those inconsistencies (VERA-BAQUERO; COLOMO-PALACIOS; MOLLOY, 2016). It is also known by companies that some studies that were difficult to conduct in the past due to lack of data can now be carried out. (LIU et al, 2016). Big Data and the mechanisms by which it is produced and disseminated introduce substantial changes in the ways information is generated and is made relevant for organizations. (CONSTANTIOU; KALLINIKOS, 2015).

Notably, industries are researching and implementing Big Data technologies intending to make their automation assets more reliable and predictive. The integration of IT systems with automation systems can become an essential tool for business users in the decision-making process. (VERA-BAQUERO; COLOMO-PALACIOS; MOLLOY, 2016). ICT technologies inside the engineering domain turn devices and equipment into intelligent systems, communicable and integrated from the field level to the operation level with a seamless data flow in both directions.

Nevertheless, even with all the progress that has been made, companies are still struggling with how to capture insights that are not obvious. It is a problem of how to discover meaningful relationships. (HURWITZ; KAUFMAN; BOWLES, 2015). For many years, plant engineers have used data collected from industrial sensors for supporting the diagnosis of failures. However, the meaning and the relationships of each specific sensor is unknown to people outside the engineering context.

Recently, data scientists are trying to use these data to make predictions on industrial processes. But, conventional approaches to create a semantic layer for industrial sensors require a rigid “term alignment” followed by a lot of manual efforts, not mentioning the difficulties to keep the semantic layer up-to-date. Hence, the problem is frequently set aside by industries.

However, this condition limits the usage of advanced analytics tools in industries, preventing the capture of potential benefits. In this context, the following research question emerges: Which approach could automate the creation of a semantic middleware for industrial sensors?

Since there are naming conventions and some other rules defined by engineers, in this study we take these standards into account and propose MINDSense - an acronym for “Middleware for Industrial Sensors”. MINDSense analyzes the metadata of sensors to automate the creation of a semantic middleware able to indicate the meaning of each sensor and its relationships with other sensors, equipments, areas, plants and other entities. MINDSense contributes to leverage advanced analytics in industries and to increase the potential of new studies on top of industrial sensors data.

This paper is structured as follows. An example of how the lack semantic prevents the development of an analytics layer on top of industrial sensor data is presented in section Running Example. In section Background, the academic foundation related to the main topic of this study is discussed. In section Methods, the details regarding the development of MINDSense are provided. The achievements reached with the usage of MINDSense are presented in Results section. Other approaches close to MINDSense approach are presented in section Related Work. The paper is concluded in the Conclusion section.

RUNNING EXAMPLE

A multi-plant industry usually has a centralized area for performing data science researches. However, sensors, typically counted on dozens per equipment, can be named in a different way in different plants. This happens because plant engineers can define a different naming convention for the sensors under his responsibility. Besides that, for several reasons, it is possible to exist naming deviations even inside a single plant. This way, to perform any study, data scientists strongly depend on plant engineers to understand the meaning of each specific sensor and its relationships.

Consider a scenario where a data scientist, examining patterns to create a predictive model, needs the full list of sensors that monitor a given equipment that in turn, is used across multiple plants. In this case, this data scientist would need to talk to several plant engineers while manually conceiving his list. The lack of term alignment between those involved hinders the process, creating an environment that is subjective and susceptible to errors. Now, suppose a new sensor is deployed on a plant after the data scientist gathered all information? The process should be restarted. Thus, the lack of semantic for industrial sensors slows down the process of development of analytics.

BACKGROUND

Most real-world data are not in a form that can be directly recorded by a computer. These quantities typically include temperature, pressure, distance, velocity, mass, and energy output (such as optical, acoustic, and electrical energy). A physical quantity must first be converted to an electrical quantity (voltage, current, or resistance) using a sensor or transducer. (AUSTERLITZ, 2003). Thus, transducers and sensors are used to convert a physical phenomenon into an electrical signal (voltage or current) that will be then converted into a digital signal used for the next stage such as a computer, digital system, or memory board. (EMILIO, 2013).

At the highest level, a sensor is something that, when stimulated, detects some aspect of physical phenomena (called input). By way of a transducer, the sensor turns the measurement into a signal so it can be electronically processed and then measured or recorded as output. (STIMMEL, 2015). Devices with input function are called sensors because they detect a physical event that changes according to some events as, for example, heat or force. Instead, device with output function are called actuators and are used in control system to monitor and compare the value of external devices. (EMILIO, 2013). The output is used as input to a further system or process that triggers some responsive action.

This is called actuation and may require yet another transducer to convert the output to yet another signal type. (STIMMEL, 2015).

Sensors can monitor the physical world by detecting and measuring different types of environmental information. By feeding suitable applications with such type of information via various types of physical world objects, the Internet would move from “interconnected computers” to “interconnected things.” (CHAQFEH; MOHAMED, 2012). Sensor types include: temperature sensors, magnetic field sensors, potentiometers, light detection sensors, among others. (EMILIO, 2013). Table 1 gives more details about sensors and transducers.

SEMANTIC MIDDLEWARE

The development of tiny sensors and actuators can realize intelligent context-aware networking in large factory environments, automotive networks, smart homes and offices, and social services support including earthquake warnings, patient monitoring and context-aware support in emergency situations. (CHAQFEH; MOHAMED, 2012).

A Middleware platform for the IoT provides an abstract layer interposed between the IT infrastructure and the applications. It aims to hide the technological details to enable the application developers to focus on the development of the IoT applications. (CHAQFEH; MOHAMED, 2012). When billions of sensors are connected to the Internet, it is not feasible for people to process all the data collected by those sensors. Context-awareness computing techniques, such as IoT middleware are proposed to better understand sensor data and help decide what data needs to be processed. (XU; HE; LI, 2014).

Besides describing all IoT infrastructure, a semantic middleware platform explains the meaning of each existing device for any application or consumer. The general idea is that semantic middleware is able to automatically discover and store metadata about IoT devices. IoT middleware may have several features, such as described in table 2.

Table 1 – Type of sensors

Sensor Type	Description
Temperature sensors	Contain electrical parameters that vary with temperature, following well-characterized transfer functions
Optical sensors	Used for detecting light intensity. Typically, they respond only to particular wavelengths or spectral bands.
Force and Pressure Transducers	A wide range of sensors are used for measuring force and pressure. Most pressure transducers rely on the movement of a diaphragm mounted across a pressure differential.
Magnetic Field Sensors	Used to measure either varying or fixed magnetic fields.
Ionizing Radiation Sensors	Can be particles produced by radioactive decay, such as alpha or beta radiation, or high-energy electromagnetic radiation, including gamma and X-rays. In many of these detectors, a radiation particle (a photon) collides with an active surface material and produces charged particles, ions, and electrons, which are then collected and counted as pulses (or events) per second or measured as an average current.
Position (Displacement) Sensors	A wide variety of transducers are used to measure mechanical displacement or the position of an object. Some require actual contact with the measured object; others do not.
Humidity Sensors	Relative humidity is the moisture content of the air compared to air completely saturated with moisture and is expressed as a percentage.
Fluid Flow Sensors	Many industrial processes use fluids and need to measure and control their flow in a system. A wide range of transducers and techniques are commonly used to measure fluid flow rates (expressed as volume per unit time passing a point).
Fiber Optic Sensors	Used to measure a wide range of quantities, including temperature, pressure, strain, displacement, vibration, and magnetic field, as well as sensing chemical and biomedical materials. They are immune from electromagnetic interference (EMI), can operate in extremely harsh environments, can be very small, and are fairly sensitive

Source: Based on Austerlitz (2003)

Table 2 – Features of a semantic middleware

Feature	Description
Interoperation	Interoperation shares information and uses the same across diverse domains of applications using diverse communication interfaces
Context detection	Context is responsible for characterizing the situation of an entity where an entity can be person, place, or object relevant to the interaction between a user and an application, including the user and applications themselves.
Security	Security and privacy are responsible for confidentiality, authenticity, and nonrepudiation.
Portability	Managing data volumes is an integral part of IoT-middleware. It is believed that there will be trillions of objects which will be part of this enormous network and hundreds of Exabytes will be stored or exchanged among the objects
Device discovery	Device discovery and management enables any device in the IoT network to detect all its neighbouring devices and make its presence known to each neighbour in the network. Device ontology is used for storing information about the heterogeneous devices.

Source: based on Bandyopadhyay et al. (2011).

INDUSTRIAL SENSORS

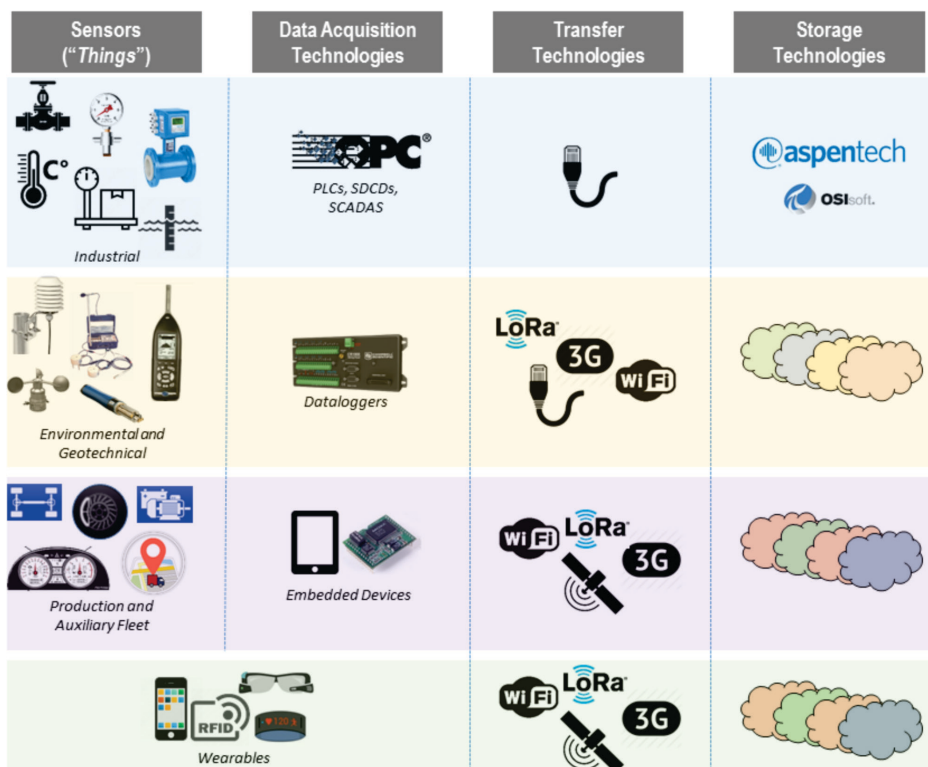
This study considers a middleware capable of describing the semantic for industrial sensors. Information Technology (IT) has been an enabling and driving force practically in all engineering domains including automation. Today, all subsystems of an automation system, from field level to operation level, has become IT enabled and driven. (SHARMA, 2017). While the term “Internet of Things” or IoT is relatively a new concept for connecting things, the basic idea of IoT is to connect all physical devices to collect their relevant data in real time to manage the “things” better and make “things” more reliable and predictive. (GUNASEKARAN et al, 2017).

Collecting data from industrial sensors is nothing new for industries. In a general way, data from industrial sensors are collected by programmable logic controllers (PLCs). Data from environmental and geotechnical sensors are usually collected by datalogger devices, positioned close to sensors.

Data from production and auxiliary fleet are retrieved by embedded devices, placed inside the fleet. Data from industrial wearables are collected from different ways. In the industrial environment, all data are routed to a supervisory software, used by operators to analyze data and make real-time plant-floor decisions. On top of this architecture, a time-series data historian system is frequently included with capabilities to store data from multiple sensors.

Softwares such as InfoPlus.21 (by AspenTech), PI System (by OSIsoft) among others, are used with this purpose. This kind of software is an industrial version of a key-value-pair (KVP) database, with specialized functions such as: data compression, ad-hoc calculations, statistics, among others. Figure 1 shows how data is collected from sensors in industrial environment.

Figure 1 – How data are collected in industrial environments



Source: Authors

However, it is difficult to identify characteristics of a specific sensor for those who are outside the engineering context, as they are not familiar to naming conventions adopted by each specific plant. This study focuses on industries with naming conventions for sensors, even when there are different versions for each plant. It is proposed an IoT middleware as an approach to deliver semantic for industrial sensors, automatically created based on naming conventions.

METHODS

The general idea of our approach is depicted in Figure 2. MINDSense (Middleware for Industrial Sensors) was designed to be accessible as an independent service and acts as an information provider not only about sensors and actuators but also on the existing relationship between machinery and sensors. Thus, analytics applications may acquire the necessary metadata to locate and interpret sensor readings while data scientists may use it as a common knowledge repository on the topic. Queries are made using the SPARQL language (Version 1.1), carried through the HTTP protocol in a REST API, while responses are encoded as JSON/XML messages. These architectural decisions aim to facilitate the system interoperability, particularly with legacy applications. The SPARQL engine interprets such queries seeking data on our ontology model stored in a triple database engine.

We have built MINDSense on top of Apache Jena¹, an open source Java framework for building semantic web applications. To express the knowledge we selected the OWL language, in the OWL DL profile which renders the maximum expressiveness possible while holding completeness and decidability, common prerequisites for descriptive logic based reasoners.

Sensor metadata collected in this study is derived from the naming patterns employed by three different port plants of a Brazilian mining company. A domain specialist yielded the necessary consulting during the process and was responsible for generating the initial data set from the company's internal database infrastructure (made available in CSV files). The data were extracted from Aspentech InfoPlus.21 database through a SQLPlus query specifically created for the purpose of this study. This preliminary data structure was used to feed MINDSense. The dataset is comprised of 42,583 rows where each row determines a sensor. Table 3 provides a sample of our data.

Table 3 – Sample of sensor metadata

Sensor tag	Description	IO Type	Unit
1GR04A_DV_HOR_01_D	Hourmeter of equipment 1GR04A	Analog	Hour
1PA0_M1_POT_01_D	Potentiometer for equipment 1PA0	Analog	Kilowatts
1GR04A_DV_DIAG_01_D	Diagnostic word	Discrete	--
EP03_LAN_ELEV_01_D	Current position - Elevation	Discrete	Millimeters

Source: Author

¹ Available at <https://jena.apache.org/index.html>

The naming conventions applied in column ‘Sensor tag’ is particularly relevant as it implicitly wraps at least six sensor characteristics. Splitting its value at the character ‘underline’ reveals the Equipment and Sub-Equipment a sensor monitors, its purpose, sequence and data source, respectively. The sensor tag at the first row, for instance, indicates that there is a sensor (1GR04A_DV_HOR_01_D) that monitors a piece of sub-equipment (DV) of a piece of equipment (1GR04A). It is a sensor read directly from an automation system (D) which indicates an hourmeter, that is, the amount of time the device is running (HOR) - the type of sensor is related to Table 1. It also informs that this sensor is the first of its type installed on the device (01). Metadata such as database connection, Port plant, organizational area and its corresponding labels, although available for processing, have been omitted from the table.

These naming patterns, though, are not strictly consistent throughout the organization. Considering our sample, a significative number of sensors (13,647 or 32.04% of the total) contained incomplete data or unstandardized values. Under these circumstances, the ability to make inferences on incomplete information while enabling structural revisions, data correction, and posterior knowledge enrichment are considered “must have” features what, on the other hand, poses modeling challenges for applications based on traditional relational databases. In this sense, the MINDSense approach takes advantage of the open-world assumption (OWA) typically made by semantic web languages such as the OWL language. The OWA states that no inferences can be drawn from statements that haven’t been made yet, what enables the work with incomplete information. To illustrate this concept, consider the following axiom: “Sensor X is a sensor of equipment Y”. If we enquire a typical RDBMS with “Is sensor X a sensor of equipment Z?”, the answer is “No”. The same question, nevertheless, would render a different answer when submitted to an OWA based system: “Unknown”.

RESULTS

Considering the common features of a semantic middleware, such as those mentioned in Table 2, the product generated in this study is interoperable, has context detection capabilities and enables device discovery. The simplified version of the ontology created for MINDSense is described in more details in Figure 3 that follows the notation produced by WebVOWL. (LOHMANN et al, 2014). The class “Sensor”, as expected, is a key element in MINDSense ontology and its individuals are associated to one of its subclasses according to its purpose (defect, elevation, status, and so on).

Sensors are responsible for monitoring Machines (instruments designed to transmit or modify the application of power, force or motion) that must be classified as an instance of equipment or sub-equipment. Equipment refers to machines that are directly controlled by a mining plant during the extraction or transport activities. Sub-equipment is a machine that acts as part of an instance of equipment. In this context, a crane is a piece of equipment while its oil pump is an instance of sub-equipment. Machines and Sensors are Spatial Things that, in turn, are allocated at places such as an areas and plants.

The application of the ontology in our sensors sample resulted in 19 classes, 516,859 axioms, 516,824 logical axioms, 28 declaration axioms, 9 object properties, 6 data type properties and 56,565 individuals.

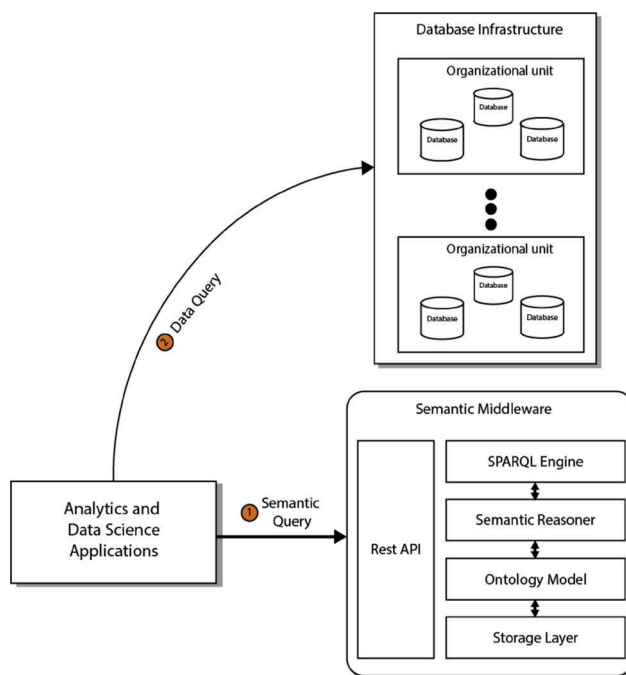
MINDSense² allowed stakeholders from outside the engineering context - including data scientists - to interact with meta data from industrial sensors, obtaining information that, until then, was restricted to domain experts of each organizational unit. The middleware usage is also useful for inducing storage standardization and term-alignment across the organization. These factors contribute to leverage advanced analytics in industries what, in turn, increases the potential of new findings on top of industrial sensors data.

² The source code of our semantic middleware is publicly available at <http://github.com/dfalci/semanticmiddleware>

Table 4 exemplifies some questions that MINDSense is naturally able to address. These questions are directly related to running example mentioned in the beginning of this paper.

MINDSense fulfills an existing lack in industrial plants as they do not have an automated data model capable of representing physical structures and relationships among industrial equipments. Our middleware also presents relationships between sensors and equipments despite the volume of industrial sensor data collected. Another advantage is that it creates a flexible model that could answer the requirements of each data consumer while creating a standardized model that could be shared across the enterprise. Depending on consumer requirements, different SPARQL queries could be written, answering specific questions but keeping the main data structure.

Figure 2 – MINDSense: a semantic middleware for industrial sensors



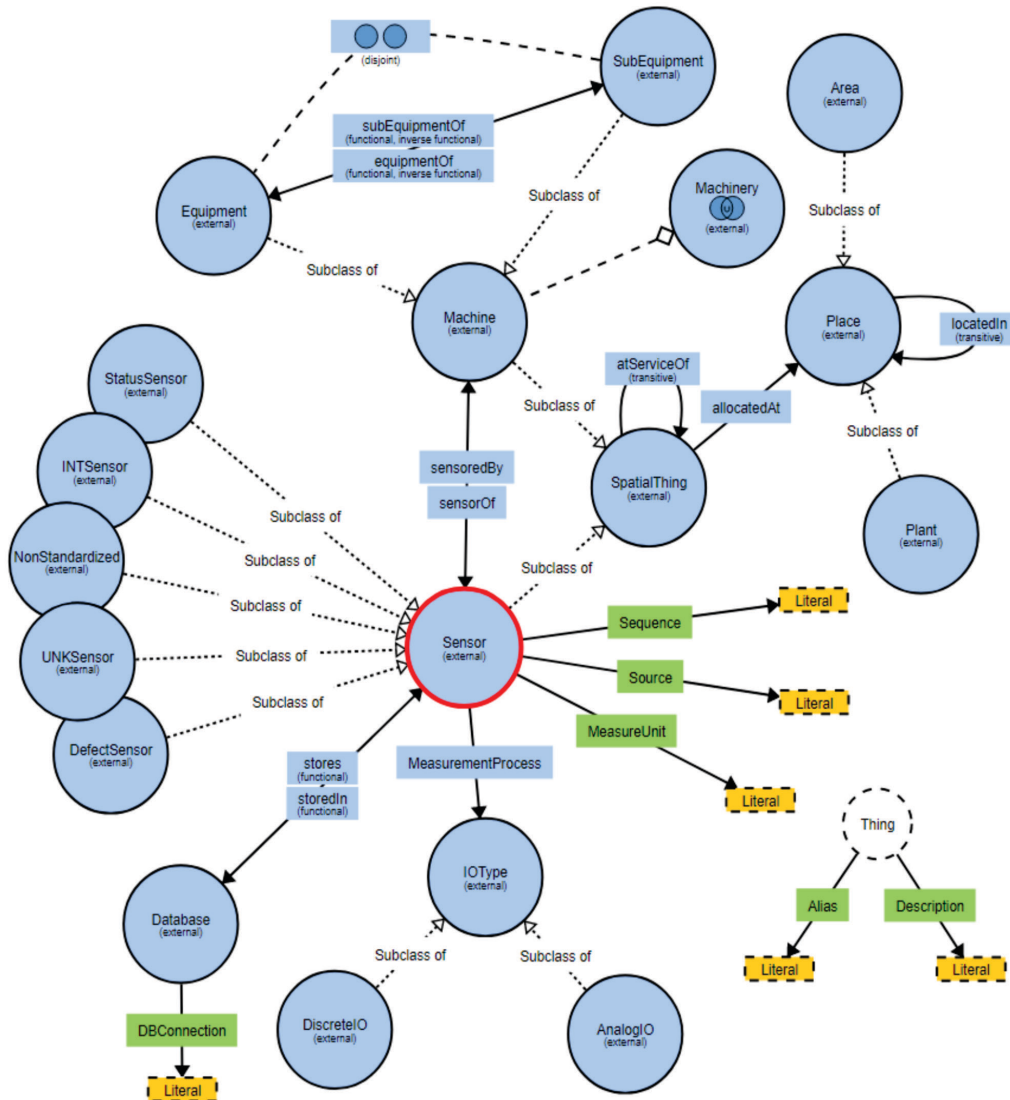
Source: Authors

Table 4 – Common questions and its SPARQL translations to MINDSense

Question	SPARQL query
In what databases one can find sensor readings from the sensor S?	<pre>SELECT ?db { ?x rdf:type/rdfs:subClassOf* sm:Sensor; sm:storedIn ?db. FILTER (?x = <sensor>) }</pre>
What are the sensors of equipment Z?	<pre>SELECT * WHERE { ?x sm:atServiceOf+ <equipment>. ?x rdf:type/rdfs:subClassOf* sm:Sensor. }</pre>
What are the name of the defect sensors of equipment Z across different port plants?	<pre>SELECT ?alias ?port WHERE { { SELECT ?d WHERE { ?d a sm:Equipment. Filter(?d = <equipment>) }} ?a rdf:type/rdfs:subClassOf* sm:DefectSensor; sm:Alias ?alias; sm:allocatedAt ?port; sm:atServiceOf ?t. FILTER (?t = ?d) }</pre>

Source: Authors

Figure 3 – Concepts and Relations of MINDSense Ontology



Source: Authors

RELATED WORK

This study is strongly related to Soignier (2017), as the author proposed a graph-based data model for industrial sensors data. The author references a product called ElementGraph which can be coupled to OsiSoft PI System. The main contribution of MINDSense is being able to implement a similar proposition capable of interacting with any industrial sensor database, including InfoPlus.21.

Diego, Martinez, Rodriguez-Molina, Cuerva (2014) and Maffei (2017) propose a semantic middleware for energy grid. The first approach considers features such as device discovery and context detection. The second approach includes those features and adds interoperability and portability features. Additionally, those studies consider specialized functions for the middleware like energy analysis and demand forecasting. Our work differs from theirs because we considered all applications, conceptually dissociated from the middleware layer, creating a more generic semantic middleware.

CONCLUSION

In this study, we presented MINDSense as a method that dynamically creates an ontology describing the meaning of industrial sensors and its relationships. We firstly presented the lack of semantics for industrial sensors as a problem to be solved. Then, we stated that the ability to work, making inferences on incomplete information, enabling revision, correction, and posterior knowledge enrichment are considered “must have” features, which poses modeling challenges for applications based on traditional relational databases. Next, we described MINDSense as a semantic middleware that could be accessible as an independent service and act as an information provider not only about sensors and actuators but also on the existing relationship between machinery and sensors.

MINDSense made possible to directly answer questions that were restricted to domain experts of each organizational unit and this contributes to leverage advanced analytics in industries and to increase the potential of new studies on top of industrial sensors data.

As a future work, our approach could be improved to include features such as security and portability. Considering we used open software tools, MINDSense can be modified by third party developers and further expanded. Also, other studies could implement our model in a real scenario, allowing the interaction of analytics tools with sensor data in real time.

REFERENCES

- AUSTERLITZ, H. *Data acquisition techniques using PCs*. San Diego: Academic Press, 2003.
- BANDYOPADHYAY, S. et al. Role of middleware for internet of things: a study. *International Journal of Computer Science & Engineering Survey*, v. 2, n. 3, p. 94-105, 2011.
- CHAQFEH, M. A.; MOHAMED, N. Challenges in middleware solutions for the internet of things. In: INTERNATIONAL CONFERENCE ON COLLABORATION TECHNOLOGIES AND SYSTEMS (CTS), 2012, Denver, CO, USA. *Proceedings...* 2012.
- CONSTANTIOU, I. D.; KALLINIKOS, J. New games, new rules: big data and the changing context of strategy. *Journal of Information Technology*, v. 30, n. 1, p. 44-57, 2015.
- DIEGO, R. D. et al. A semantic middleware architecture focused on data and heterogeneity management within the Smart Grid. *Energies*, v. 7, n. 9, p. 5953-5994, 2014.
- EMILIO, M. D. *Data acquisition systems: from fundamentals to applied design*. New York: Springer, 2013.
- GUNASEKARAN, A. et al. Big data and predictive analytics for supply chain and organizational performance. *Journal of Business Research*, v. 70, p. 308-317, 2017.
- HURWITZ, J. S.; KAUFMAN, M.; BOWLES, A. *Cognitive computing and big data analytics*. Indianapolis: Wiley, 2015.
- LIU, J et al. Rethinking big data: a review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 115, p. 134-142, 2016.
- LOHMANN, S. et al. WebVOWL: Web-based visualization of ontologies. *International Conference on Knowledge Engineering and Knowledge Management*, p. 154-158, 2014.
- MAFFEI, A. et al. A Semantic middleware supported receding horizon optimal power flow in energy grids. *IEEE Transactions on Industrial Informatics*, 2017.
- SHARMA, K. L. *Overview of industrial process automation*. Amsterdam: Elsevier, 2017.
- SOIGNIER, A. *Using asset data models to empower your industrial organization*. 1982. Disponível em: <<http://goo.gl/ApQQpg>>. Acesso em: 05 jun. 2017.
- STIMMEL, C. L. “*Building smart cities: analytics, ICT, and design thinking*”. 2015.
- VERA-BAQUERO, A.; COLOMO-PALACIOS, R.; MOLLOY, O. Real-time business activity monitoring and analysis of process performance on big-data domains. *Telematics and Informatics*, v. 0, n. 3, p. 793-807, 2016.
- XU, L. D.; HE, W.; LI, S. Internet of things in industries: a survey. *IEEE Transactions on Industrial Informatics*, v. 10, n. 4, p. 2233-2243, 2014.

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.; Josbäcker, F.; Kimm, N.; Kradeppohl, 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

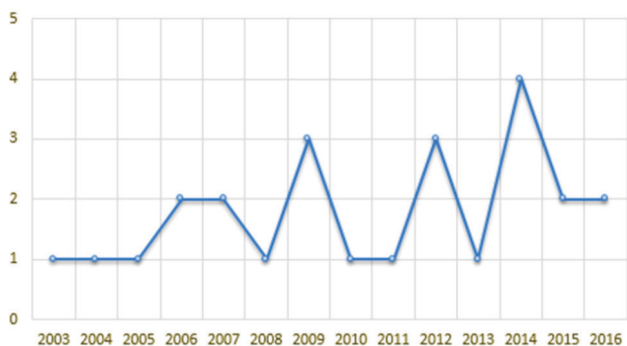


Figura 2 – Diagrama de extremos e quartis da distribuição temporal

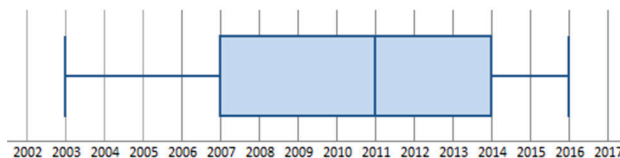


Figura 3 – Distribuição do corpus de análise atendendo ao país associado à afiliação dos autores

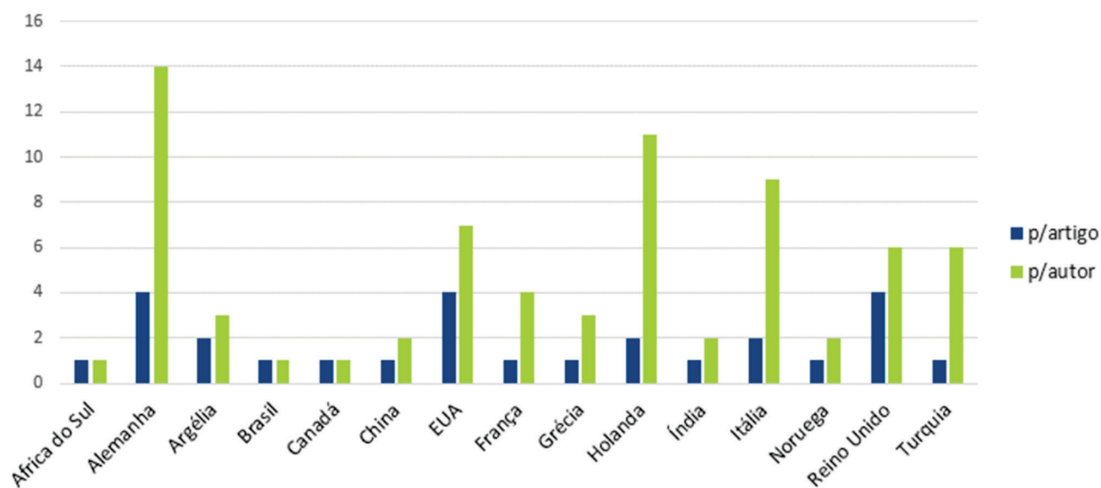


Figura 4 – Distribuição percentual do tipo de abordagem identificada nos 25 artigos

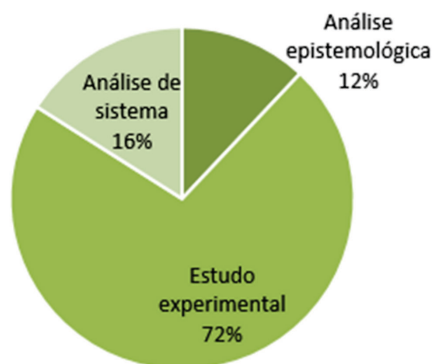


Figura 5 – Distribuição percentual, por nível de aproximação, da abordagem temática aos dois conceitos operacionais (indexação automática e ontologias)

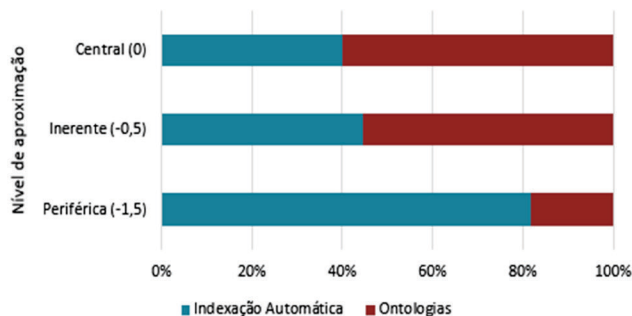
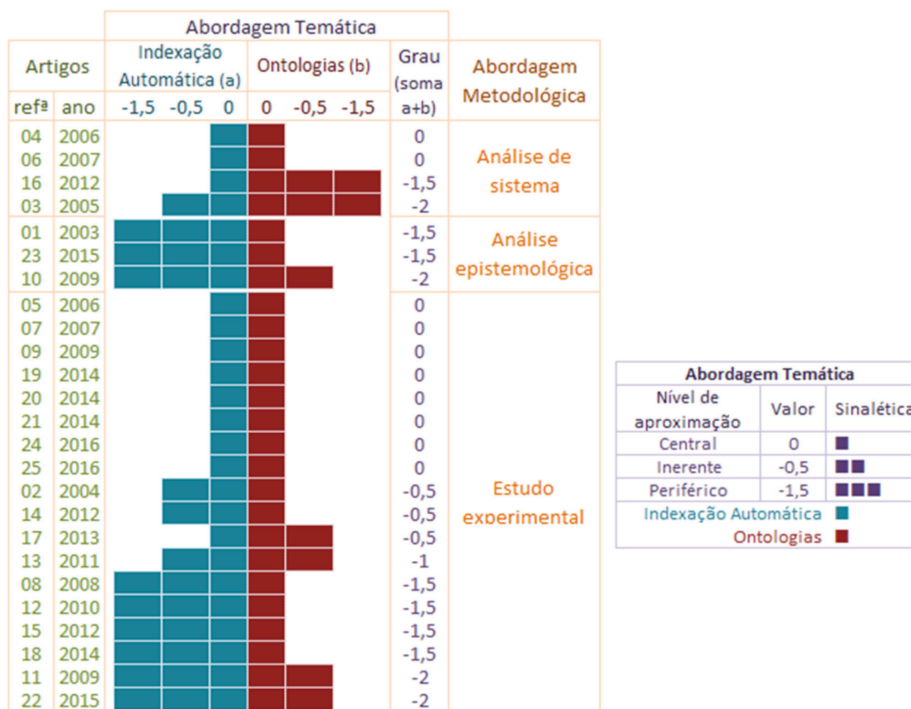


Figura 6 – Comparação entre a abordagem temática e a abordagem metodológica identificadas nos 25 dos artigos



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.

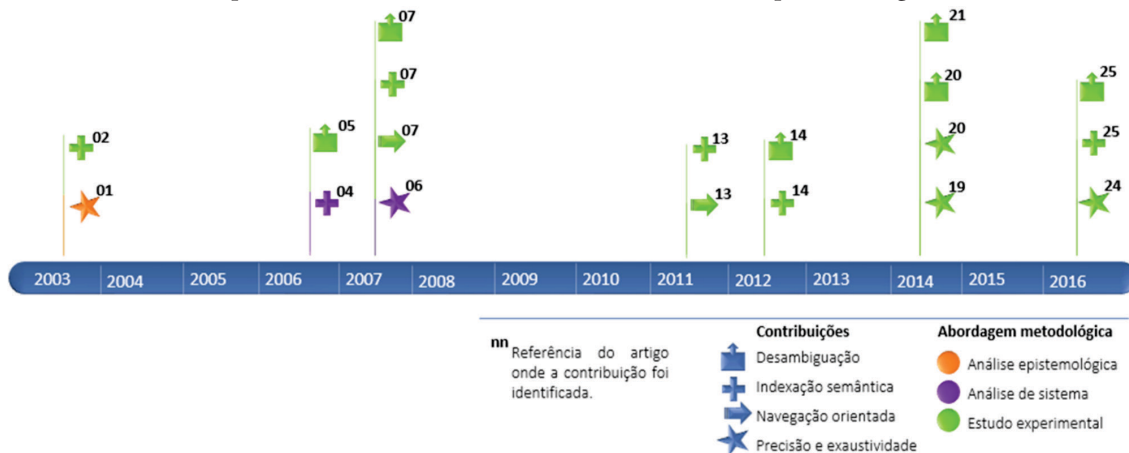
Por fim, a figura 7 apresenta a distribuição temporal das contribuições identificadas, de acordo com o ano de publicação dos artigos de onde foram coligidas. No referido esquema é possível constatar uma dispersão no tempo das várias contribuições, não se verificando, assim, concentração de qualquer tipologia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

APÊNCICE 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.
- CHIARAVALLOTTI, 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

Aplicação e uso de ontologias musicais

Lígia Maria Arruda Café

Doutora em Linguística, Université Laval (ULaval), Québec, QC, Canadá.

Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Santa Catarina (PGCIN/UFSC), Florianópolis, SC, Brasil.

Bolsista de Produtividade em Pesquisa (PQ/CNPQ).

<http://lattes.cnpq.br/9103205400963118>

ligia.cafe@ufsc.br

Camila Monteiro de Barros

Doutora em Ciência da Informação, Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, Brasil.

Departamento de Ciência da Informação, Universidade Federal de Santa Catarina (CIN/UFSC), Florianópolis, SC, Brasil.

<http://lattes.cnpq.br/1818999454664723>

camila.c.m.b@ufsc.br

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

RESUMO

As ontologias são instrumentos fundamentais para a interoperabilidade no contexto da Web Semântica, pois estão fundamentadas na descrição explícita das relações semânticas em domínios de conhecimento. O domínio da música, especificamente, apresenta conceitos implicados em dimensões de naturezas distintas, como sonora, bibliográfica, cultural (incluindo emoção, gênero musical, informações de redes sociais), de produção, de performance, etc. Este estudo, de caráter exploratório, procura mostrar as várias experiências relacionadas ao uso das ontologias musicais no mundo registradas na literatura científica internacional. Para tanto, realizou-se um levantamento nas bases LISA e Scopus e também em trabalhos oriundos das conferências do ISMIR. O corpus total é formado por 102 documentos, e a análise qualitativa ocorreu por meio do resumo dos documentos. Os resultados mostram cinco tipos de usos e aplicações das ontologias: 1) âmbito comercial; 2) compartilhamento da representação conceitual sem mencionar aplicação específica; 3) recomendação de músicas; 4) recuperação da informação; 5) outros. Conclui-se que para cada uso e aplicação pretendidos para determinada ontologia musical, certas dimensões da música são mais ou menos relevantes. Nesse sentido, do ponto de vista da interconexão de dados no contexto da Web Semântica, o que parece ser um desafio para o futuro das pesquisas em ontologias musicais é a interação, de fato, entre as distintas propostas de modo a cobrir, ainda que não completamente, o heterogêneo universo conceitual do domínio da música.

Palavras-chave: Sistema de Organização do Conhecimento. Ontologia. Música.

Application and use of music ontologies

ABSTRACT

Ontologies are fundamental tools for interoperability in the context of the Semantic Web, since they are based on the explicit description of semantic relations in knowledge domains. The domain of music, specifically, presents concepts implied in dimensions of different natures, such as sound, bibliographical, cultural (including emotion, musical genre, information of social networks), production, performance, etc. This exploratory study seeks to show the various experiences related to the use of musical ontologies in the world registered in the international scientific literature. To do so, we conducted a survey on the LISA and Scopus databases and also on papers from the ISMIR conferences. The total corpus consists of 102 documents and the qualitative analysis took place through a summary of the documents. The results show five types of uses and applications of the ontologies: 1) commercial scope; 2) sharing conceptual representation without mentioning specific application; 3) recommendation of songs; 4) information retrieval; 5) others. We conclude that for each intended use and application for a given musical ontology, certain dimensions of music are more or less relevant. In this sense, from the point of view of data interconnection in the context of the semantic web, what seems to be a challenge for the future of research on musical ontologies is the interaction, in fact, between the different proposals in order to cover, although not completely, the heterogeneous conceptual universe of the domain of music.

Keywords: Knowledge Organization System. Ontology. Music.

Aplicación y uso de ontologías musicales

RESUMEN

Las ontologías son instrumentos fundamentales para la interoperabilidad en el contexto de la Web Semántica, pues están fundamentadas en la descripción explícita de las relaciones semánticas en dominios de conocimiento. El dominio de la música, específicamente, presenta conceptos implicados en dimensiones de naturalezas distintas, como sonora, bibliográfica, cultural (incluyendo emoción, género musical, informaciones de redes sociales), de producción, de rendimiento, etc. Este estudio, de carácter exploratorio, busca mostrar las varias experiencias relacionadas al uso de las ontologías musicales en el mundo registradas en la literatura científica internacional. Para tanto, realizamos un levantamiento en las bases LISA y Scopus y también en trabajos oriundos de las conferencias del ISMIR. El corpus total está formado por 102 documentos y el análisis cualitativo ha ocurrido por medio del resumen de los documentos. Los resultados muestran cinco tipos de usos y aplicaciones de las ontologías: 1) ámbito comercial; 2) compartir la representación conceptual sin mencionar aplicación específica; 3) recomendación de músicas; 4) recuperación de la información; 5) otros. Concluimos que, para cada uso y aplicación pretendidos para una determinada ontología musical, ciertas dimensiones de la música son más o menos relevantes. En este sentido, desde el punto de vista de la interconexión de datos en el contexto de la web semántica, lo que parece ser un desafío para el futuro de las investigaciones en ontologías musicales es la interacción, de hecho, entre las distintas propuestas para cubrir, aunque no completamente, el heterogéneo universo conceptual del dominio de la música.

Palabras clave: Sistema de Organización del Conocimiento. Ontología. Música.

INTRODUÇÃO

As tecnologias relacionadas à Web Semântica são soluções que favorecem a estruturação de dados a fim de promover a interligação entre recursos de informação (W3C, 2017), constituindo as práticas de *linked data* (BIZER; HEATH; BERNERS-LEE, 2009). Essa estrutura de dados, legível por máquina, é fundamental no desenvolvimento e gestão de redes que envolvem grandes quantidades de informação, já que oferece suporte para seu processamento automático.

As ontologias são instrumentos fundamentais para a interoperabilidade no contexto da Web Semântica, pois estão fundamentadas na descrição explícita das relações semânticas em domínios de conhecimento. A explicitação torna possível o processamento automático de informações, diferentemente da informação não estruturada, legível por humanos, em que as relações semânticas estão implícitas (MOTTA; SIQUEIRA; ANDREATTA, 2009). A estrutura de uma ontologia traz conceitos definidos, relações conceituais e suas associações a instâncias (LACY, 2005), i. e., ocorrências de fatos de acordo com os princípios classificatórios definidos na ontologia. De acordo com Gruber (1993), ontologia é a especificação formal de uma conceitualização, legível por máquina.

São variadas as ferramentas para desenvolvimento de ontologias; apresentaremos algumas mais amplamente utilizadas, de acordo com o World Wide Web Consortium¹ (W3C). A linguagem Ontology Web Language (OWL) tem sua sintaxe representada com base em eXtensible Markup Language (XML), o W3C também recomenda o uso do Resource Description Framework (RDF) (linguagem que especifica instâncias da OWL), entre outros padrões de desenvolvimento Web.

Para Gruber (1993), a construção de ontologias também deve visar sua reutilização em diferentes aplicações. A portabilidade das ontologias está relacionada ao tipo de formalismo utilizado na definição do sistema e da linguagem utilizada na

representação. Esses instrumentos também podem ser classificados quanto a sua função, aplicação e estrutura como ontologia de domínio, ontologia de tarefa, etc. (ALMEIDA; BAX, 2003). A ontologia dá o suporte para que computadores realizem inferências lógicas semanticamente relevantes e em distintas aplicações (W3C, 2017b), além de minimizar os problemas de polissemia e sinonímia dos sistemas baseados na sintaxe dos termos (RODRÍGUEZ-GARCÍA et al., 2015).

Para que de fato seja um instrumento eficiente, a conceitualização proposta em uma ontologia precisa representar adequadamente o domínio de conhecimento a que se refere; sua estruturação deve ser, portanto, contextualmente orientada (SMIRAGLA, 2015). Nesse sentido, as relações ali explicitadas envolvem a definição do lugar que cada conceito ocupa no domínio representado (BRASCHER, CAFÉ, 2010). Ora, cada domínio de conhecimento apresenta uma dinâmica particular, além disso, a visão que se tem desse domínio também vai variar dependendo do objetivo para o qual a ontologia está sendo desenvolvida. Por exemplo: uma ontologia da área da música, que dá suporte a recomendações de músicas em um *site* comercial, é construída com base em um modelo de abstração conceitual diferente da ontologia que recomenda músicas de acordo com as relações sociais que uma pessoa estabelece na Web. Isso ocorre porque a própria abstração conceitual que resulta do processo de significação social da música varia, conferindo à noção de “domínio de conhecimento” um estatuto cultural complexo que influencia suas formas de representação.

O domínio da música, especificamente, apresenta conceitos implicados em dimensões de naturezas distintas, como sonora, bibliográfica, cultural (incluindo emoção, gênero musical, informações de redes sociais), de produção, de performance, etc. The Music Ontology² é uma ontologia que se propõe a representar, com alto nível de generalidade, conceitos e relações do domínio da música referentes a todas essas dimensões. De

¹ Disponível em: <<https://www.w3.org>>.

² Disponível em: <<http://musicontology.com/>>.

acordo com Raimond et al. (2007), a ontologia é desenvolvida com base em outras ontologias, como a *Friend-of-a-friend* (FOAF) para a definição das classes “pessoa” e “grupo”, e a Event ontology para a descrição de eventos como performances envolvendo músicos e seus instrumentos. Adota também os conceitos de obra, manifestação, item e expressão definidos pelo Functional Requirements for Bibliographic Records (FRBR), além de estar conectada a outras fontes de dados. The Music Ontology foi idealizada com o intuito de contribuir para o movimento do *open linked data*, sendo extensível a domínios mais específicos, e já foi adotada em diversas aplicações (RAIMOND et al., 2007). Recursos que também disponibilizam seus dados para interconexão por meio de ontologias são o MusicBrainz³, Wikipedia⁴, entre outros. Além da The Music Ontology, existem outras ontologias de música como a Musical Performance Ontology (SÉBASTIEN; SÉBASTIEN; CONRUYT, 2013) e a Studio Ontology Framework (FAZEKAS; SANDLER, 2011).

Alguns exemplos dos possíveis usos de ontologias tanto na música quanto em outros domínios são: a) aumentar a eficiência na recuperação da informação quando, por exemplo, o sistema de busca baseado em ontologia realiza correspondências semânticas entre termos e entre fontes, expandindo a busca do usuário e/ou os termos de indexação (FERNEDA; DIAS, 2017, SOUZA, 2015); b) interconectar fontes de dados heterogêneas, como anotações manuais, diferentes formatos de metadados e informações extraídas automaticamente do recurso de informação (RAIMOND et al., 2007); c) estabelecer perfis de usuários, relacionando-os a recursos de informação e permitindo aplicações de recomendação de informações (KIM, 2013), tanto para descoberta de novas músicas, quanto para manutenção de certo gosto musical. Tais usos podem estar relacionados a aplicações de cunho comercial ou não, em contextos públicos ou particulares.

Este estudo, de caráter exploratório, procura mostrar as várias experiências relacionadas ao uso das ontologias musicais no mundo registradas na literatura científica internacional.

METODOLOGIA

O levantamento bibliográfico foi realizado nas bases Library and Information Science (LISA) e Scopus, por meio da busca “music AND ontology” no resumo, sem recorte temporal, com a exclusão de áreas não afins como bioquímica, medicina, neurociência, resultando em 132 documentos. Os documentos foram avaliados qualitativamente quanto à pertinência das temáticas, tendo sido excluídos 39 artigos. Também foram levantados trabalhos oriundos das conferências do International Society of Music Information Retrieval (ISMIR), que continham o termo “ontology” no título, resultando em nove trabalhos. O *corpus* total é formado por 102 documentos.

A análise qualitativa ocorreu por meio do resumo dos documentos em que foram mapeados os usos e aplicações de ontologias musicais. Analisamos tanto o que se refere ao desenvolvimento de ontologias específicas para o domínio da música quanto aquelas aplicações mais gerais que tiveram sua eficiência testada no âmbito da informação musical.

³ Disponível em: <<https://musicbrainz.org/>>

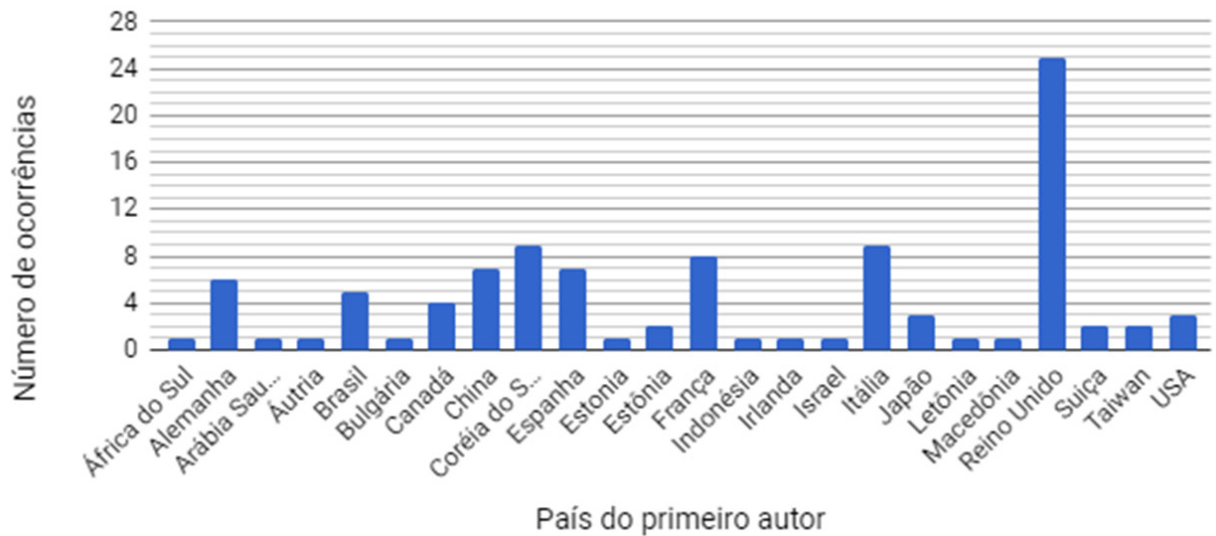
⁴ Disponível em: <<https://www.wikipedia.org/>>

Figura 1- Número de publicações por ano



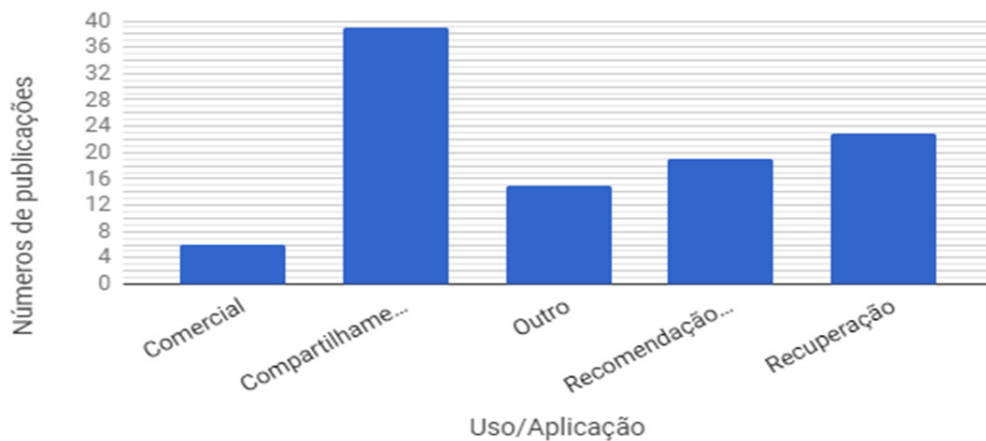
Fonte: Dados da pesquisa, 2017.

Figura 2- Número de publicações por país do primeiro autor



Fonte: Dados da pesquisa, 2017.

Figura 3- Usos e aplicações das ontologias musicais



Fonte: Dados da pesquisa, 2017.

RESULTADOS E DISCUSSÃO

O panorama geral do *corpus* analisado mostra publicações no período de 1996 a 2016, sendo os maiores números nos anos de 2013 e 2016 (figura 1). Observa-se também uma evolução gradual na quantidade de artigos a partir de 2006.

Com relação ao país do primeiro autor, a figura 2 mostra destaque para o Reino Unido. O Brasil, ainda que apresente tímida participação, registra uma publicação maior que vários países.

Wilmering, T. e Allik, A. aparecem como primeiro autor de quatro publicações cada, sendo os nomes com maior ocorrência. Ambos os autores estão filiados à Queen Mary Universidade de Londres (Reino Unido). Aparecendo como primeiro autor em três publicações estão Albuquerque, M., Rho, S. e Wang, J, respectivamente filiados à Universidade Federal do Rio de Janeiro (Brasil), Universidade de Sungkyul (Coreia do Sul) e Universidade de Pequim (China).

Os usos e aplicações das ontologias foram agrupados em cinco tipos (figura 3): 1) âmbito comercial; 2) compartilhamento da representação conceitual sem mencionar aplicação específica; 3) recomendação de músicas; 4) recuperação da informação; 5) outros. Esses tipos são intercambiáveis, ou seja, uma ontologia desenvolvida com a finalidade de compartilhamento de vocabulário pode ser utilizada em uma aplicação comercial. No entanto, a classificação foi realizada de acordo com os resumos nos quais informações apresentavam níveis distintos de especificidade.

Com relação aos estudos voltados ao âmbito comercial da música, os autores sugerem ontologias que mapeiam os atores da cadeia de negócios voltados à música em ambiente Web peer-to-peer (ex. RUPP; ESTIER, 2003) a fim de reduzir o tempo gasto pelo cliente em busca de serviços (ex. POON, 2011), também é mapeado o fluxo de trabalho desde a produção até o consumo com o intuito de aprimorar a comunicação entre pessoas e setores da indústria da música (ex. SCHUMACHER; GEY; KLINGNERE, 2014).

As pesquisas que visam compartilhamento concentram o maior volume de publicações e trazem representações do domínio da música em diferentes aspectos, como os intervalos tonais (SOMMARUGA; BAGGI, 2009) e a estrutura harmônica (WEYDE, WISSMANN, 2007). Outros estudos objetivam investigar a correspondência entre formatos de metadados por meio de ontologias (ex. TIAN et al., 2013, SOHN; KIM; CHUNG, 2012) ou ainda mapear automaticamente relações do tipo “is a” para um novo termo na Web (ex. ESPINOSA-ANKE, 2016). Trata-se basicamente de ontologias que oferecem uma estrutura conceitual para formatos de descrição de dados (como *tags* e outros), visando compartilhamento de vocabulários que possibilitem a realização da rede de *linked data*, como é o caso da The Music Ontology.

No que se refere às aplicações para recuperação da informação, as pesquisas apresentam ontologias para classificação da música baseada no seu contexto, combinada a uma atualização do perfil de busca do usuário (ex. REFORMAT; GOLMOHAMMADI, 2009). A adaptação à OWL de metadados MPEG7 para incorporação em uma ontologia (GARCÍA; CELMA, 2005) também é apontada como solução para eficiência na recuperação da informação. Ontologias também são apresentadas para conversão de dados em formato MARC para RDF visando interligar bibliotecas de música (LISENA et al., 2016).

O uso de ontologias em aplicações de recomendação de músicas inclui, geralmente, características do perfil do usuário, sejam hábitos de escuta (ex. KIM; KIM, 2014) ou de busca em diferentes mecanismos (ex. MOCHOLI et al., 2012). As ontologias também apoiam a análise de co-ocorrências de *tags* de gênero musical (ex. SCHREIBER, 2016) ou ainda da semântica das *tags* para construção de outras ontologias, como a que descreve emoções relacionadas à música (KIM, 2013). No último caso, essas ontologias lidam com um aspecto de alta complexidade: a dimensão emocional e cultural do domínio da música. Elas devem ser capazes de registrar conceitos em linguagem verbal que representem a experiência simbólica baseada na melodia.

A classe “outros” engloba os estudos que não especificam uso ou aplicação como o desenvolvimento de ontologia para calcular a consistência semântica de outras ontologias (BARINSKIS; BARZDINS, 2007), descrição para arquivamento em longo prazo de trabalhos acusmáticos (ESPOSITO; GESLIN, 2008) e descrição de emoções para uso em qualquer aplicação (KIM; KWON, 2011, WANG et al., 2010).

Vale ressaltar que nem todos os estudos aqui analisados tinham especificamente a música como foco. Em alguns casos, o domínio da música figurou apenas como âmbito de teste da aplicação proposta (ex. PEDRINACI et al., 2005; PYSHKIN; KUZNETSOV, 2015; ACHICHI et al., 2016).

CONCLUSÃO

A música é uma linguagem expressiva e não referencial (SVENONIUS, 1994), assim, para cada uso e aplicação pretendidos para determinada ontologia musical, certas dimensões da música são mais ou menos relevantes. A complexidade de se construir uma ontologia neste domínio fica evidente, já que exige a interação entre essas diferentes dimensões.

Este estudo mostra que as pesquisas até aqui realizadas sobre ontologias musicais se aproximam muito dessa tentativa de lidar com variadas características da música: tanto aquelas provindas da produção musical, como da percepção dos usuários e do próprio som. De fato, as ontologias são instrumentos que permitem que tais práticas sejam exploradas, como foi possível notar nas aplicações de recomendação de música que precisam ponderar definições de gêneros musicais, hábitos de escuta e percepção dos usuários. Também as ontologias que visam oferecer representações do domínio da música podem atuar neste sentido, favorecendo o compartilhamento de uma linguagem comum referente a determinados aspectos da informação musical.

Do ponto de vista da interconexão de dados no contexto da Web Semântica, o que parece ser um desafio para o futuro das pesquisas em ontologias musicais é a interação, de fato, entre as distintas propostas de modo a cobrir, ainda que não completamente, o heterogêneo universo conceitual do domínio da música.

REFERÊNCIAS

- ALMEIDA, M.B.; BAX, M.P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, Brasília, v. 32, n. 3, p. 7-20, set./dez. 2003. Disponível em: <<http://www.scielo.br/pdf/ci/v32n3/19019.pdf>>. Acesso em: 30 ago. 2015.
- BIZER, C; HEATH, T.; BERNERS-LEE, T. Linked data: the story so far. *International Journal on Semantic Web and Information Systems*, v. 5, n. 3, 2009. Disponível em: <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>>. Acesso em: 28 ago. 2011.
- BRÄSCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? In: LARA, M.L.G. de; SMIT, J.W. (Org.) *Temas de Pesquisa em Ciência da Informação no Brasil*. São Paulo: Escola de Comunicação e Artes/USP, 2010. p. 87-103.
- FAZEKAS, G.; SANDLER, M.B. Novel methods in information management for advanced audio workflows. In: INTERNATIONAL CONFERENCE ON DIGITAL AUDIO EFFECTS, 12., DAFx 2009, sep. 1-4, 2009, Como, Italy. *Proceedings...* Como, Italy: ISPG, 2009. p. 464-470.
- FERNEDA, E.; DIAS, G.A. OntoSmart: um modelo de recuperação de informação baseado em ontologia. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 22, n. 2, p.170-187, abr./jun. 2017. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2081/1882>>. Acesso em: 07 jul. 2017.
- GRUBER, T.R. A Translation approach to portable ontology specifications. *Appeared in Knowledge Acquisition*, v. 5, n. 2, p. 199-220, 1993. Disponível em: <<http://tomgruber.org/writing/ontolingua-kaj-1993.pdf>>. Acesso em: 30 jul. 2015.
- KIM, H.H. A semantically enhanced tag-based music recommendation using emotion ontology. In: SELAMAT, A. et al. (Ed.). *Asian Conference on Intelligent Information and Database Systems (ACIIDS)*, 2013, Part II, LNAI 7803, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), p. 119-128, 2013.

MOTTA, E.; SIQUEIRA, S.; ANDREATTA, A. Populating a domain ontology from a web biographical dictionary of music: an unsupervised rule-based method to handle Brazilian Portuguese texts. In: INTERNATIONAL CONFERENCE ON WEB INFORMATION SYSTEMS AND TECHNOLOGIES, 5., mar. 23-26, 2009, Lisbon, Portugal. *Proceedings...* Portugal: Springer, 2009. p. 192-199.

RAIMOND, Y. et al. The music ontology. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL, 8., 2007. *Proceedings...* Austrian Computer Society, 2007. Disponível em: <http://ismir2007.ismir.net/proceedings/ISMIR2007_p417_raimond.pdf>. Acesso em: 10 ago. 2015.

RODRÍGUEZ-GARCÍA, M.Á. et al. Ontology-based music recommender system. In: INTERNATIONAL CONFERENCE ON DISTRIBUTED COMPUTING AND ARTIFICIAL INTELLIGENCE, 12., v. 373, 2015. *Proceedings...* 2015. p. 39-46. (Series Advances in Intelligent Systems and Computing). Disponível em: <http://link.springer.com/chapter/10.1007/978-3-319-19638-1_5>. Acesso em: 7 jul. 2017.

SEBASTIEN, V.; SEBASTIEN, D.; CONRUYT, N. An ontology for musical performances analysis: application to a collaborative platform dedicated to instrumental practice. In: INTERNATIONAL CONFERENCE ON INTERNET AND WEB APPLICATIONS AND SERVICES, 5., 2010, Barcelona, Spain. *Proceedings...* Barcelona, Spain: IEEE, 2010. p. 538-543.

SMIRAGLIA, R. *Domain analysis for knowledge organization: tools for ontology extraction*. Elsevier, 2015.

SOUSA, E.E. de. Uso de ontologia para recuperação da informação disponibilizada em vídeos por meio de indexação multimodal. *Revista Ibero-americana de Ciência da Informação*, Brasília, v. 8, n. 2, p. 293, 2015. Disponível em: <<http://periodicos.unb.br/index.php/RICI/article/view/15869>>. Acesso em: 7 jul. 2017.

WORLD WIDE WEB CONSORTIUM (W3C). Disponível em: <<https://www.w3.org>>. Acesso em: 7 jul. 2017.

REFERÊNCIAS DO CORPUS DE PESQUISA

ACHICHI, M. et al. Automatic key selection for data linking. In: KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT INTERNATIONAL CONFERENCE, 20., 2016, nov. 19-23, 2016, Bologna, Italy. *Proceedings...* Bologna, Italy: Springer International Publishing, 2016. p. 3-18. Disponível em: <http://link.springer.com/chapter/10.1007/978-3-319-49004-5_1>. Acesso em: 7 jul. 2017.

ADRIAN, B. et al. *Epiphany: adaptable RDFa generation linking the web of documents to the web of data*. Knowledge Engineering and Management by the Masses, p. 178-192, 2010. Disponível em: <<http://www.springerlink.com/index/c7716g3nk1727682.pdf>>. Acesso em: 7 jul. 2017.

ALBUQUERQUE, M. de O. et al. *Cataloguing and searching musical sound recordings in an Ontology-Based Information System*. In: IGI Global. *Governance, communication, and innovation in a knowledge intensive society*. EUA: IGI Global, 2011. p. 292-306. cap. 24.

_____; SIQUEIRA, S.W.M.; BRAZ, M.H.L.B. A conceptualization of music sound recordings and its representation in an ontology. *International Journal of Knowledge Society Research (IJKSR)*, v. 4, n. 1, p. 73-89, 2013. Disponível em: <<http://www.igi-global.com/article/a-conceptualization-of-music-sound-recordings-and-its-representation-in-an-ontology/84993>>. Acesso em: 7 jul. 2017.

ALLIK, A. et al. A shared vocabulary for audio features. In: EXTENDED SEMANTIC WEB CONFERENCE: THE SEMANTIC WEB, 2013 (Satellite Events). Berlin, Heidelberg: Springer, 2013. p. 285-286. *Series Lecture Notes in Computer Science*. v. 7955. Disponível em: <http://link.springer.com/chapter/10.1007/978-3-642-41242-4_44>. Acesso em: 07 jul. 2017.

_____. et al. Facilitating music information research with shared open vocabularies. In: EXTENDED SEMANTIC WEB CONFERENCE: THE SEMANTIC WEB, 2013 (Satellite Events). Berlin, Heidelberg: Springer, 2013. p. 178-183. *Series Lecture Notes in Computer Science*, v. 7955. Disponível em: <http://link.springer.com/chapter/10.1007/978-3-642-41242-4_20>. Acesso em: 7 jul. 2017.

_____; FAZEKAS, G.; SANDLER, M.B. An ontology for audio features. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE, 17., aug. 7-11, 2016, New York City, USA. *Proceedings...* ISMIR Conference: New York City, USA 2016. p. 73-79.

_____. Ontological representation of audio features. In: INTERNATIONAL SEMANTIC WEB CONFERENCE: THE SEMANTIC WEB, 2016. p. 3-11. *Series Lecture Notes in Computer Science*, v. 9982. Disponível em: <http://link.springer.com/chapter/10.1007/978-3-319-46547-0_1>. Acesso em: 7 jul. 2017.

- BALABAN, M. The music structures approach to knowledge representation for music processing. *Computer Music Journal*, v. 20, n. 2, p. 96-111, 1996. Disponível em: <<http://www.jstor.org/stable/3681334>>. Acesso em: 7 jul. 2017.
- BARBERA, M. et al. Building a semantic web digital library for the municipality of Milan. In: RETHINKING ELECTRONIC PUBLISHING: INNOVATION IN COMMUNICATION PARADIGMS AND TECHNOLOGIES, 13., 2009. *Proceedings...* International Conference on Electronic Publishing, 2009. p. 133-154. Disponível em: <http://elpub.scix.net/data/works/att/156_elpub2009.content.pdf>. Acesso em: 7 jul. 2017.
- BARINSKIS, M.; BARZDINS, G. Satisfiability model visualization plugin for deep consistency checking of OWL ontologies. In: OWL: EXPERIENCES AND DIRECTIONS (OWLED), 3., jun. 6-7, 2007, Innsbruck, Austria. *Proceedings...* Innsbruck, Austria: OWLED, 2007. p. 524. Disponível em: <<http://www.academia.edu/download/42350257/paper40.pdf>>. Acesso em: 7 jul. 2017.
- BAUMANN, S.; DE ROSNAY, M.D. Music life cycle support through ontologies. In: INTERNATIONAL FLORIDA ARTIFICIAL INTELLIGENCE RESEARCH SOCIETY CONFERENCE, 15., may 14-16, 2002, Pensacola Beach, Florida, USA. *Proceedings...* Pensacola Beach, Florida, USA: AAAI Press, 2002. p. 27-32. Disponível em: <<https://ocs.aaai.org/Papers/FLAIRS/2004/Flairs04-009.pdf>>. Acesso em: 7 jul. 2017.
- BAUMANN, S.; KLUTER, A.; NORLIEN, M. Using natural language input and audio analysis for a human-oriented MIR system. In: WEB DELIVERING OF MUSIC, dec. 11, 2002, Darmstadt, Germany. *Proceedings...* Darmstadt, Germany: IEEE, 2002. p. 74-81. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/1176196/>>. Acesso em: 1 jul. 2017.
- BROWN, S. et al. Building listening experience linked data through crowd-sourcing and reuse of library data. In: INTERNATIONAL WORKSHOP ON DIGITAL LIBRARIES FOR MUSICOLOGY, 1., 2014. *Proceedings...* ACM, 2014. p. 1-8. Disponível em: <<http://dl.acm.org/citation.cfm?id=2660172>>. Acesso em: 1 jul. 2017.
- BURRED, J.J. et al. Audio content analysis. In: KOMPATSIARIS, Y.; HOBSON, P.(Ed.). *Semantic multimedia and ontologies: theory and applications*, 2008. p. 123-162.
- CANO, P. et al. Mtg-db: a repository for music audio processing. In: WEB DELIVERING OF MUSIC, 4., 2004, Barcelona, Spain. *Proceedings...* Barcelona, Spain: IEEE, 2004. p. 2-9. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/1358093/>>. Acesso em: 1 jul. 2017.
- CASTANO, S.; FERRARA, A.; MONTANELLI, S. Ontology-based classification and retrieval of music resources. In: ITALIAN SYMPOSIUM ON ADVANCED DATABASE SYSTEMS, 14., jan. 2006, Portonovo. *Proceedings...* Portonovo: SEBD, 2006. p. 225-232. Disponível em: <https://www.researchgate.net/profile/Stefano_Montanelli/publication/220973995_Ontology-based_Classification_and_Retrieval_of_Music_Resources/links/5423fbb0cf26120b7a70b51.pdf>. Acesso em: 7 jul. 2017.
- CELINO, I. et al. Squiggle: a semantic search engine for indexing and retrieval of multimedia content. In: INTERNATIONAL CONFERENCE ON SEMANTIC-ENHANCED MULTIMEDIA PRESENTATION SYSTEMS, 1., v. 228, 2006, Athens, Greece. *Proceedings...* Athens, Greece: CEUR-WS. org, 2006. p. 40-54. Disponível em: <<http://dl.acm.org/citation.cfm?id=2889982>>. Acesso em: 7 jul. 2017.
- CELMA, Ò.; SERRA, X. FOAFing the music: bridging the semantic gap in music recommendation. *Web Semantics: Science, Services and Agents on the World Wide Web*, v. 6, n. 4, p. 250-256, 2008. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1570826808000711>>. Acesso em: 7 jul. 2017.
- DE OLIVEIRA ALBUQUERQUE, M. et al. An ontology for musical phonographic records: contributing with a representation model. In: WORLD SUMMIT ON KNOWLEDGE SOCIETY: BEST PRACTICES FOR THE KNOWLEDGE SOCIETY, v. 49, 2009, Berlin, Heidelberg. *Proceedings...* Berlin, Heidelberg: Springer, 2009. p. 495-502. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-642-04757-2_53>. Acesso em: 7 jul. 2017.
- DI, W. et al. Using OWL to construct and parse ontology for learning resources in network education environment. In: WRI GLOBAL CONGRESS ON INTELLIGENT SYSTEMS, 2., 2010, Wuhan, China. *Proceedings...* Wuhan, China: IEEE, 2010. p. 400-402. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/5709402/>>. Acesso em: 7 jul. 2017.
- DIEFENBACH, D. et al. Computing the semantic relatedness of music genre using semantic web data. In: INTERNATIONAL CONFERENCE ON SEMANTIC SYSTEMS, SEMANTiCS: Posters and Demos Track, 12., sep. 13-15, 2016, Leipzig, Germany. *Proceedings...* Leipzig, Germany, 2016. Disponível em: <<http://ceur-ws.org/Vol-1695/paper23.pdf>>. Acesso em: 7 jul. 2017.
- DIMITROVA, V. et al. Exploring exploratory search: a user study with linked semantic data. In: INTERNATIONAL WORKSHOP ON INTELLIGENT EXPLORATION OF SEMANTIC DATA, 2., 2013, New York, USA. *Proceedings...* New York, USA: ACM, 2013. p. 2. Disponível em: <<http://dl.acm.org/citation.cfm?id=2462199>>. Acesso em: 7 jul. 2017.
- ESPINOSA-ANKE, L. et al. *Finding and expanding hypernymic relations in the music domain*. In: INTERNATIONAL CONFERENCE OF THE CATALAN ASSOCIATION FOR ARTIFICIAL INTELLIGENCE, 19., out. 2016, Barcelona, Spain. Disponível em: <http://www.users.di.uniroma1.it/~collados/papers/CCIA16_EspinosaAnke.pdf>. Acesso em: 7 jul. 2017.
- ESPOSITO, N.; GESLIN, Y. Long-term preservation of acousmatic works: toward a generic model of description. In: IEEE Mediterranean Electrotechnical Conference, 14., MELECON, may 5-7, 2008, Ajaccio, France. *Proceedings...* Ajaccio, France: IEEE, 2008. p. 270-274. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/4618446/>>. Acesso em: 7 jul. 2017.

- FAZEKAS, G.; SANDLER, M.B. *The studio ontology framework*. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE, 12., oct. 24-28, 2011, Miami, Florida, USA. *Proceedings...* Miami, Florida, USA, 2011. p. 471-476. Disponível em: <<https://pdfs.semanticscholar.org/e093/9283174fa9343c931f5de56946dfee32f554.pdf>>. Acesso em: 7 jul. 2017.
- _____. Novel methods in information management for advanced audio workflows. In: INTERNATIONAL CONFERENCE ON DIGITAL AUDIO EFFECTS, 12., DAFx 2009, sep. 1-4, 2009, Como, Italy. *Proceedings...* Como, Italy: ISPG, 2009. p. 464-470. Disponível em: <http://www.academia.edu/download/42578102/paper_93.pdf>. Acesso em: 7 jul. 2017.
- FERRARA, A. et al. A semantic web ontology for context-based classification and retrieval of music resources. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, v. 2, n. 3, p. 177-198, 2006. Disponível em: <<http://dl.acm.org/citation.cfm?id=1152151>>. Acesso em: 1 jul. 2017.
- FIELDS, B. et al. The segment ontology: bridging music-generic and domain-specific. In: IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO, jul. 11-15, 2011, Barcelona, Spain. *Proceedings...* Barcelona, Spain: IEEE, 2011. p. 1-6. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/6012204/>>. Acesso em: 1 jul. 2017.
- FIELDS, B.; PHIPPEN, S.; COHEN, B. A case study in pragmatism: exploring the practical failure modes of linked data as applied to classical music catalogues. In: INTERNATIONAL WORKSHOP ON DIGITAL LIBRARIES FOR MUSICOLOGY, 2., ACM, jun. 25, 2015, Knoxville, TN, USA. *Proceedings...* Knoxville, TN, USA, 2015. p. 21-24. Disponível em: <<http://dl.acm.org/citation.cfm?id=2785531>>. Acesso em: 1 jul. 2017.
- FU, J.; XU, J.; JIA, K. Domain ontology based automatic question answering. In: INTERNATIONAL CONFERENCE ON COMPUTER ENGINEERING AND TECHNOLOGY, 2009. Singapore. *Proceedings...* Singapore: IEEE, 2009. p. 346-349. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/4769619/>>. Acesso em: 1 jul. 2017.
- GARCIA, R. et al. Multimedia content description using semantic web languages. In: KOMPATSIARIS, Y.; HOBSON, P. (Ed.). *Semantic multimedia and ontologies: theory and applications*, 2008. p. 17-54. Disponível em: <http://link.springer.com/content/pdf/10.1007/978-1-84800-076-6_2.pdf>. Acesso em: 1 jul. 2017.
- _____; CELMA, O. Semantic integration and retrieval of multimedia metadata. In: INTERNATIONAL WORKSHOP ON KNOWLEDGE MARKUP AND SEMANTIC ANNOTATION, 5., 2005. p. 69-80. Disponível em: <<http://ceur-ws.org/Vol-185/semAnnot05-07.pdf>>. Acesso em: 7 jul. 2017.
- GILLET, P. et al. Complex correspondences for query patterns rewriting. In: INTERNATIONAL CONFERENCE ON ONTOLOGY MATCHING, 8., v. 1111, 2013, Sydney, Australia. *Proceedings...* Sydney, Australia: CEUR-WS.org, 2013. p. 49-60. Disponível em: <<http://dl.acm.org/citation.cfm?id=2874498>>. Acesso em: 7 jul. 2017.
- GILSON, O. et al. From web data to visualization via ontology mapping. *Computer Graphics Forum*, v. 27, n. 3, p. 959-966, may 2008. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8659.2008.01230.x/full>>. Acesso em: 7 jul. 2017.
- GIORDANO, D. et al. A semantic-based and adaptive architecture for automatic multimedia retrieval composition. In: INTERNATIONAL WORKSHOP ON CONTENT-BASED MULTIMEDIA INDEXING, 9., jun. 13-15, 2011, Madrid, Spain. *Proceedings...* Madrid, Spain: IEEE, 2011. p. 181-186. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/5972542/>>. Acesso em: 7 jul. 2017.
- GRACY, K.F.; ZENG, M.L.; SKIRVIN, L. Exploring methods to improve access to music resources by aligning library data with linked data: a report of methodologies and preliminary findings. *Journal of the Association for Information Science and Technology*, v. 64, n. 10, p. 2078-2099, 2013. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.22914/full>>. Acesso em: 7 jul. 2017.
- HAN, B.-j. et al. Music emotion classification and context-based music recommendation. *Multimedia Tools and Applications*, v. 47, n. 3, p. 433-460, 2010. Disponível em: <<http://link.springer.com/article/10.1007/s11042-009-0332-6>>. Acesso em: 7 jul. 2017.
- HARISPE, S. et al. Semantic measures based on RDF projections: application to content-based recommendation systems. In: OTM CONFEDERATED INTERNATIONAL CONFERENCES "ON THE MOVE TO MEANINGFUL INTERNET SYSTEMS", 2013, Berlin, Heidelberg. Berlin, Heidelberg: Springer, 2013. p. 606-615. Disponível em: <http://link.springer.com/chapter/10.1007/978-3-642-41030-7_44>. Acesso em: 7 jul. 2017.
- HEEPS, S. et al. Dynamic ontology mapping for interacting autonomous systems. In: INTERNATIONAL WORKSHOP ON SELF-ORGANIZING SYSTEMS, 2., 2007, sep. 11-13, 2007, The Lake District, UK. *Proceedings...* The Lake District, UK: Springer, 2007. p. 255-263. Disponível em: <<http://link.springer.com/content/pdf/10.1007/978-3-540-74917-2.pdf#page=266>>. Acesso em: 7 jul. 2017.
- JOVANOVIK, M. et al. Linked music data from global music charts. In: INTERNATIONAL CONFERENCE ON SEMANTIC SYSTEMS, 10., sep. 4-5, 2014, Leipzig, Germany. *Proceedings...* Leipzig, Germany: ACM, 2014. p. 108-115. Disponível em: <<http://dl.acm.org/citation.cfm?id=2660536>>. Acesso em: 7 jul. 2017.
- JUNG, R. Take your smart music with you and be up to date. In: INTERNATIONAL CONFERENCE ON INTELLIGENT ENVIRONMENTS (IE 08), 4., jul. 21-22, 2008, Seattle, WA, USA. p. 41. Disponível em: <http://digital-library.theiet.org/content/conferences/10.1049/cp_20081171>. Acesso em: 7 jul. 2017.
- JURIC, D.; FAZEKAS, G. Knowledge extraction from audio content service providers' API descriptions. In: INTERNATIONAL CONFERENCE METADATA AND SEMANTICS RESEARCH, 10., nov. 22-25, 2016, Göttingen, Germany. *Proceedings...* Göttingen, Germany: Springer International Publishing, 2016. p. 55-66. Disponível em: <http://link.springer.com/chapter/10.1007/978-3-319-49157-8_5>. Acesso em: 7 jul. 2017.

- KAIYA, H.; SAEKI, M. Using domain ontology as domain knowledge for requirements elicitation. In: IEEE INTERNATIONAL CONFERENCE REQUIREMENTS ENGINEERING, 14., sep. 11-15, 2006, Minneapolis/St. Paul, MN, USA. p. 189-198. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/1704062/>>. Acesso em: 7 jul. 2017.
- KIM, H.H. A semantically enhanced tag-based music recommendation using emotion ontology. In: ASIAN CONFERENCE ON INTELLIGENT INFORMATION AND DATABASE SYSTEMS, 5., v. Part II, mar. 18-20, 2013, Kuala Lumpur, Malaysia. *Proceedings...* Kuala Lumpur, Malaysia: Springer-Verlag, 2013. p. 119-128. Disponível em: <<http://dl.acm.org/citation.cfm?id=2451031>>. Acesso em: 7 jul. 2017.
- _____.; KIM, D.; JO, J. A unified music recommender system using listening habits and semantics of tags. *International Journal of Intelligent Information and Database Systems*, v. 8, n. 1, p. 14-30, 2014. Disponível em: <<http://www.inderscienceonline.com/doi/abs/10.1504/IJIDS.2014.060460>>. Acesso em: 7 jul. 2017.
- KIM, M.; KWON, H.-C. Lyrics-based emotion classification using feature selection by partial syntactic analysis. In: IEEE INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE, 23., nov. 7-9, 2011, Boca Raton, FL, USA. *Proceedings...* Boca Raton, FL, USA: IEEE, 2011. p. 960-964. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/6103456/>>. Acesso em: 7 jul. 2017.
- KIM, S.K.; KIM, J.-M.; SHIN, P.S. Designing a music content retrieval system supporting Korean. In: INTERNATIONAL CONFERENCE ON INFORMATION SCIENCE AND APPLICATIONS, jun. 24-26, 2013, Suwon, South Korea. *Proceedings...* Suwon, South Korea: IEEE, 2013. p. 1-4. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/6579492/>>. Acesso em: 7 jul. 2017.
- KODURI, G.K. Culture-aware approaches to modeling and description of intonation using multimodal data. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT, EKAW (Satellite Events), 2014. p. 209-217. Disponível em: <<http://mtg.upf.edu/system/files/publications/doctoral-consortium.pdf>>. Acesso em: 7 jul. 2017.
- KOLOZALI, S. et al. Knowledge representation issues in musical instrument ontology design. INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE, 12., oct. 24-28, 2011, Miami, Florida, USA. *Proceedings...* Miami, Florida, USA, 2011. p. 465-470. Disponível em: <<http://ismir2011.ismir.net/papers/PS3-19.pdf>>. Acesso em: 7 jul. 2017.
- _____. et al. Towards the automatic generation of a semantic web ontology for musical instruments. In: INTERNATIONAL CONFERENCE ON SEMANTIC AND DIGITAL MEDIA TECHNOLOGIES. Berlin, Heidelberg: Springer, 2010. p. 186-187. Disponível em: <http://link.springer.com/10.1007%2F978-3-642-23017-2_13>. Acesso em: 7 jul. 2017.
- LEE, J.H.; CHO, H.; KIM, Y.-S. Users' music information needs and behaviors: design implications for music information retrieval systems. *Journal of the Association for Information Science and Technology*, v. 67, n. 6, p. 1301-1330, 2016. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.23471/full>>. Acesso em: 7 jul. 2017.
- LISENA, P. et al. Exploring linked classical music catalogs with OVERTURE. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 15., Kobe, Japan, Poster Track, oct. 17-21, 2016. Disponível em: <<https://pdfs.semanticscholar.org/b684/d094a00fc7bfe20314daa835cf07d6d21a5f.pdf>>. Acesso em: 7 jul. 2017.
- LUGER, M. et al. EASAIER semantic music retrieval portal. In: SEMANTICS AND DIGITAL MEDIA TECHNOLOGY: Posters and Demos, Genoa, Italy, 2007. p.13-14. Disponível em: <http://info.slis.indiana.edu/~dingying/Publication/mluger_samt07.pdf>. Acesso em: 10 jun. 2017.
- LUZZI, C. ManUScript Italian poEtry in muSiC (1500-1700) interoperable model: towards an application of FRBRoo, Linked Open Data and Semantic Web technology. In: INTERNATIONAL WORKSHOP ON DIGITAL LIBRARIES FOR MUSICOLOGY, 1., 2014, London, United Kingdom. *Proceedings...* London, United Kingdom: ACM, 2014. p. 1-3. Disponível em: <<http://dl.acm.org/citation.cfm?id=2660189>>. Acesso em: 10 jun. 2017.
- MARQUES, C. M.; VON ZUBEN, J.; GUILHERME, I.R. FTMOntology: an ontology to fill the semantic gap between music, mood, personality, and human physiology. In: OTM CONFEDERATED INTERNATIONAL CONFERENCES "ON THE MOVE TO MEANINGFUL INTERNET SYSTEMS", 2011, Berlin, Heidelberg. *Proceedings...* Berlin, Heidelberg: Springer, 2011. p. 27-28. Disponível em: <https://link.springer.com/chapter/10.1007%2F978-3-642-25126-9_6>. Acesso em: 10 jun. 2017.
- MIRYLENKA, D.; PASSERINI, A.; SERAFINI, L. Bootstrapping domain ontologies from Wikipedia: a uniform approach. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 24., 2015, Buenos Aires, Argentina. *Proceedings...* Buenos Aires, Argentina: AAIA, 2015. p. 1464-1470. Disponível em: <<http://www.aaai.org/ocs/index.php/IJCAI/IJCAI15/paper/download/11006/10867>>. Acesso em: 7 jul. 2017.
- MOCHOLI, J.A. et al. A multicriteria ant colony algorithm for generating music playlists. *Expert Systems with Applications*, v. 39, n. 3, p. 2270-2278, 2012. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417411011043>>. Acesso em: 7 jul. 2017.
- MOTTA, E.; SIQUEIRA, S.W. M.; ANDREATTA, A. Populating a domain ontology from a web biographical dictionary of music: an unsupervised rule-based method to handle Brazilian Portuguese texts. In: INTERNATIONAL CONFERENCE ON WEB INFORMATION SYSTEMS AND TECHNOLOGIES WEBIST, 2009, Lisbon, Portugal. *Proceedings...* Lisbon, Portugal, 2009. p. 192-199. Disponível em: <<http://www.scitepress.org/DigitalLibrary/PublicationsDetail.aspx?ID=2VJ83RagfoI%3d&t=1>>. Acesso em: 7 jul. 2017.

- NAKATSUJI, M.; MIYOSHI, Y.; OTSUKA, Y. Innovation detection based on user-interest ontology of blog community. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 5., nov. 5-9, 2006, Athens, GA, USA. *Proceedings...* Athens, GA, USA, 2006. p. 515-528. Disponível em: <<http://link.springer.com/content/pdf/10.1007/11926078.pdf#page=538>>. Acesso em: 7 jul. 2017.
- NISHEVA-PAVLOVA, M.; PAVLOV, P. Ontology-based search and document retrieval in a digital library with folk songs. *Information Services & Use*, v. 31, n. 3-4, p. 157-166, 2011. Disponível em: <<http://content.iospress.com/articles/information-services-and-use/isu645>>. Acesso em: 7 jul. 2017.
- PAREDES-VALVERDE, M.A. et al. A semantic-based approach for querying linked data using natural language. *Journal of Information Science*, v. 42, n. 6, p. 851-862, 2016. Disponível em: <<http://journals.sagepub.com/doi/abs/10.1177/0165551515616311>>. Acesso em: 7 jul. 2017.
- PASSANT, A.; DECKER, S. Hey! ho! let's go! explanatory music recommendations with dbrec. In: EXTENDED SEMANTIC WEB CONFERENCE: THE SEMANTIC WEB: RESEARCH AND APPLICATIONS, Berlin, Heidelberg, 2010, p. 411-415. Disponível em: <<http://www.springerlink.com/index/X715525387672351.pdf>>. Acesso em: 7 jul. 2017.
- PEDRINACI, C. et al. Music rights clearance business analysis and delivery. In: INTERNATIONAL CONFERENCE, EC-Web, aug. 23-26. 2005, Copenhagen, Denmark. *Proceedings...* Copenhagen, Denmark, 2005. p. 198-207. Disponível em: <<http://link.springer.com/content/pdf/10.1007/11545163.pdf#page=210>>. Acesso em: 7 jul. 2017.
- POON, F. et al. Semantically enhanced matchmaking of consumers and providers: a Canadian real estate case study. In: INTERNATIONAL CONFERENCE ON INFORMATION INTEGRATION AND WEB-BASED APPLICATIONS AND SERVICES, 13., dec. 05-07, 2011, Ho Chi Minh City, Vietnam. *Proceedings...* Ho Chi Minh City, Vietnam: ACM, 2011. p. 198-205. Disponível em: <<http://dl.acm.org/citation.cfm?id=2095570>>. Acesso em: 7 jul. 2017.
- PRESUTTI, V.; GANGEMI, A. Content ontology design patterns as practical building blocks for web ontologies. In: INTERNATIONAL CONFERENCE ON CONCEPTUAL MODELING, ER 2008, 27., oct. 20-24, 2008, Barcelona, Spain. *Proceedings...* Barcelona, Spain, 2008. p. 128-141, 2008. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-540-87877-3_11>. Acesso em: 7 jul. 2017.
- PYSHKIN, E.; KUZNETSOV, A. Approach to building a web-based expert system interface and its application for software provisioning in clouds. In: FEDERATED CONFERENCE ON COMPUTER SCIENCE AND INFORMATION SYSTEMS, sep. 13-16, 2015, Lodz, Poland. *Proceedings...* Lodz, Poland: IEEE, 2015. p. 343-354. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/7321464/>>. Acesso em: 7 jul. 2017.
- RAHMAN, F.; SIDDIQI, J. Semantic annotation of digital music. *Journal of Computer and System Sciences*, v. 78, n. 4, p. 1219-1231, 2012. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022000011001164>>. Acesso em: 7 jul. 2017.
- RAIMOND, Y. et al. The music ontology. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL, 8., 2007. *Proceedings...* Austrian Computer Society, 2007. Disponível em: <http://ismir2007.ismir.net/proceedings/ISMIR2007_p417_raimond.pdf>. Acesso em: 7 jul. 2017.
- _____; SANDLER, Mark. Evaluation of the music ontology framework. In: EXTENDED SEMANTIC WEB CONFERENCE: THE SEMANTIC WEB: RESEARCH AND APPLICATIONS, 2012, Berlin, Heidelberg. *Proceedings...* Berlin, Heidelberg: Springer, 2012, p. 255-269. Disponível em: <<http://www.springerlink.com/index/K7NV56134N6503H0.pdf>>. Acesso em: 7 jul. 2017.
- REFORMAT, M.Z.; GOLMOHAMMADI, S.K. Rule- and OWA-based Semantic Similarity for User Profiling. *International Journal of Fuzzy Systems*, v. 12, n. 2, p. 87-102, 2010. Disponível em: <http://www.ijfs.org.tw/ePublication/2010_paper_2/ijfs10-2-r-1-ReformatGolmohaFinal-proof%20o.pdf>. Acesso em: 7 jul. 2017.
- _____; _____. Updating user profile using ontology-based semantic similarity. In: IEEE INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS, aug. 20-24, 2009, p. 1062-1067. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/5277205/>>. Acesso em: 7 jul. 2017.
- RESTAGNO, L. et al. A semantic web annotation tool for a web-based audio sequencer. In: INTERNATIONAL CONFERENCE ON WEB ENGINEERING, jun. 20-24, 2011, Paphos, Cyprus. *Proceedings...* Paphos, Cyprus: Springer, 2011. p. 289-303. Disponível em: <<http://link.springer.com/content/pdf/10.1007/978-3-642-22233-7.pdf#page=302>>. Acesso em: 7 jul. 2017.
- RHO, S. et al. COMUS: Ontological and rule-based reasoning for music recommendation system. In: PACIFIC-ASIA CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING: ADVANCES IN KNOWLEDGE DISCOVERY AND DATA MINING. Berlin, Heidelberg: Springer, 2009. p. 859-866. Disponível em: <<http://www.springerlink.com/index/D7W3118388690846.pdf>>. Acesso em: 7 jul. 2017.
- _____. et al. Implementing situation-aware and user-adaptive music recommendation service in semantic web and real-time multimedia computing environment. *Multimedia tools and applications*, v. 65, n. 2, p. 259-282, 2013. Disponível em: <<http://link.springer.com/article/10.1007/s11042-011-0803-4>>. Acesso em: 7 jul. 2017.
- _____; HAN, B.-j.; HWANG, E. SVR-based music mood classification and context-based music recommendation. In: INTERNATIONAL CONFERENCE ON MULTIMEDIA, 17., oct. 19-24, 2009, Beijing, China. *Proceedings...* Beijing, China: ACM, 2009. p. 713-716. Disponível em: <<http://dl.acm.org/citation.cfm?id=1631395>>. Acesso em: 7 jul. 2017.

- RODRÍGUEZ-GARCÍA, M.Á. et al. Ontology-based music recommender system. In: INTERNATIONAL CONFERENCE ON DISTRIBUTED COMPUTING AND ARTIFICIAL INTELLIGENCE, 12., v. 373, 2015. *Proceedings...* 2015. p. 39-46. (Series Advances in Intelligent Systems and Computing). Disponível em: <http://link.springer.com/chapter/10.1007/978-3-319-19638-1_5>. Acesso em: 7 jul. 2017.
- RUPP, P.; ESTIER, T. A model for a better understanding of the digital distribution of music in a peer-to-peer environment. In: ANNUAL HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 36., jan. 6-9, 2003, Big Island, HI, USA. *Proceedings...* Big Island, HI, USA: IEEE, 2003. p. 11. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/1174819/>>. Acesso em: 7 jul. 2017.
- SCHREIBER, H. Genre ontology learning: comparing curated with crowd-sourced ontologies. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE, 17., aug. 7-11, 2016, New York City, USA. *Proceedings...* New York City, USA, 2016. p. 400-406. Disponível em: <http://www.tagtraum.com/download/schreiber_learnedgenreontologies_ismir2016.pdf>. Acesso em: 7 jul. 2017
- SCHUMACHER, F.; GEY, R.; KLINGNER, S. Semantics for the music industry: the development of the music business ontology (MBO). In: INTERNATIONAL CONFERENCE ON SEMANTIC SYSTEMS, 10., 2014, Leipzig, Germany. *Proceedings...* Leipzig, Germany: ACM, 2014. p. 77-83. Disponível em: <<http://dl.acm.org/citation.cfm?id=2660531>>. Acesso em: 7 jul. 2017.
- SEBASTIEN, V.; SEBASTIEN, D.; CONRUYT, N. An ontology for musical performances analysis: application to a collaborative platform dedicated to instrumental practice. In: INTERNATIONAL CONFERENCE ON INTERNET AND WEB APPLICATIONS AND SERVICES, 5., may 9-15, 2010, Barcelona, Spain. *Proceedings...* Barcelona, Spain: IEEE, 2010. p. 538-543. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/5476485/>>. Acesso em: 7 jul. 2017.
- _____.; _____.; _____. Annotating works for music education: propositions for a musical forms and structures ontology and a musical performance ontology. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE, 14., 2013. Disponível em: <http://www.ppgia.pucpr.br/ismir2013/wp-content/uploads/2013/09/83_Paper.pdf>. Acesso em: 7 jul. 2017.
- SOHN, J.-S.; KIM, D.-H.; CHUNG, I.-J. Ontology-based ID3 tag management system. In: INTERNATIONAL CONFERENCE ON ICT CONVERGENCE, oct. 15-17, 2012, Jeju Island, South Korea. *Proceedings...* Jeju Island, South Korea: IEEE, 2012. p. 255-257. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/6386833/>>. Acesso em: 7 jul. 2017.
- SOMMARUGA, L.; BAGGI, D. Towards a semantic web based model for the tonal system in standard IEEE 1599. *Journal of Multimedia*, v. 4, n. 1, feb. 2009. Disponível em: <<https://pdfs.semanticscholar.org/4ea8/be43e5f12ac54c906407d182327fbb9d0fd0.pdf>>. Acesso em: 7 jul. 2017.
- STEYN, J. The information architecture of music. In: _____. *Structuring music through markup language: designs and architectures*. Scopus, 2012. Cap. 1, p. 1-28.
- THALMANN, F. et al. The mobile audio ontology: experiencing dynamic music objects on mobile devices. In: INTERNATIONAL CONFERENCE ON SEMANTIC COMPUTING, 10., 2016, Laguna Hills, CA, USA. *Proceedings...* Laguna Hills, CA, USA: IEEE, 2016. p. 47-54. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/7439304/>>. Acesso em: 7 jul. 2017.
- TIAN, M. et al. Towards the representation of Chinese traditional music: a state of the art review of music metadata standards. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, sep. 2-6, 2013, Lisbon, Portugal. *Proceedings...* Lisbon, Portugal, 2013. p. 71-81. Disponível em: <<http://dcpapers.dublincore.org/pubs/article/viewFile/3672/1895>>. Acesso em: 7 jul. 2017.
- TIAN, Y. et al. An ontology-based model driven approach for a music learning system. In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND KNOWLEDGE ENGINEERING, 2009, Boston, Massachusetts, USA. *Proceedings...* Boston, Massachusetts, USA: 2009. p. 739-744.
- TUNG, W.-F. et al. Collaborative service system design for music content creation. *Information Systems Frontiers*, v. 16, n. 2, p. 291-302, apr. 2014. Disponível em: <<http://link.springer.com/article/10.1007/s10796-012-9346-0>>. Acesso em: 7 jul. 2017.
- WANG, J. et al. A collaborative model of low-level and high-level descriptors for semantics-based music information retrieval. In: IEEE INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE AND INTELLIGENT AGENT TECHNOLOGY, dec. 9-12, 2008, Sydney, NSW, Australia. *Proceedings...* Sydney, NSW, Australia: ACM; IEEE, 2008. p. 532-535. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/4740503/>>. Acesso em: 7 jul. 2017.
- _____. et al. Enriching music mood annotation by semantic association reasoning. In: IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO, jul. 19-23, 2010, Suntec City, Singapore. *Proceedings...* Suntec City, Singapore: IEEE, 2010. p. 1445-1450. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/5583243/>>. Acesso em: 7 jul. 2017.
- _____. et al. Predicting high-level music semantics using social tags via on-tology-based reasoning. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE, 11., aug. 9-13, 2010, Utrecht, Netherlands. *Proceedings...* Utrecht, Netherlands, 2010, p. 405-410. Disponível em: <<http://www.ismir2010.ismir.net/proceedings/ismir2010-69.pdf>>. Acesso em: 7 jul. 2017.
- WARDHANA, A.T.A.; NUGROHO, H.T. Combining FOAF and music ontology for music concerts recommendation on Facebook application. In: CONFERENCE ON NEW MEDIA STUDIES, nov. 27-28, 2013, Tangerang, Indonesia. *Proceedings...* Tangerang, Indonesia: IEEE, 2013. p. 1-5. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/6708544/>>. Acesso em: 7 jul. 2017.

WENG, J.-F. et al. Constructing an Immersive Poetry Learning multimedia environment using ontology-based approach. In: IEEE INTERNATIONAL CONFERENCE ON UBI-MEDIA COMPUTING, 1., 2008, Lanzhou, China. *Proceedings...* Lanzhou, China: IEEE, 2008. p. 308-313. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/4570908/>>. Acesso em: 7 jul. 2017.

WEYDE, T. Automatic semantic annotation of music with harmonic structure. In: SOUND AND MUSIC COMPUTING CONFERENCE, 4., jul. 11-13, 2007, Lefkada, Greece. 2007. Disponível em: <<http://openaccess.city.ac.uk/2967/>>. Acesso em: 7 jul. 2017.

WILMERING, T.; FAZEKAS, G.; SANDLER, M. Towards ontological representations of digital audio effects. In: INTERNATIONAL CONFERENCE ON DIGITAL AUDIO EFFECTS, 14., sep. 19-23, 2011, Paris, France. *Proceedings...* Paris, France, 2011. P. 119-122. Disponível em: <http://recherche.ircam.fr/pub/dafx11/Papers/64_e.pdf>. Acesso em: 7 jul. 2017.

_____; _____. AUFX-O: Novel methods for the representation of audio processing workflows. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 15., oct. 17-21, 2016, Kobe, Japan. *Proceedings...* Kobe, Japan: Springer International Publishing; Cham, 2016. p. 229-237. Disponível em: <http://link.springer.com/chapter/10.1007/978-3-319-46547-0_24>. Acesso em: 7 jul. 2017.

_____; _____. The audio effects ontology. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL, nov. 4-8, 2013, Curitiba, Brazil. *Proceedings...* Curitiba, Brazil, 2013. p. 215-220. Disponível em: <<http://ai2-s2-dfs.s3.amazonaws.com/5036/7aa19c4ad07f9fd17eac15215e46addb1194.pdf>>. Acesso em: 7 jul. 2017.

_____; _____. Interdisciplinary classification of audio effects in the audio effect ontology. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 15., 2016, Kobe, Japan. *Proceedings...* Kobe, Japan: Springer International Publishing; Cham, 2016. Disponível em: <<http://ceur-ws.org/Vol-1690/paper60.pdf>>. Acesso em: 7 jul. 2017.

WU, D.; SHI, J. Classical music recording ontology used in a library catalog. *Knowledge Organization*, v. 43, n. 6, 2016. Disponível em: <<http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=09437444&AN=117778144&h=eNZ9zXiUKlil8qeZjzOr-vsDQ6%2FonePmZvs1rYkjqkYgY8ohwuz93rvputFkSjAUFEO5l-BNnDZUq57%2ByXdnNe2w%3D%3D&crl=c>>. Acesso em: 7 jul. 2017.

YANG, L.; GAO, J. The construction of music domain ontology. In: WORKSHOP ON CHINESE LEXICAL SEMANTICS: CHINESE LEXICAL SEMANTICS, 2012. Berlin, Heidelberg: Springer, 2012. p. 829-835. Disponível em: <http://link.springer.com/chapter/10.1007/978-3-642-36337-5_84>. Acesso em: 7 jul. 2017.

Ontologias como suporte à modelagem da informação na arquitetura, engenharia e construção

Renata M. A. Baracho

Professora Associada do Programa de Pós Graduação em Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais (PPG-GOC/UFMG), Belo Horizonte, MG, Brasil.

<http://lattes.cnpq.br/4218954956709188>

E-mail: renatabaracho@ufmg.br

Livia Marangon Duffles Teixeira

Doutoranda em Ciência da Informação pela Universidade Federal de Minas Gerais - Programa de Pós Graduação em Gestão e Organização do Conhecimento (PPG-GOC/UFMG), Belo Horizonte, MG, Brasil.

<http://lattes.cnpq.br/5511349023940518>

E-mail: liviamarangon@gmail.com

Mário Lucio Pereira Junior

Doutorando em Ciência da Informação pela Universidade Federal de Minas Gerais - Programa de Pós Graduação em Gestão e Organização do Conhecimento (PPG-GOC/UFMG), Belo Horizonte, MG, Brasil.

<http://lattes.cnpq.br/5892357969061565>

E-mail: mariolpereiraj@gmail.com

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

RESUMO

Ontologias são estruturas de organização do conhecimento que podem ser aplicadas a diversos domínios, inclusive arquitetura, engenharia e construção civil (AEC). Para que se estabeleçam relações existentes entre as entidades em determinado contexto, é necessário vinculá-las às representações da modelagem gráfica, bem como elaborar suas definições subsidiadas por um vocabulário controlado. O Building Information Modeling ou Modelagem de Informação da Construção (BIM) é um ambiente tecnológico para a gestão de informações em processos colaborativos, interdisciplinares e interoperáveis, que sinaliza uma adequação semântica das entidades (sejam materiais, objetos, representações) através da validação de compromissos ontológicos. A vinculação Internet of Things ou Internet das Coisas (IoT) a modelos elaborados via BIM permite identificar e localizar as entidades/objetos, evidenciar a interconexão entre eles e incrementar a modelagem das informações. O presente artigo investiga a conexão teórica entre a tecnologia BIM, as ontologias como vocabulários controlados e IoT no âmbito da AEC. Para tal, apresenta revisão de literatura sobre os temas mencionados e elabora descrições empíricas para validação do arcabouço teórico preliminar. Espera-se que a utilização simultânea da modelagem gráfica e da modelagem da informação resulte na criação de uma perspectiva de empreendimento em AEC com correspondência fidedigna ao mundo real, ao mesmo tempo que se adequa à tecnologia IoT Internet das Coisas.

Palavras-chave: Ontologia. BIM. Recuperação da informação. Representação da informação. Internet das coisas.

Ontologies as support for information modeling in architecture, engineering and construction

ABSTRACT

Ontologies are structures of knowledge organization that can be applied to several domains, including architecture, engineering and civil construction (AEC). In order to establish existing relationships between entities in a given context, it is necessary to link them to representations of graphic modeling, as well as to elaborate their definitions subsidized by a controlled vocabulary. Building Information Modeling (BIM) is a technological environment for the management of information in collaborative, interdisciplinary and interoperable processes, which requires the semantic adequacy of entities (be it materials, objects, perspectives) through the validation of commitments Ontological. The connection of Internet of Things (IoT) to the models elaborated via BIM allows to identify and to locate of the entities / objects, to show the interconnection between them and to increase the modeling of the information. This article investigates the theoretical connection between BIM technology, ontologies as controlled vocabularies and IoT within the scope of the AEC. To this end, it presents a review of the literature on the topics mentioned and elaborates empirical descriptions for validation of the preliminary theoretical framework. It is hoped that the simultaneous use of graphical modeling and information modeling will result in the creation of a business perspective in AEC with true correspondence to the real world, while terminologically adequate for IoT.

Keywords: *Ontology. BIM technology. Information Retrieval. Representation of Information. Internet of things.*

Ontologías como soporte para modelar la información en arquitectura, ingeniería y construcción

RESUMEN

Las ontologías son estructuras de organización del conocimiento que pueden aplicarse a diversos dominios, incluyendo arquitectura, ingeniería y construcción civil (AEC). Para que se establezcan relaciones existentes entre las entidades en determinado contexto es necesario vincularlas a las representaciones del modelado gráfico, así como elaborar sus definiciones subsidiadas por un vocabulario controlado. El Building Information Modeling o Modelado de Información de la Construcción (BIM) es un entorno tecnológico para la gestión de la información en procesos colaborativos, interdisciplinarios e interoperables, que señalan una adecuación semántica de las entidades (sean materiales, objetos, representaciones) mediante la validación de compromisos ontológicos. La vinculación Internet of Things o Internet de las cosas (IoT) a modelos elaborados vía BIM permite identificar y localizar las entidades/objetos, evidenciar la interconexión entre ellos e incrementar el modelado de las informaciones. El presente artículo investiga la conexión teórica entre la tecnología BIM, las ontologías como vocabularios controlados e IoT en el ámbito de la AEC. Para ello, presenta revisión de literatura sobre los temas mencionados y elabora descripciones empíricas para validación del marco teórico preliminar. Se espera que la utilización simultánea del modelado gráfico y del modelado de la información resulte en la creación de una perspectiva de emprendimiento en AEC con correspondencia fidedigna al mundo real, al tiempo que se adecua a la tecnología IoT.

Palabras clave: *Ontología. Modelado de información de la construcción. Recuperación de la información. Representación de la información. Internet de las cosas.*

INTRODUÇÃO

A difusão da internet e do trabalho em rede propiciou o aumento da competitividade, possibilitou automação e compartilhamento de atividades e trouxe novas formas de controle e gestão. A partir desse contexto, identificam-se no mercado demandas por melhorias nos processos de negócios, baseadas nos recursos informacionais das organizações. A informação, portanto, torna-se um ativo indispensável para a melhoria das atividades e dos processos que se desencadeiam em produtos e/ou serviços.

As tecnologias de informação e as novas formas de representação do mundo, real e imaginário, têm influenciado profundamente a sociedade atual, indo em direção a um contexto de convergência e integração de tecnologias. Nesse cenário, os limites entre real e virtual, físico e digital estão cada vez mais frágeis e sutis.

A eficiência da gestão de informações, independentemente do domínio de aplicação, é sustentada por teorias que culminam na prática de representação da informação e do conhecimento. O resultado da prática consiste em um modelo representativo com maior similaridade a uma parte da realidade e às demandas explicitadas.

O estudo da prática de modelagem se justifica como apontado por Campos (2004), uma vez que os mecanismos de representação permitem que os processos de formalização sobre os objetos e suas relações em contextos predefinidos possam ser facilmente representados. Outrossim, os modelos são uma maneira de especificar as entidades e as relações entre as entidades dentro do domínio específico de conhecimento.

Existem diferentes metodologias e estruturas para se realizar a representação de modelos, como as linguagens de representação gráfica ou as estruturas terminológicas. Para a representação de processos de negócios, exemplifica-se com a notação BPMN (*Business Process Model and Notation*) que, através dos seus construtos, representa as atividades que acontecem em determinado ambiente com determinado fim.

As ontologias são outra forma para se representar um domínio do conhecimento, estabelecendo uma estrutura que define as regras e critérios que regulam a combinação entre termos e relações em um domínio do conhecimento. As linguagens documentárias são estruturas terminológicas que listam e organizam os conceitos representativos de áreas do conhecimento, de forma mais genérica ou específica. No campo de Arquitetura, Engenharia e Construção (AEC) tem-se a tecnologia BIM (*Building Information Modeling* ou Modelagem da Informação da Construção) para a representação e gestão de informações, tendo como base a simulação do empreendimento em todo seu ciclo de vida.

O aumento da conexão das máquinas e equipamentos significa um volume de dados produzidos que ultrapassa a capacidade de processamento com tecnologias tradicionais. Assim, verifica-se a necessidade de novos processos de representação e organização da informação e do conhecimento que possibilitem a otimização da recuperação de informação. Nesse contexto, vocabulários controlados como as ontologias surgem como soluções promissoras.

As ontologias podem ser aplicadas a diversos domínios, inclusive arquitetura, engenharia e construção. Para estabelecer as relações entre as entidades de um contexto é necessário vinculá-las às representações da modelagem gráfica, além de elaborar suas definições subsidiadas por um vocabulário controlado. A tecnologia BIM, que se configura como um ambiente para a gestão de informações em processos colaborativos, interdisciplinares e interoperáveis, carece de adequação semântica das suas entidades (sejam materiais, objetos, representações) através da validação de compromissos ontológicos. A inserção da *Internet of Things* (IoT) ou Internet das Coisas a modelos elaborados via BIM permite identificar e localizar as entidades/objetos, evidenciar a interconexão entre eles e incrementar a modelagem das informações.

A recuperação da informação de um projeto de engenharia se processa, principalmente, através de três fontes: descrição textual, gráfica e tabelas. Mesmo que todas estejam dentro da mesma tecnologia, precisam estar representadas em uniformidade semântica para que resultem satisfatoriamente como demandas de pesquisa. Logo, o trabalho investiga a conexão teórica entre a tecnologia BIM, as ontologias com os vocabulários controlados e IoT, no âmbito da AEC. Para tal, apresenta uma revisão de literatura sobre os temas mencionados e elabora descrições empíricas para validação do arcabouço teórico preliminar. Espera-se que a utilização simultânea da modelagem gráfica e da modelagem da informação resulte na criação de uma perspectiva de empreendimento em AEC com correspondência fidedigna ao mundo real, ao mesmo tempo que se adequa à tecnologia IoT Internet das Coisas.

Parte-se do princípio que a interseção entre esses três temas poderá trazer avanços significativos para todo o ciclo de vida da construção de edificações. Isso ocorrerá principalmente na melhoria dos processos por meio da representação, recuperação e comunicação da informação entre as etapas. Visando redução ou eliminação de erros, desperdícios, otimização de tempo e melhoria na qualidade, espera-se obter ganhos globais, desde a criação do projeto até a utilização da obra edificada.

Os procedimentos metodológicos adotados nesta pesquisa foram os seguintes:

- i. apresentação de arcabouço teórico dos conceitos, a fim de sustentar as proposições apresentadas pelo propósito do trabalho, envolvendo a modelagem gráfica, a modelagem espacial e a modelagem da informação:
 - o primeiro apontamento consistiu na fundamentação sobre modelos de informação da construção com a tecnologia BIM;
 - o segundo apontamento consistiu na fundamentação sobre modelos em ontologias;
 - o terceiro apontamento consistiu na

fundamentação sobre modelos no contexto de vocabulário controlado;

- o quarto apontamento consistiu na fundamentação sobre a IoT internet das coisas;
 - o quinto apontamento consistiu em vincular os apontamentos e propor o *framework*;
- ii. relato da experiência em formato prototípico, com apresentação de resultados.

O trabalho está organizado conforme segue. A primeira seção apresenta a introdução e a metodologia desenvolvida no arrolar da pesquisa. A segunda seção apresenta o *background* sobre os conceitos abordados: tecnologia BIM, ontologias e IoT. A seção 3 relata a aplicação empírica da proposta e apresenta os resultados encontrados. E finalmente, a seção 4 oferece as considerações finais.

BACKGROUND

As ontologias são estruturas representacionais baseadas nos conceitos e nas relações entre eles, dentro de um domínio específico do conhecimento. Assim, elas permitem representar um recorte do mundo e prover soluções semânticas, propiciando e melhorando a interoperabilidade entre sistemas, além de automatizar processos de inferências (TEIXEIRA E AGANETTE, 2016). A partir desse entendimento, justifica-se a relação entre os conceitos abordados neste trabalho.

Propõe-se nesta seção evidenciar a vinculação dos conceitos enquanto solução à geração de modelos representacionais semanticamente consistentes. Teixeira e Baracho (2017) identificaram o *gap* terminológico nos modelos gráficos de empreendimentos, enxergaram os ganhos dessa relação para a AEC e propuseram um *framework* metodológico para futura validação empírica.

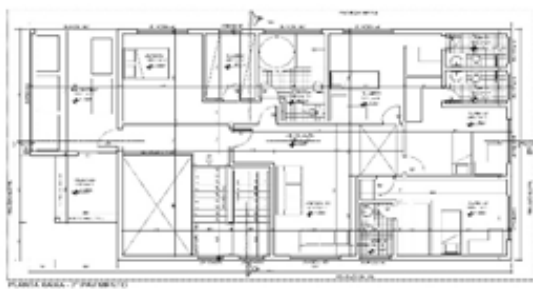
MODELAGEM DA INFORMAÇÃO DA CONSTRUÇÃO

Modelos são abstrações para representar parte da realidade. As diferentes metodologias e as diferentes notações utilizadas resultam em modelos representativos similares ou fidedignos ao mundo real. A representação de modelos em AEC demanda tanto em projeto, quanto na execução, tanto no *as built* do empreendimento, a simulação do mundo real. Isso se justifica, pois ele se torna a referência para o planejamento da obra, para a sua construção e para a manutenção da edificação.

A inserção de tecnologias e do trabalho em rede proporcionou mudanças no trabalho de engenheiros, arquitetos e construtores, como no sequenciamento de ações em planejamento e na gestão das informações textuais e gráficas de modo conjunto. A tecnologia BIM, além de atender essas demandas, proporciona recursos de representação e visualização, permitindo a interação direta e intuitiva, bastante similar a uma obra virtual, simulando virtualmente a construção (LEE et al., 2006).

A modelagem arquitetônica de um empreendimento a ser construído, num método tradicional, se inicia pela representação geométrica através de plantas e cortes (figura 1), ou seja, são desenhos geométricos sem semântica ou informação alguma vinculada. Posteriormente, a atividade evolui para a etapa de construção da perspectiva, ou seja, à modelagem tridimensional e simulação do mundo real (figura 2).

Figura 1 – Exemplo de representação geométrica



Fonte: Teixeira e Baracho (2017).

Figura 2 – Exemplo de simulação do mundo real



Fonte: Teixeira e Baracho (2017).

A tecnologia BIM apresenta princípios norteadores que mudam a forma linear e individual do processo de planejamento, execução e manutenção. Considera a edificação como um processo vivo, desde o levantamento dos requisitos feitos pelo arquiteto para iniciar a proposta até o possível descarte da construção, e o trabalho em arquivo digital dinâmico e colaborativo, com o desenvolvimento integrado de todos os projetos, do arquitetônico até os complementares, estrutura, elétrico, hidráulico, lógico, detalhamento. Propõe uma ação integrada e inteligente de desenvolvimento de projeto. Paralelamente, o projeto é criado em três dimensões, simulando virtualmente a situação real. Isso possibilita não só a visão espacial, mas também a compatibilização dos projetos e trabalho colaborativo dos profissionais. Outro ponto da tecnologia BIM é que os elementos são estruturados de maneira inter-relacionada. Existem entidades e relacionamentos nos objetos criados, propondo uma rede de iterações entre os elementos construtivos. Todos os elementos construtivos estão associados a banco de dados.

Conforme Eastman et al. (2008), com a tecnologia BIM um modelo virtual preciso do edifício é construído digitalmente e, quando completo, diferentemente de um simples modelo tridimensional ou desenho bidimensional da tecnologia CAD (Computer Aided Design), contém não só a geometria, mas também diversos dados ou informações relevantes necessárias à construção, fabricação, manutenção e demais atividades para realização da construção.

Assim, é uma nova filosofia de trabalho que integra arquitetos, engenheiros e construtores, que passam a trabalhar de modo colaborativo, alterando e dinamizando o ciclo da informação.

De acordo com Pereira Junior e Baracho (2015), a tecnologia BIM pode ser considerada uma evolução em todo o processo construtivo, na medida em que permite novas possibilidades de construção, projeto, planejamento, visualização, representação, processamento, uso e recuperação da informação, contida na representação geométrica e na ligação com os atributos do bando de dados.

Do ponto de vista da gestão do conhecimento, a tecnologia BIM apresenta uma evolução da tecnologia CAD ao trabalhar com informações semânticas sobre os objetos e suas relações, aproximando, assim, a fase de criação, projeto, das fases de execução e monitoramento, durante toda a vida útil do edifício. Um modelo BIM é tridimensional, como um modelo CAD pode ser, mas também é paramétrico e baseado em objetos, o que possibilita a criação de entidades editáveis em suas propriedades nativas e reais, que podem ser alteradas e que se relacionam entre si.

De acordo com Lee et al. (2006), os sistemas BIM oferecem recursos que favorecem a representação e a visualização da edificação e que permitem a modificação dos elementos de forma direta e intuitiva. Eles garantem a centralização da informação, com a utilização de modelo único através de arquivos integrados e ligados *on-line*, e possibilitam que as atualizações sejam facilmente registradas pelos diversos profissionais. Assim, as modificações em um projeto específico, ou em uma parte do projeto, propagam automaticamente atualizações em outros projetos ou outras partes. Usado em sua plenitude, um modelo BIM deve ser integrado e consistente, e idealmente construído de maneira colaborativa entre todos os profissionais parceiros: arquitetura, estrutura e instalações.

Na associação entre os elementos da modelagem tridimensional e as suas relações semânticas, “partes do modelo podem ser representadas de diferentes maneiras pelos diferentes aplicativos dos atores do processo” (CORREA e SANTOS, 2014, p.7).

Essa é uma característica da tecnologia BIM, uma vez que o formato padrão de arquivo IFC - *Industry Foundation Classes* - não representa todas as classes de um empreendimento, já que considera os diferentes profissionais com os diferentes tipos de informações e tipos de representação no modelo - culminando em dificuldades na interoperabilidade e gerando retrabalho. Nesse sentido, as ontologias são propostas como a otimização do processo de modelagem da tecnologia BIM ao possibilitarem a integração semântica que poderá regulamentar as trocas de informações demandas no processo.

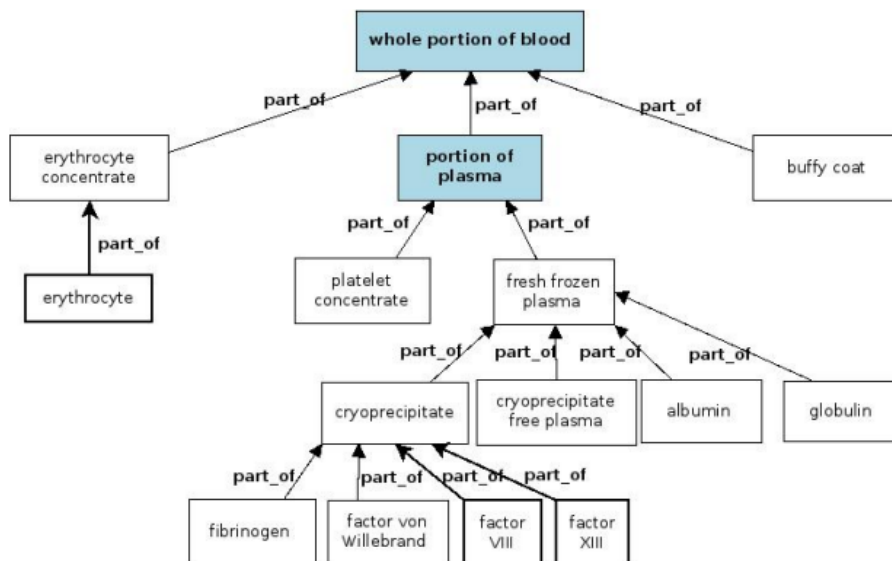
ONTOLOGIA E VOCABULÁRIO CONTROLADO

O tema ontologia foi e é estudado em várias áreas de conhecimento, com abordagens e objetivos diferentes, entre elas a filosofia, a ciência da computação e a ciência da informação, e com aplicações práticas nas mais diversas profissões como medicina, biologia, engenharia e arquitetura. Neste trabalho, ontologia diz respeito a uma representação de parte de determinado domínio, legível por máquina, para fins de representação e recuperação de informação, com possibilidades de inferência automática.

Grenon e Smith (2004, p.138) explicam que as ontologias podem ser definidas como um conjunto de entidades que existem no mundo, organizado através das categorias sob as quais essas entidades simbólicas pendem, e dos diferentes tipos de relações que são mantidas entre elas. Enquanto ferramenta de modelagem, Gassen (2014, p.13) atribui às ontologias os seguintes aprimoramentos, dentre outros existentes: interoperabilidade, validação, tradução, enriquecimento semântico, auxílio na heterogeneidade semântica, integração e assim por diante.


No contexto da AEC, a instânciação é um processo proveniente da modelagem gráfica (como já explicado na seção 2. 1), cujo termo se refere à existência real do objeto. A natureza classificatória das ontologias, ao possibilitar as relações “tipos de” e “partes de” e outras além dessas entre as instâncias e os seus conceitos e termos, confere para si a competência de validar como verdadeiro os compromissos ontológicos existentes entre eles (figura 3).

Figura 3 – Exemplo de identificação de componentes da uma ontologia



Fonte: Mendonça e Almeida (2013, p.169).

Figura 4 – Exemplo de consulta ao termo “brick cement”









 **brick cement** (<cement by property>, cement (construction material), ... Materials (hierarchy name))

Note: Waterproof cement used in masonry work.

Terms:
brick cement (preferred,C,U,English-P,D,U,N)
cement, brick (C,U,English,UF,U,N)
baksteencement (C,U,Dutch-P,D,U,U)
cemento hidrófugo (C,U,Spanish-P,D,U,U)

Facet/Hierarchy Code: M.MT

Hierarchical Position:

-  Materials Facet
-  Materials (hierarchy name) (G)
-  materials (substances) (G)
-  <materials by composition> (G)
-  inorganic material (G)
-  cement (construction material) (G)
-  <cement by property> (G)
-  brick cement (G)

Fonte: Art & Architecture Thesaurus® Online

A análise detalhada das entidades de um domínio, denominada “análise de princípios ontológicos”, pode ser utilizada como fonte de conhecimento para profissionais que lidam com a modelagem gráfica. Outrossim, considera-se a necessidade de se vincular a semântica e pragmática (Cabré, 2005) ao *frame* do projeto, para que se atenda e satisfaça à complexidade representativa e comunicativa do modelo em elaboração e/ou gerado.

O uso de um vocabulário controlado como base conceitual através da qual uma área do conhecimento está representada pode fundamentar a construção do conhecimento do empreendimento por meio da ontologia.

As linguagens documentárias são instrumentos utilizados para representar, organizar e conseqüentemente, para recuperar informação e conteúdo.

O vocabulário controlado é um tipo de linguagem documentária, assim como os sistemas de classificação bibliográfica, os cabeçalhos de assunto e os tesouros. Os tesouros são utilizados em sistema de informação para representar padronizadamente o conteúdo de documentos através de um controle terminológico (CAFÉ; BRASCHER; SUJII, 1990). Eles organizam um domínio do conhecimento através de seus termos, principalmente em relações hierárquicas (relação gênero/espécie - figura 4) e relações partitivas (relação todo/parte).

Além da dinâmica e fluidez da língua, o contexto da tecnologia BIM é caracterizado pela interdisciplinaridade em um mesmo projeto de empreendimento. Assim, os termos são passíveis de alteração em relação a situações geográficas, em relação às disciplinas que os utilizam, podendo-se existir mais de um conceito para um único termo, e ainda, necessitar a adequação semântica de tradução, dentre outras possibilidades.

A rigidez terminológica embutida em um tesouro busca alcançar uma linguagem comum entre os especialistas através da padronização, otimizando a comunicação naquela área do conhecimento. Na AEC, através da tecnologia BIM, o instanciamento que é gerado pelo banco de dados (como explicado na seção 2.1), naturalmente está relacionado a uma classe e traz consigo os seus atributos e o seu conceito. Entretanto, a padronização e normalização do termo não necessariamente reflete a especificidade da realidade de um empreendimento, e, portanto, demanda a sua construção colaborativa.

O desafio apontado para a gestão de informações em processos colaborativos, interdisciplinares e interoperáveis - características intrínsecas aos processos em BIM - é ressaltado na necessidade da adequação semântica das entidades (sejam materiais, objetos, representações). As incompatibilidades geradas pelas diferenças semânticas, conceituais e terminológicas das entidades do projeto resultam em dificuldades também na recuperação de informação e na comunicação entre os profissionais participantes.

INTERNET DAS COISAS (IOT)

A Internet das Coisas (IoT) conecta objetos e serviços físicos à internet, permitindo a comunicação entre objeto-objeto e entre pessoa-objeto. Para Yang e outros (2010, p. 358), “ao permitir novas formas de comunicação entre pessoas e coisas, e coisas entre si, a IoT adicionaria uma nova dimensão para o mundo de informação e comunicação da mesma forma que a internet fez anteriormente”. Para Lacerda e Lima-Marques (2015, p.159), a IoT “são sistemas, no sentido amplo, interligados entre si em diferentes escalas, formando ecossistemas com componentes biológicos, materiais, urbanos – tendo em comum a informação como substrato”.

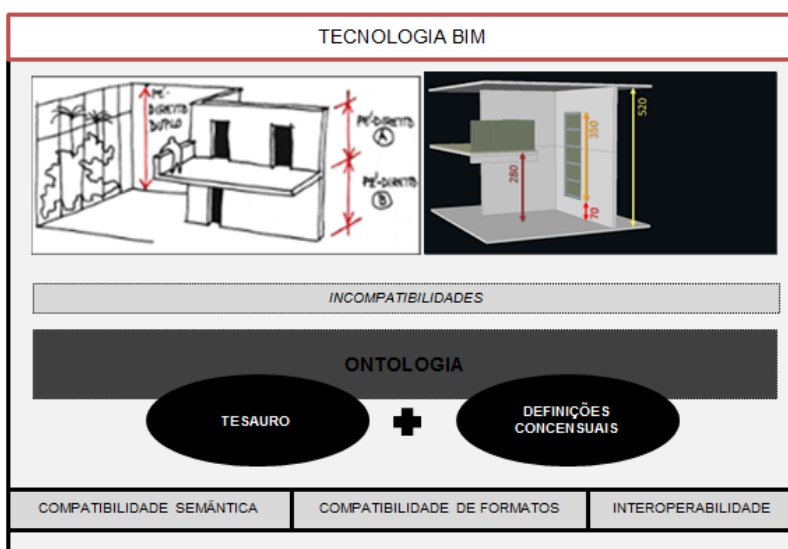
A IoT foi possível a partir do desenvolvimento e ampliação das capacidades das redes de comunicação e dos processadores e circuitos eletrônicos atuais. Os equipamentos, máquinas e produtos industriais que até então funcionavam isolados passam a ter capacidade de processamento e comunicação de dados, conexão, comunicação e acesso à internet embutidas. Assim, automóveis, telefones, televisores, geladeiras e outros tipos de objetos e pessoas, por meio de câmeras, sensores e diferentes tipos de medidas e monitoramento, passam a apresentar capacidade de sentir o ambiente, perceber seu estado e o de outros, além de intercambiar, solicitar, fornecer, delegar, gerenciar e trocar informações, possibilitando novos usos, personalização, automação e análises de desempenho.

Considerando a colocação de Cerp-IoT (2009), a IoT propõe uma rede de informação integrada, uma infraestrutura de rede dinâmica e global, baseada em protocolos de comunicação padronizados e interoperáveis nos quais as coisas físicas e virtuais tem identidade, atributos físicos e personalidades virtuais. Essa definição faz referência direta à necessidade de utilização de ontologias para a representação dessa informação na busca da padronização e interoperabilidade. Outro ponto importante é a utilização de interfaces inteligentes. Acrescenta-se a capacidade de autoconfiguração A com a possibilidade de aprendizagem e autodesenvolvimento.

Com a IoT espera-se que as coisas se tornem participantes ativas, insiram aprendizagem e inteligência, participem dos processos informacionais e sociais com a capacidade de interagir e comunicar com o ambiente por meio dos dados e informações recebidas desse próprio ambiente.

Na AEC, a integração da IoT e tecnologias BIM é tema recente de pesquisa, com pouca produção se comparado às produções em isolado. (BARACHO, PEREIRA JUNIOR E ALMEIDA, 2017). Para além do processo de projeto e construção das edificações, pode-se considerar determinado empreendimento, ou uma cidade com suas edificações, representadas tridimensionalmente por meio da modelagem de informações da construção (tecnologia BIM) e monitorada de maneira dinâmica (IoT), considerada a modelagem da informação de acordo com o espaço, o uso desse espaço, a hierarquia e relacionamento dos objetos com as ontologias. A IoT pode facilitar a implementação plena da tecnologia BIM, auxiliando na passagem das fases de projeto, planejamento e construção, atingindo integralmente os trabalhos nas fases de monitoramento, detecção de falhas, manutenção, gerenciamento de consumo, reforma, revitalização e até mesmo definição da necessidade de descarte de uma construção.

Figura 5 – Proposta de *framework*



Fonte: elaborado pelos autores

CONSIDERAÇÕES PRELIMINARES

O ATT - The Art & Architecture Thesaurus¹ é um tesouro que inclui os termos, as descrições e outros metadados a respeito de conceitos relacionados à arte, arquitetura, conservação, arqueologia e patrimônio cultural, incluindo ainda tipos de trabalhos, estilos, materiais, técnicas e outros. De acordo com Alexiev, Isaac e Lindenthal (2015), a maioria das relações entre os termos no referido tesouro são de TG (Termo Genérico), embora também existam TRP (Termos de Relação Partitiva - parte/todo), ressaltando ainda que essas relações são do tipo hierárquicas (generalização e especialização). Já as ontologias são capazes de especificar explicitamente a semântica dos termos de um domínio do conhecimento.

Essa característica possibilita sua utilização enquanto apoio à definição dos referidos termos naquele contexto - otimizando o processo de comunicação, além de apresentar uma solução a problemas de interoperabilidade semântica entre sistemas. Por isso, elas podem proporcionar a troca de informação entre sistemas e até mesmo entre pessoas (JASPER; USCHOLD, 1999). As relações que se estabelecem entre os termos em uma ontologia são resultantes do estudo do domínio.

¹ <http://www.getty.edu/research/tools/portal/index.html>

A vinculação da modelagem da informação da construção e do uso de IoT em AEC às ontologias e linguagens documentárias fundamenta a proposta de solução à geração de modelos representacionais semanticamente consistentes. O vínculo se justifica pela necessidade de estabelecer as relações entre as entidades de um contexto e vincular as mesmas às suas representações na modelagem gráfica incorporando as suas definições subsidiadas por um vocabulário controlado específico. Ou seja, utilizando-se do tesouro se dará subsídio à construção da definição consensual para o conceito na ontologia, onde também serão estabelecidos os tipos de relações associativas existentes entre os conceitos e respectivas descrições no contexto do empreendimento (figura 5).

RELATO DA EXPERIÊNCIA

A expectativa da prática está relacionada à utilização simultânea da modelagem gráfica e da modelagem conceitual resultando na criação de um modelo de empreendimento bem fundamentado, com correspondência executável no mundo real e documentada de forma terminologicamente adequada. Teixeira e Baracho (2017) propuseram um *framework* metodológico que será validado através de definições empíricas.

Para testar e validar o *framework* proposto foi seguido o seguinte roteiro metodológico:

1. Escolhido o termo em português e seu equivalente na língua inglesa.

TERMO	EM PORTUGUÊS	pé direito
	EM INGLÊS	ceiling height; headroom

2. Posteriormente, procedeu-se à pesquisa pelo termo em inglês no Tesouro.

DEFINIÇÃO PADRONIZADA “AAT”

Research
 Research Home > Tools > Art & Architecture Thesaurus > Search Results
 Art & Architecture Thesaurus® Online
 Search Results
 Find Name: **side wall**
 Logic:
 Note: 1 result
 Click the icon to view the hierarchy. Check boxes to view multiple records at once.

1. **fill papers**
 (<wall coverings by location or context>, wall coverings, ... Furnishings and Equipment (hierarchy name)) [300204057]
 sidewalls
 side walls
 walls- side

Research
 Research Home > Tools > Art & Architecture Thesaurus > Search Results
 Art & Architecture Thesaurus® Online
 Search Results
 Find Name: **headroom**
 Logic:
 Note: 0 results
 Your search has produced no results. Please refine your search by clicking on New Search.

Research
 Research Home > Tools > Art & Architecture Thesaurus > Search Results
 Art & Architecture Thesaurus® Online
 Search Results
 Find Name: **ceiling height**
 Logic:
 Note: 0 results
 Your search has produced no results. Please refine your search by clicking on New Search.

Research
 Research Home > Tools > Art & Architecture Thesaurus > Search Results
 Art & Architecture Thesaurus® Online
 Search Results
 Find Name: **ceiling**
 Logic: **AND**
 Note: **height**
 0 results
 Your search has produced no results. Please refine your search by clicking on New Search.

Research
 Research Home > Tools > Art & Architecture Thesaurus > Search Results
 Art & Architecture Thesaurus® Online
 Search Results
 Find Name: **wall**
 Logic: **AND**
 Note: **ceiling**
 4 results
 View Selected Records Select All Records Clear All
 Page: 1
 Click the icon to view the hierarchy. Check boxes to view multiple records at once.

1. **mural painting (image-making)**
 (painting (image-making), <painting and painting techniques>, ... Processes and Techniques (hierarchy name)) [300178677]
 wall painting (image-making)
 wall-painting (image-making)
2. **mural paintings (visual works)**
 (<paintings by location or context>, paintings (visual works), ... Visual and Verbal Communication (hierarchy name)) [300033644]
 painting, wall (object)
 wall-painting (object)
 wall painting (object)
 wall paintings (objects)
 wall-paintings (objects)
3. **wallboard**
 (building board, building materials, ... Materials (hierarchy name)) [300132980]
 board, wall
 dry-wall (wallboard)
 dry wall (wallboard)
 wall-board
 wall board
4. **wallpapers**
 (<wall coverings by form>, wall coverings, ... Furnishings and Equipment (hierarchy name)) [300037366]
 wall-paper
 wall-papers

Diante da indisponibilidade de resultados pelos termos originais, optou-se pela busca booleana. Ainda assim, buscou-se intuitivamente alcançar o conceito “do piso ao teto”, sem sucesso.

3. A terceira etapa consistiu em levantar de profissionais de diferentes formações e que atuam nos mesmos projetos dentro de uma mesma organização, as definições em linguagem natural para os termos em língua portuguesa.
4. Para a construção do termo para um empreendimento específico, utilizou-se como referência uma representação gráfica, e a partir da observação da imagem por ambas, dos seus conhecimentos e das definições levantadas, construiu-se a definição no contexto. No caso em específico, como não houve sucesso na pesquisa no ATT, realizou-se a busca outra fonte, para que exista uma referência padronizada.

DEFINIÇÃO EM LINGUAGEM NATURAL	
POR ARQUITETA 1	“Pé direito refere-se à altura do piso ao teto de um único pavimento.”
POR ARQUITETA 2	“Pé direito: altura do piso ao teto em um pavimento. Admite-se pé direito duplo ou triplo.”
POR ENGENHEIRA	“Na engenharia, usamos a expressão “pé direito” quando nos referimos a altura do piso até o teto.”

CONSTRUÇÃO DO TERMO NO EMPREENDIMENTO			
TERMO	REPRESENTAÇÃO GRÁFICA	DEFINIÇÕES UTILIZADAS	DEFINIÇÃO CONSENSUAL
EM PORTUGUÊS: pé direito		ATT: termo indisponível	<p>“Pé direito refere-se à altura do piso ao teto de um único pavimento.”</p> <p>Neste empreendimento, consideram-se então vários “pé-direitos”, uma vez que ele apresenta vários blocos, com vários andares e alturas diferentes.</p>
EM INGLÊS: ceiling height; headroom		ARQUITETA 1: “Pé direito refere-se à altura do piso ao teto de um único pavimento.”	
		ARQUITETA 2: “Pé direito: altura do piso ao teto em um pavimento. Admite-se pé direito duplo ou triplo.”	
		ENGENHEIRA: “Na engenharia, usamos a expressão “pé direito” quando nos referimos a altura do piso até o teto.”	
		DICIONÁRIO ESPECIALISTA: Diferença entre a cota do piso e a cota do tecto de uma casa. Diferença de cota entre o piso inferior e o piso superior, incluindo a espessura da laje superior.	

CONSIDERAÇÕES FINAIS

Identificado o *gap* semântico que resulta nas dificuldades de interoperabilidade e comunicação em processos de modelagem de empreendimentos em tecnologia BIM e IoT, foi proposto vincular diferentes metodologias de representação da informação e do conhecimento como solução.

Para tal, foi realizada a revisão de literatura dos temas, a saber: i) tecnologia BIM; ii) ontologias e vocabulário controlado; e iii) internet das coisas.

Os estudos teóricos apontaram as contribuições dos instrumentos de representação da informação, aqui os tesouros e ontologias, quando se assume a relevância das naturezas terminológicas (conceitos e relações), assertivas (axiomas aplicados aos conceitos e relações) e pragmáticas (conhecimento consensual) para o contexto da AEC, utilizando-se da plataforma BIM e IoT.

Com a aplicação da metodologia observou-se que a amplitude do ATT não cobre todos os termos e que uma alternativa deve ser disponibilizada para essas situações.

Buscou-se por outros tesouros em arquitetura, sendo encontradas duas iniciativas brasileiras: *Thesaurus Experimental de Arquitetura*, um projeto descontinuado da USP, e *Thesaurus BRASED*, também descontinuado. Como o ATT é o único em que se obteve resultado na pesquisa, optou-se por utilizar a definição do termo extraído de um dicionário da arquitetura² disponível na internet.

O levantamento de definições junto às especialistas iniciou-se de maneira bem curiosa. Ambas buscaram o entendimento global da pesquisa e demonstraram bastante interesse pelo conhecimento prévio em relação à tecnologia BIM e por haver consciência de algumas de suas deficiências em relação à classificação de objetos e adequação de objeto à especificação da engenharia. A elaboração individual das definições realizou-se bem rapidamente. Ao realizar a leitura dessas definições, alguns questionamentos em relação “ao que seria o *teto*” apareceram: é o telhado, a alvenaria ou o acabamento (forro)? Para elaborar a definição, tomou-se como referência um projeto de empreendimento para que pudessem se basear. Procedeu-se então à observação da imagem, à leitura das definições em linguagem natural e em seguida à leitura da definição apresentada pelo dicionário. Assim, após trabalho conjunto, chegou-se ao resultado final.

O processo demonstrou muita riqueza e reflexão das profissionais, e foi capaz de definir, em relação ao projeto em específico, o que de fato significava “pé direito” naquele empreendimento. Como se trata de um projeto com alturas diferentes, a definição deveria abarcar as duas possibilidades. Observa-se que um resultado na tecnologia, por exemplo, seria a não possibilidade de atribuir uma janela mais alta que o pé direito do imóvel no projeto.

Através do modelo bem referenciado, será possível minimizar inconsistências com a demanda original do desenvolvimento (cliente); maior assertividade na identificação, descrição e relacionamento das entidades utilizadas no empreendimento em seu domínio real; e sustentar a comunicação eficiente

entre diferentes áreas do conhecimento, diferentes línguas e conseqüentemente, diferentes denominações terminológicas dos mesmos conceitos.

A aplicação tecnológica do trabalho apresentado e validado empiricamente depende ainda de desdobramentos e suporte para desenvolvimento, o que desde o princípio não se configurou como um objetivo.

Considera-se aqui ontologia como a forma de representar os objetos e suas relações para ser utilizada em sistemas BIM e de Internet das coisas. Sistemas de Internet das Coisas como sistemas robustos que, por meio dos sensores, captam, transmitem e monitoram as informações dos ambientes ou coisas. E a tecnologia BIM, que possibilita a simulação das edificações em três dimensões com a interligação de informações construtivas ligadas a banco de dados

Finalmente, fica uma observação de que a união desses conceitos poderá trazer inovações para o processo, e a necessidade de desenvolvimento de pesquisas teóricas e aplicações práticas que podem confirmar a premissa estabelecida.

REFERÊNCIAS

ALEXIEV, V.; ISAAC, A.; LINDENTHAL, J. On the composition of ISO25964 hierarchical relations (BTG, BTP, BTI). *International Journal Digital Library*, v. 17, n. 1, p. 39-48, 2015. Disponível em: <<https://link.springer.com/article/10.1007%2Fs00799-015-0162-2>>. Acesso em: 29 abr. 2017.

BARACHO, R. M. A.; PEREIRA JUNIOR, M. L.; ALMEIDA, M. B. Ontologia, internet das coisas e modelagem da informação da construção (BIM): estudo exploratório e a inter-relação entre as tecnologias. In: ONTOBRAS, 2017. *Anais...* 2017.

CAFÉ, L. M. de M. de; BRASCHER, M. B. M.; SUJII, M. K. Elaboração de tesouros utilizando-se o programa de elaboração de tesouros em microcomputador. *Revista de Biblioteconomia*, v. 18, n. 2, p. 185-192, jun./dez. 1990. Disponível em: <<http://hdl.handle.net/123456789/339>>. Acesso em: 20 abr. 2014.

CAMPOS, M. L. A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. *Ciência da Informação*, v. 33, n. 1, 2004.

² <https://www.engenhariacivil.com/dicionario/dicionario-arquitetura>

- CORREA, F. R.; SANTOS, E. T. Ontologias na construção civil: uma alternativa para o problema de interoperabilidade com o uso do IFC. *Gestão e Tecnologia de Projetos*, v. 9, n. 2, p. 7-22, jul./dez. 2014. Disponível em: <<http://www.revistas.usp.br/gestaodeprojetos/article/view/69141/93381>>. Acesso em: 16 jul. 2017.
- EASTMAN, C. et al. *BIM handbook: a guide to building information modeling for owners, managers, designers, engineers, and contractors*. New Jersey: John Wiley & Sons, 2008.
- GRENON, P.; SMITH, B.; GOLDBERG, L. Biodynamic ontology: applying BFO in the biomedical domain. In: PISANELLI, D. M. *Ontologies in Medicine*. Amsterdam: IOS Press, 2004.
- INTERNET OF THINGS EUROPEAN RESEARCH CLUSTER - CERP IOT. *Internet of things: strategic reserach roadmap*. 2009. Disponível em: <http://www.internet-of-things-research.eu/pdf/IoT_Cluster_Strategic_Research_Agenda_2009.pdf>. Acesso em: 18 maio 2016.
- JASPER, R.; USCHOLD, M. *A framework for understanding and classifying ontology applications*. 1999. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.39.6456>>. Acesso em: 11 maio 2017.
- LACERDA, F.; LIMA-MARQUES, G. Da necessidade de princípios de arquitetura da informação para a internet das coisas. *Perspectivas em Ciência da Informação*, v. 20, n. 2, p. 158-171, 2015.
- LEE, G. et al. Specifying parametric building project behavior (BOB) for a building information modeling system. *Automation in Construction*, n. 15, p.758-776, 2006. Disponível em: <www.elsevier.com/locate/autcon>. Acesso em: 14 jul. 2017.
- MENDONÇA, F. M.; ALMEIDA, M. B. Um estudo de caso na construção de ontologias biomédicas: uma ontologia de domínio sobre hemoterapia. In: ONTOBRAS, 2013. *Anais...* 2013.
- PEREIRA JUNIOR, M. L.; BARACHO, R. M. A. Relações entre a gestão da informação e do conhecimento e uso de sistema BIM por arquitetos e engenheiros. In: SEMINÁRIO IBERO-AMERICANO ARQUITETURA E DOCUMENTAÇÃO, 4., 2015, Belo Horizonte. *Anais...* Belo Horizonte: [s.n.], 2015.
- TEIXEIRA, L. M. D.; AGANETTE, E. C. Ontologias no suporte à caracterização de processos uma proposta de estudo. In: CONFERÊNCIAS IBERO-AMERICANAS WWW/INTERNET E COMPUTAÇÃO APLICADA, 2016. *Actas...* 2016. P. 149-159.
- _____; BARACHO, R. M. A. Considerações ontológicas sobre modelar em AEC. In: ONTOBRAS, 2017. *Anais...* 2017.
- YANG, D.; Liu, F.; LIANG, Y. A survey of the internet of things. In: INTERNATIONAL CONFERENCE ON E-BUSINESS INTELLIGENCE, 2010. *Proceedings...* 2010.

Ontologia aplicada a um modelo de gestão organizacional: contribuições da ciência da informação

Cláudio Gottschalg Duque

Doutor em Ciência da Informação, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brasil.

Programa de Pós-Graduação em Ciência da Informação, Universidade de Brasília (PGCINF/UnB), Brasília, DF, Brasil.

<http://lattes.cnpq.br/8531105272766089>

E-mail: klauss@unb.br

Geraldo Gonçalves Bastos

Mestre em Gestão do Conhecimento e da Tecnologia da Informação, Universidade Católica de Brasília (UCB), Brasília, DF, Brasil.

Programa de Pós-Graduação em Ciência da Informação, Universidade de Brasília (PGCINF/UnB), Brasília, DF, Brasil.

<http://lattes.cnpq.br/8531105272766089>

E-mail: ggbdino@gmail.com

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

RESUMO

O impacto do volume, velocidade e variedade da informação a que ficaram expostas as organizações exige resposta inovadora para tomada de decisão, especialmente relacionada com a recuperação de informação e extração automática de conhecimento, que consolidam a importância da gestão de alto desempenho. Assim, o desafio é compreender as disciplinas que estão integradas no espaço informacional da gestão nas organizações, a fim de arquitetar soluções que permitam melhorias na tomada de decisão a partir de uma estrutura informacional madura, aqui denominada Gestão de Alta Performance (GAP). O GAP é uma estrutura informacional que suporta ontologias de domínio com o compromisso ontológico de determinar o modelo de plano de gestão para a organização, conforme o nível de maturidade informacional e de gestão. Um sistema foi desenvolvido para a construção de uma ontologia específica do domínio através das análises sintáticas e semânticas dos textos que compõem uma coleção de gestão organizacional. A ontologia, construída semiautomaticamente, gera conhecimento específico de domínio e propicia, através de sua rede semântica, as condições para que o gestor encontre efetivamente o modelo de plano de gestão personalizado para o negócio. Em vez de se basear em teorias estatísticas, propõe-se um sistema de recuperação de informação consolidado que utiliza a Teoria da Relevância. O modelo ontológico está arquitetado considerando a interdisciplinaridade, ponderando uma estrutura lógica que fundamenta as relações entre as disciplinas que compõem a ontologia a partir das suas propriedades, características e conteúdo.

Palavras-chave: Ontologia. Gestão organizacional.

Ontology applied to an organizational management model: contributions of information science

ABSTRACT

The impact of the volume, speed and variety of information to which organizations have been exposed requires innovative response to decision making, especially related to information retrieval and automatic knowledge extraction, which consolidate the importance of high performance management. Thus, the challenge is to understand the disciplines that are integrated in the informational space of the management in the organizations, in order to architect solutions that allow improvements in decision making from a mature information structure, here called High Performance Management (GAP). The GAP is an informational framework that supports domain ontologies with the ontological commitment to determine the management plan template for the organization, according to the level of informational and management maturity. A system was developed for the construction of a domain-specific ontology through the syntactic and semantic analyzes of the texts that make up a collection of organizational management. The ontology, built semi-automatically, generates domain-specific knowledge and provides, through its semantic network, the conditions for the manager to effectively find the business management model for the business. Instead of being based on statistical theories, a Consolidated Information Recovery System is proposed that uses Relevance theory. The ontological model is structured considering interdisciplinarity, pondering a logical structure that bases the relations between the disciplines that compose the ontology from the properties, characteristics and content of the same ones.

Keywords: *ontology. organizational management.*

Ontología aplicada a un modelo de gestión organizacional: contribuciones de la ciencia de la información

RESUMEN

El impacto del volumen, la velocidad y la variedad de información a la que se han expuesto las organizaciones requiere una respuesta innovadora a la toma de decisiones, especialmente en relación con la recuperación de la información y la extracción automática del conocimiento, lo que consolida la importancia de la gestión de alto rendimiento. Por lo tanto, el desafío es entender las disciplinas que se integran en el espacio informativo de la gestión en las organizaciones, con el fin de diseñar soluciones que permitan mejorar la toma de decisiones de una estructura de información madura, denominada Gestión de Alto Rendimiento (GAP). El GAP es un marco informativo que soporta ontologías de dominio con el compromiso ontológico de determinar la plantilla del plan de gestión para la organización, de acuerdo con el nivel de madurez informacional y de gestión. Se desarrolló un sistema para la construcción de una ontología específica de dominio a través de los análisis sintácticos y semánticos de los textos que conforman una colección de gestión organizacional. La ontología, construida de forma semiautomática, genera conocimiento específico de dominio y proporciona, a través de su red semántica, las condiciones para que el gestor encuentre eficazmente el modelo de gestión empresarial para el negocio. En lugar de basarse en teorías estadísticas, se propone un Sistema Consolidado de Recuperación de Información que utiliza la teoría de la Relevancia. El modelo ontológico está estructurado considerando la interdisciplinaria, ponderando una estructura lógica que basa las relaciones entre las disciplinas que componen la ontología de las propiedades, características y contenidos de las mismas.

Palabras clave: *Ontología. Gestión organizacional.*

VISÃO GERAL DO *FRAMEWORK GAP*

Uma das principais dificuldades que encontramos no âmbito das organizações é a tomada de decisão. O ato de administrar que envolve diversas decisões em seu curso, é fundamental para a perenidade das organizações, exige conhecimento especializado de seus praticantes e envolve um conjunto de informações que muitas vezes não estão à disposição dos gestores. Existem diversos exemplos de profissionais que, içados ao posto de comando, encontram dificuldades de transformar suas percepções sobre os caminhos da organização em prática. E muitos fracassaram, fracassam, e fracassarão sem mesmo perceber a importância do ato de gerir. Tampouco lhe atribuirão valor e necessidade de empenho em suas longas carreiras de gestores.

O GAP nasceu da percepção da dificuldade dos gestores de colocar em prática suas orientações e do entendimento de que essa dificuldade é inerente ao tipo de organização, ou modelo de gestão empregado, ocorrendo na maioria das vezes a partir do momento que um novo gestor é empossado, e passa a dar orientações a sua equipe. Outro aspecto motivador para o desenvolvimento do GAP é a dificuldade inerente que qualquer gestor tem para gerir, não interessa a experiência que tenha, ou mesmo sua formação. Fazer com que o time trabalhe em consonância com as orientações do gestor, seja ele de alta ou média gerência, é sempre um desafio para as organizações, e envolve riscos, ativos e capital que podem se desvanecer sem se identificar a real causa e motivação.

A epistemologia do GAP está baseada na perspectiva de que o sistema capitalista pode gerar riquezas de forma mais perene se houver efetivo aporte de conhecimentos que suportem o ato de gerir. Logo, as organizações, sejam elas privadas ou públicas, devem buscar o desenvolvimento do modelo capitalista em prol da distribuição de renda gerada pelo trabalho em organizações que sejam bem administradas, minimizando os insucessos inerentes ao empreendedorismo.

Conceber uma plataforma que permita minimizar esse impacto, direcionando o processo de planejamento da gestão e especificando as etapas de sua implementação, com mecanismos de monitoramento sobre o processo e o impacto da gestão são os resultados que o GAP se propõe a desenvolver na organização. Ressalte-se que a orientação da plataforma está voltada ao processo de implementação do modelo de gestão, contudo não há ruptura desse objetivo com os objetivos estratégicos organizacionais, vez que esses têm alta relevância para o desenvolvimento do modelo de gestão e, efetivamente, são fonte de suprimento para o desenvolvimento organizacional.

Assim, todo o *framework GAP* está suportado por um modelo ontológico que tem o compromisso de suportar a estruturação de um modelo de gestão com o uso das diversas disciplinas que envolvem o domínio da gestão. Essas disciplinas compõem uma base teórica sobre gestão fundamentada sob o aspecto acadêmico, arranjadas considerando a necessidade de seu conteúdo informacional para desenvolver planos de gestão.

Em uma estruturação diacrônica, o GAP revela a evolução dos termos ligados à gestão, buscando predicar a Gestão de Alta Performance como uma disciplina que trata do desenvolvimento dos recursos humanos para ampliar o potencial de sua participação nas organizações. Esse entendimento foi incrementado no GAP da perspectiva de predicar a Gestão de Alta Performance com argumentos da administração que pudessem consolidar o comportamento de gestão, e mais especificamente orientar o trabalho da média e alta gerência. Nesta perspectiva, o uso da expressão gestão para os predicados deve ser entendida como termo singular (ZUPPOLINI, 2014), que integra o sentido do sujeito Gestão de Alta Performance, sem ser o GAP, e como tal deve ser compreendido como complemento dos padrões de gestão que pretendemos orientar com o *framework*, sendo as disciplinas do modelo apresentado na figura 1, a seguir.

1. **Gestão do Conhecimento** - Estratégia de gestão administrativa que busca detectar o conhecimento organizacional, focos de desenvolvimento, processos de estruturação, armazenamento e utilização do conhecimento para construção de inovações que permitam melhorar os resultados da organização. Base fundamental para a gerência de alta performance, aborda também o uso das ferramentas de gestão do conhecimento para conseguir modelar o conhecimento em prol do desenvolvimento da alta performance;
2. **Gestão da Informação** - A estruturação da informação organizacional buscando estabelecer em que contexto a informação interfere nos processos organizacionais, na condução das atividades organizacionais e no desenvolvimento de conhecimento organizacional são os elementos necessários para o desenvolvimento da performance sistêmica da organização. Neste sentido, estudar os processos, o fluxo informacional, redundâncias e os elementos estruturantes da organização possibilitará identificar os fundamentos para um programa de gestão de alta performance;
3. **Epistemologia Organizacional** - Filosofia não é ciência nem teoria, mas sem filosofia não haveria ciência, e sem esta não haveria tecnologia (ROBREDO, 2008). A espiral hermenêutica do conhecimento, da inovação e da descoberta. Contextualização dos conceitos de informação e conhecimento em ambientes corporativos de alta competitividade. Outros conceitos relacionados: ecologia da informação, informação e conhecimento intramuros, fluxos informacionais e sistemas e subsistemas de informação e de comunicação. Certeza e probabilidade. Fontes de informação internas e externas. Vida útil da informação. A análise crítica da informação. Revisitação dos conceitos de conhecimento tácito, implícito e explícito. Centralidade do usuário. Vetores de inovação. Nichos de oportunidades.
4. **Planejamento Estratégico** - Tem por objetivo traçar as diretrizes e estratégias a serem adotadas em toda a organização. Para tanto, serão apresentados conceitos inovadores do processo de estabelecimento de estratégias considerando o conjunto de disciplinas que possibilitam caminhos para melhoria de performance. O estabelecimento do contexto da alta performance para a organização, a definição de indicadores, métricas e níveis de excelência orientados a partir dos conceitos de gestão do conhecimento e demais disciplinas do *Framework* GAP deverão consubstanciar todos os elementos necessários para orientar o processo de planejamento estratégico da organização.;
5. **Gestão da Inovação** - A gestão da inovação visa atingir o balanceamento adequado entre conhecimento disponível externa e internamente e o conhecimento em criação, isto é, criatividade. É a garantia de que o processo de resolução de problemas estará indo a favor das expectativas e da estratégia geral da organização (ZAWISLAK, 1996).
6. **Gestão de Agentes Complexos** - A análise de todas as forças que exercem tensão sobre as decisões organizacionais em face dos múltiplos agentes envolvidos no ambiente organizacional dificulta em muito a visão dos sistemas caórticos que interagem no contexto organizacional. Neste sentido, a disciplina buscar esboçar a formação dos arquétipos organizacionais que são fundamentados pela resultante de forças com complexa interação entre elas. A instrumentação das ferramentas de visualização dessas forças, definição dos elementos de interação de forças para estabelecer resultantes organizacionais mais performáticas é o tema central desta disciplina. Para tanto, serão analisados os multiagentes organizacionais, a potência das forças e dos sistemas arquetípicos no âmbito da empresa e as ferramentas de condução do potencial dessas forças para viabilizar a melhoria de performance organizacional;

7. **Gestão de Marketing** – Entendida como ferramenta competitiva que pode alavancar o resultado da melhoria de performance gerencial, estabelece-se em um processo social e gerencial pelo qual os atores organizacionais obtêm o que necessitam e desejam para desenvolver inovações de produtos e serviços, com a criação, oferta e troca de produtos de valor junto aos clientes (KOTLER, 1998). Estruturada com direcionamento dos quatro Ps (produto, preço, praça e promoção) propõe a estruturação da segmentação de mercado, a identificação do consumidor alvo e o melhor posicionamento para obter vantagem competitiva e fidelização do cliente, resultando em ganho de performance organizacional.
8. **Gestão da Sustentabilidade** – A responsabilidade social das organizações é tida como comportamento promotor de desenvolvimento sustentável, sendo a melhoria da performance da gestão, segundo Clapp (2005), resultante da conscientização socialmente responsável do papel da organização. O GAP compreende o papel social da organização e reforça em seu *framework* mecanismo de confluência do conceito no ato de gestão praticado pela média e alta gestão das organizações;
9. **Gestão de Talentos** - A gestão de pessoas por competências é um dos principais pilares da implantação da gestão de alta performance, visto que o êxito em todas as ações a serem realizadas pelos colaboradores depende basicamente de dois fatores: competências técnico-comportamentais devidamente mapeadas e alinhadas às competências organizacionais, e o comprometimento, sendo o último obtido através da boa gestão dos recursos humanos e aplicação de políticas meritocráticas;
10. **Governança Corporativa** - A identificação de elementos essenciais para o controle organizacional está prevista na disciplina de governança corporativa, sendo fundamental esclarecer que o objetivo principal é formular um processo de comunicação de informações financeiras que permita à sociedade reconhecer as organizações que estão efetivamente preparadas para atuar buscando a alta performance. O estudo da congruência das ações de gestão com as ações de governança corporativa está justificado, visto que desta forma é que efetivamente as empresas manifestam seu apreço para comunicação do desenvolvimento e do estabelecimento de metas de alta performance. As boas práticas de governança corporativa têm a finalidade de aumentar o valor da sociedade, facilitar seu acesso ao capital e contribuir para a sua perenidade contextos fundamentados pela estruturação da alta performance organizacional prevista no *framework* do GAP.
11. **Teoria dos Jogos** - A disciplina busca fundamentar o raciocínio estratégico dos alunos, visto que através dos jogos pode-se estabelecer a lógica entre os processos organizacionais e a intuição, a fim de propor soluções hipotéticas para circunstâncias do mundo real e de cenários futuros de integração das forças organizacionais. O conhecimento sobre a necessidade de desenvolvimento de uma visão de contexto do jogo organizacional permite aculturar a constante busca de uma visão sistêmica e o traço de estratégias para atingir os objetivos propostos no âmbito da organização e com respectivos *stakeholders*;
12. **Gestão da Qualidade** - A definição dos modelos de qualidade é aspecto essencial para o estabelecimento dos parâmetros de performance. A partir da avaliação da qualidade, definição do modelo de gestão de qualidade e análise dos distintos modelos de qualidade o aluno estará habilitado à identificação dos requisitos de qualidade organizacional. Em complementação à avaliação do histórico do conceito de qualidade, estruturação da qualidade organizacional, métricas e medidas de qualidade, ciclo de controle e auditoria da qualidade, inteligência dos indicadores de

qualidade e o ciclo de coordenação de ações são argumentos explorados nesta disciplina, que buscam suportar a fundamentação dos juízos de qualidade, confiança e tolerância organizacional.;

13. Ambiente Organizacional - O estudo do ambiente organizacional busca identificar os atores da organização que contribuem para estruturar o processo de gestão de alta performance. Agregamos neste contexto o estudo da comunicação organizacional, a informação para o negócio, definição de clima organizacional, psicossociologia das organizações e fatores que interferem no ambiente organizacional.

Esse espectro de disciplinas foi organizado de acordo com o compromisso ontológico e está arranjado em uma ontologia que permite a inserção incremental de novas disciplinas, todas considerando três relações fundamentais para a modelagem do conhecimento neste domínio:

a. Conceito disciplinar – Busca o estado da arte sobre o conceito da disciplina, revelando suas correlações com as demais disciplinas do *framework*, e permitindo popular a ontologia de acordo com a evolução científica do termo dentro do domínio da gestão organizacional. As disciplinas são tratadas como classes que agregam os termos que definem o domínio da Gestão de Alta Performance (GAP).

b. Interdisciplinaridade – O GAP considera que existe no âmbito das ciências sociais uma conexão horizontal entre as ciências envolvidas, sendo a informação o elemento basilar dessa percepção, vez que todas as ciências sociais, têm em seu objeto a informação em maior ou menor escala. A integração também ocorre sobre a perspectiva do objeto principal das ciências sociais, que é o ser humano e suas inter-relações. A organização, foco do GAP, é criada, mantida e desenvolvida pelos seres humanos que interagem entre si, com outras organizações e com seus clientes. Portanto, pensar sob o prisma da interdisciplinaridade é também uma inovação tecnológica proposta na modelagem do GAP, enquanto sistema de informação voltado para ampliar o potencial do ato de gerir.

c. Ferramentas da disciplina – A identificação das ferramentas, aqui entendidas como técnicas, métodos, processos, metodologias e *expertises*, todas consideradas para determinar o melhor arranjo entre as disciplinas, buscando a completude de ações necessárias para robustecer o ato de gestão. Importante observar que as ferramentas estão devidamente definidas na ontologia do GAP, sendo a fonte de informação que permite estabelecer, na prática, qual o melhor plano considerando o nível de maturidade da organização e dos gestores trabalhados com o *framework*.

Figura 1 – Plataforma de disciplinas do GAP

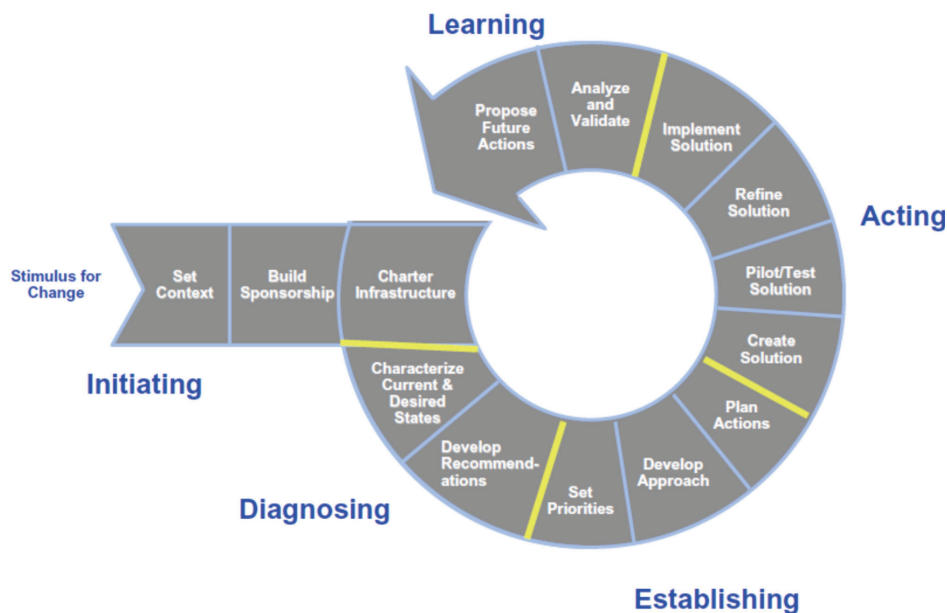


O *framework* do GAP utiliza a metodologia IDEAL (McFEELEY, 1996), desenvolvido pela SEI – Software Engineering Institute da Universidade Carnegie Mellon dos EUA, que é uma metodologia de melhoria contínua de processos. Mesmo considerando o alto nível de abstração apresentado na metodologia, buscamos traçar atividades práticas que são implementadas em todas as etapas de desenvolvimento da alta performance no GAP. O IDEAL, conforme figura 2, trabalha com as seguintes etapas macro:

- » *Initiating*: Preparação das bases necessárias para que o esforço de melhoria de processos seja bem-sucedido. Durante essa etapa o GAP utiliza instrumentos da Teoria dos Jogos para mapear as forças que interagem no ambiente organizacional, identificando *stakeholders* do projeto, áreas envolvidas, massa crítica, patrocinadores, multiplicadores e gestores.
- » *Diagnosing*: Levantamento do estado atual e definição do estado desejado. Nessa etapa o GAP compõe o mapa de processo a ser trabalhado, rotinas defensivas, lacunas informacionais, fontes de informação e o modelo “AS IS” – estado atual do macroprocesso de gestão.

- » *Establishing*: Planejamento detalhado de como alcançar as melhorias desejadas. Momento em que o GAP constrói uma matriz de responsabilidades, define gestor do projeto e integra com a gestão de agentes complexos os fluxos de relevância e motivação para alcance do objetivo do projeto de melhoria de gestão.
- » *Acting*: Execução do planejamento – Durante esta etapa são estabelecidos o arranjo ontológico de disciplinas necessárias à implantação do plano de gestão, e executam-se as ações necessárias para seu aprendizado organizacional.
- » *Learning*: Aprendizado adquirido durante o processo de melhoria. Na última etapa são detectadas e explicitadas as lições aprendidas, além de mapear os riscos e obstáculos encontrados. Toda a equipe envolvida é trabalhada com técnicas de *Coach* e Mentoria, para reduzir os impactos dos riscos e obstáculos e ampliar o potencial das vantagens do projeto.

Figura 2 – IDEAL – Iniciação, Diagnóstico, Planejamento, Ação e Aprendizagem



Fonte: GREMBA; MYERS (1997).

O GAP prevê o uso do IDEAL como ferramenta de planejamento para estabelecer um diálogo propositivo e de controle junto aos patrocinadores do projeto de melhoria da performance de gestão, contudo, utiliza a mesma ferramenta para desenvolver competências conversacionais em todas as atividades do projeto. Echeverria (1997) defende que o desenvolvimento de competências conversacionais permite o cultivo da tolerância entre o cliente e o fornecedor, em todas as atividades que necessitem de um ciclo coordenado de ações (CCA). Para o autor, a atividade humana é interdependente, e exige para sua efetiva conclusão a integração de ações de outras pessoas criando o que denominou Dança das Promessas.

Reconhecendo e valorizando o ato linguístico da Promessa, Echeverria, (1997), esclarece que toda relação para desenvolver algum objetivo vai requerer que os atores exerçam os papéis de cliente e fornecedor, inclusive permutando esses papéis até o término da atividade. A figura 3 explicita a interação e nos permite observar durante o desenvolvimento dos processos organizacionais todas as etapas descritas. O CCA se divide em quatro fases (BASTOS, 2003, apud ECHEVERRIA, 1997):

- a. Criação do Contexto – Na perspectiva de Creswell (2007), o indivíduo procura entender o mundo em que trabalha e vive, o que denominamos inquietude. Reconhecendo essa tendência, o GAP busca identificar os elementos relevantes e as inquietudes para a equipe que desenvolverá a atividade, criando um ambiente colaborativo com a técnica de Gestão de Talento, que permite o envolvimento gradativo do time de trabalho. Estabelece-se o contexto do projeto e como a atividade vai corroborar para construir o objetivo final.
- b. Negociação – Utilizando técnicas de negociação propositiva em ambiente colaborativo, o GAP prevê o estabelecimento de negociações que determinem o esforço dos atores envolvidos, a disponibilidade, a responsabilidade e o comprometimento do time. Nessa etapa é permitido recusar, contraofertar e adiar; contudo,

ajustado o processo de trabalho, o GAP passa a controlar efetivamente o desenvolvimento das tarefas derivadas e gerir com indicadores sua conclusão. Ao final dessa etapa é desenvolvida uma declaração de aceite e explicitada a promessa para cada membro do time;

- c. Realização - Momento de executar o que está planejado, inclusive de gerar novos ciclos de coordenação de ações com outras áreas e pessoas necessárias a cada tarefa, podendo ocorrer o cancelamento e revogação da mesma em casos não previstos, ou de urgência e emergência. Ao final dessa etapa será elaborada a declaração de cumprimento do que foi executado, esclarecendo inclusive o que não se concluiu, com os respectivos motivos;
- d. Avaliação – Realizada por todos os membros da equipe de trabalho, conta com um momento interno para entendimento das ocorrências de projeto, momento em que as lições aprendidas são armazenadas para permitir o aprendizado organizacional. No segundo momento, o trabalho de avaliação é realizado junto ao patrocinador da atividade, sendo analisadas reclamações e satisfações quanto ao que foi realizado. O resultado final dessa etapa é a declaração de satisfação, que é fonte de captação de indicadores de tolerância e inquietudes manifestadas para o desenvolvimento da confiança.

Figura 3 – CCA – Ciclo de Coordenação de Ações

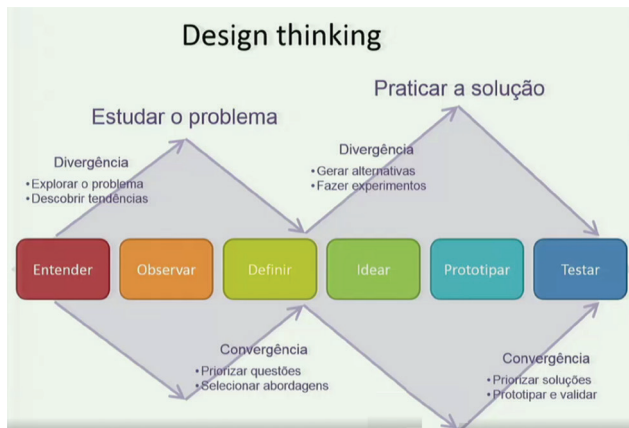


Fonte: BASTOS (2003)

Outro ferramental utilizado pelo GAP para consolidar o *framework* é o Design Thinking (DT), ferramenta do pensamento criativo que objetiva trabalhar de forma estruturada na resolução de problemas, na inspiração da criatividade e no desenvolvimento de inovações. Martin (2009) esclarece o conceito de DT como a atividade de formatar o contexto em que se insere determinado problema na organização, sem, contudo fazer parte do mesmo, corroborando Senge (2002) com formulação de uma visão sistêmica para o gestor. Assim DT exige interação, colaboração e abordagens práticas para encontrar novas ideias e soluções inovadoras.

A figura 4 apresenta o esquema do Design Thinking, que no GAP é utilizado como ferramenta para abordagem do processo de gestão nas unidades que serão trabalhadas. Para tanto, a equipe é convidada sob a perspectiva do processo definido no DT, e suas etapas são desenvolvidas com indicadores específicos que revelam, por fim, a quantidade de novas ideias, as inovações projetadas e o nível de participação e envolvimento dos atores.

Figura 4 – Processo de Design Thinking



Fonte: do autor

Durante a abordagem do problema da gestão na organização, são realizadas oficinas que objetivam permitir ao gestor expor a atual situação do processo de gestão, as dificuldades enfrentadas, lacunas de habilidades e competências, trabalho em grupo, visão sistêmica, integração com o planejamento

estratégico, nível de execução do plano de trabalho, e expectativas quanto ao trabalho que será desenvolvido no contexto do GAP. Durante as oficinas trabalhamos com o entendimento do problema, bem como observamos o mesmo nas atividades de acompanhamento das oficinas, para então conseguir definir a partir das divergências e convergências qual é realmente o problema, e o que será foco de nosso DT para propor melhorias, prototipar o processo com mentoria e testar a solução idealizada. Certamente as divergências e convergências nessa etapa serão tratadas para que possamos criar rotinas de ajuste e priorizar soluções criadas pelo grupo de trabalho.

Por fim, o *framework* do GAP se completa com a Ontologia de Gestão de Alta Performance, instrumento que permite popular o domínio de gestão tratado no projeto de melhoria da performance de gestão, e que é consolidado pela Ontologia de Gestão do GAP para identificar o conjunto de disciplinas necessárias à melhoria de gestão identificada, validade e homologada pelo patrocinador do projeto. Esse componente exige tratamento em separado, o que fazemos no item a seguir.

GESTÃO DE ALTA PERFORMANCE – MODELO ONTOLÓGICO

COMPROMISSO ONTOLÓGICO

O GAP – Gestão de Alta Performance é um modelo de gestão fundamentado na gestão do conhecimento desenvolvido em 2003 a partir de estudos promovidos por Bastos(2003) para melhorar a performance dos processos de gestão das organizações. Inicialmente foi ofertado ao Ministério dos Transportes para organizar processo de gestão da informação na Secretaria Nacional de Políticas de Transporte. À época possibilitou determinar um modelo de gestão para aquela secretaria, em especial para organizar as informações estratégicas de transporte modeladas em uma arquitetura na plataforma digital que derivou posteriormente em uma intranet da secretaria.

Outras incursões ocorreram em empresas públicas e privadas que permitiram maturar o modelo, fazendo com que se estruturasse em uma ontologia de domínio, e assim fosse utilizado para identificar o nível de maturidade gerencial da organização, seus objetivos estratégicos, o mapa estratégico, e por fim, o modelo de gestão necessário para atingir os objetivos organizacionais, tudo pautado em um modelo de monitoramento de indicadores que se traduz a partir das ferramentas, métodos e estratégias de gestão orquestradas pelo GAP para concepção dos resultados pretendidos.

O modelo sofreu maturação concomitante na execução de projetos acadêmicos para estruturação de cursos de especialização para formação de multiplicadores da metodologia, de tal sorte que um time de consultores, professores, mestres e doutores proporcionaram o robustecimento do GAP.

O compromisso ontológico é de agregar e estruturar disciplinas que podem orientar a formulação de um plano de gestão capaz de organizar a atividade de gestão em qualquer tipo de organização, a partir da interdisciplinaridade e do arranjo das ferramentas contidas no domínio, criando um fluxo diretivo adequado ao estágio de desenvolvimento da organização trabalhada. A população inicial se estabeleceu com disciplinas do domínio da gestão e da gestão de alta performance, e em seguida, buscou-se inserir relações com termos e ferramentas do domínio, todos predicados com as propriedades e funções no estado da arte da gestão.

MODELAGEM ONTOLÓGICA

A construção de um modelo ontológico que agregue disciplinas da gestão para consolidar o conceito de Gestão de Alta Performance deve seguir alguns princípios basilares que Gómez-Pérez et al (1967) definiram como essenciais no trato das ontologias:

» Clareza - É fundamental que a definição das classes ocorra com a utilização de termos que estejam devidamente definidos, evitando redundância, repetição e proximidade. Ao predicar um termo, deve-se observar que o sujeito da predicação não

se limite apenas à designação de uma instância do predicado, mas sim à designação de um tipo que esclareça todo o universo da aplicação do atributo.

- » Codificação objetiva - A conceituação deve ser especificada em um nível de conhecimento capaz de ser representado e codificado em linguagem formal (GOMÉZ-PÉREZ, 1967). O GAP foi modelado utilizando o Protégè, portanto, em um nível de codificação objetivo, que buscou representar as relações entre as entidades envolvidas no domínio da alta performance considerando conceito, interdisciplinaridade e ferramentas;
- » Extensibilidade - O *framework* está estruturado para permitir o incremento de novas disciplinas, estendendo o espectro da modelagem para se habilitar a integrar com novas tecnologias, métodos, práticas e disciplinas (GRUBER, 1993);
- » Foco no compromisso ontológico - Através do uso de um vocabulário especializado em gestão, o GAP estrutura os termos essenciais para o entendimento da alta performance (GRUBER, 1993). O compromisso ontológico deve ser perseguido durante o desenvolvimento da ontologia, contudo para sua evolução pode ser adequado considerando o uso de novos termos no domínio trabalhado.
- » Rigor terminológico - Os termos utilizados na modelagem ontológica do GAP buscam representar conhecimentos circunscritos ao domínio da gestão de alta performance, portanto, termos captados no vocabulário técnico da área.

A modelagem ontológica do GAP está fundamentada na lógica de primeira ordem (LPO), visto que outros diversos tipos de lógicas de representação do conhecimento podem se traduzir na lógica de primeira ordem (SOWA, 2000, p.42), e que ela é a melhor maneira de representar conhecimento a partir da linguagem natural. A LPO permite considerar que o mundo consiste de objetos com

certas relações entre eles que são válidas ou não válidas. Assim, a predicação, que busca qualificar o termo utilizando propriedades, relações e funções se faz de forma orientada, permitindo inclusive seu formalismo através de linguagem de representação.

Para tanto, o domínio da Gestão de Alta Performance foi classificado como um sistema, o qual interpretado nos permitiu observar as disciplinas de gestão do GAP. Essas disciplinas foram estruturadas em função de suas propriedades, relações e funções que se consolidam na expressão a seguir:

GAP(GC, AO, GQ, TJ, GCorp, GComp, GS, MK, GAC, Glnov, PE, EO, Glnf)	
onde:	GC - Gestão do Conhecimento;
	AO - Aprendizagem Organizacional;
	GQ - Gestão da Qualidade;
	GCorp - Governança Corporativa;
	GComp - Gestão de Competência;
	GS - Gestão da Sustentabilidade;
	MK - Marketing;
	GAC - Gestão de Agentes Complexos;
	Glnov - Gestão da Inovação;
	PE - Planejamento Estratégico;
	EO - Epistemologia Organizacional;
	Glnf - Gestão da Informação.

Sendo que para todo x do GAP temos uma disciplina do domínio de gestão,

$$\forall \alpha \Rightarrow \text{Disciplina}(\alpha);$$

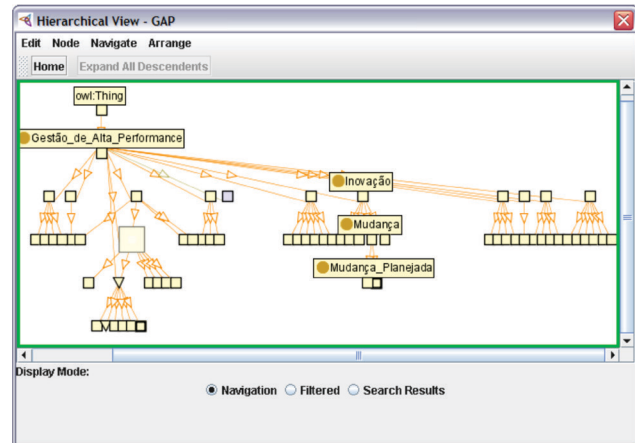
e que para qualquer das disciplinas sempre estaremos tratando da gestão,

$$\exists \text{GAP}(\alpha) \Rightarrow \text{Gestão}(\alpha);$$

$$\text{GAP}(\alpha) \Rightarrow \neg \forall \text{Gestão}(\alpha);$$

Assim, se estabelecem alguns dos axiomas por qualificadores da LPO, os quais orientam toda a modelagem ontológica do *framework*. O modelo foi formalizado no Protégè, sendo a figura 05 sua representação com o uso do applet Jambalaya.

Figura 5 – Ontologia do GAP

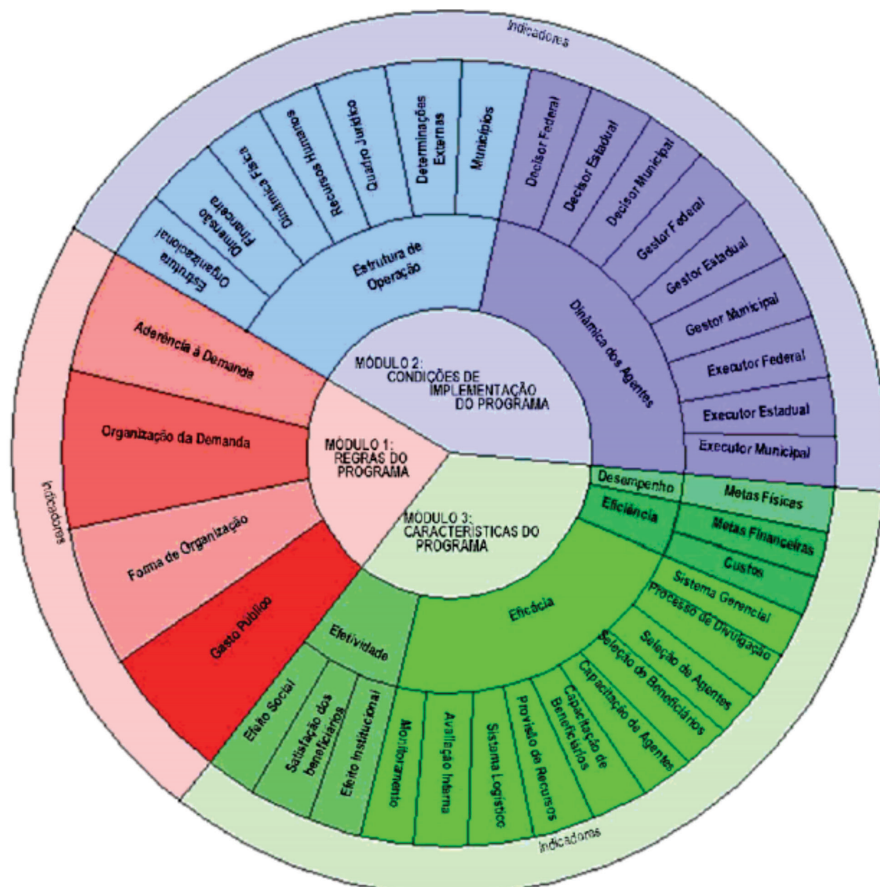


Fonte: dos autores

Entendida como um instrumento da Arquitetura da Informação (AI), a ontologia nos permite recuperar informações sobre o domínio da Gestão de Alta Performance, e com isso pavimentar a melhor estruturação de informações sobre gestão que podem objetivamente alinhar um plano de gestão para as organizações. A modelagem em AI proporciona, além da recuperação da informação, o entendimento, compartilhamento e reusabilidade dos esquemas de representação e das estruturas de armazenamento das informações e dos dados. (CARTAXO, 2016).

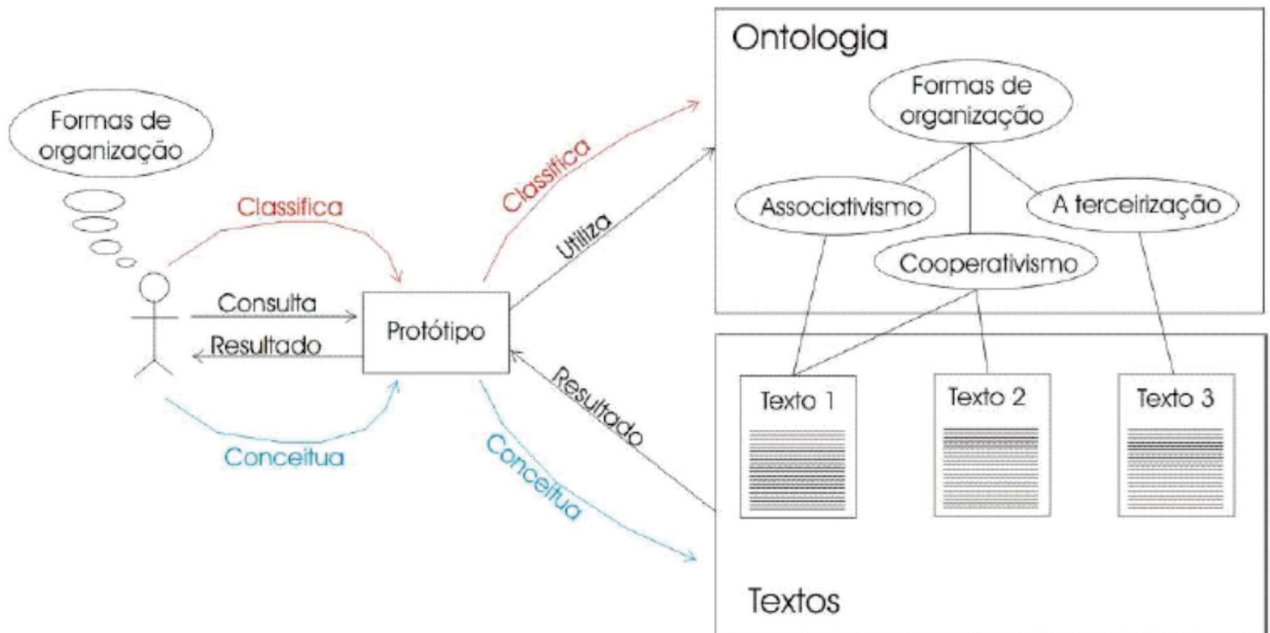
Nossas pesquisas sobre a importância da utilização prática da ontologia permitiram identificar no trabalho de Mori (2009) o uso de uma ontologia fraca para formar uma base de conhecimento sobre o Modelo Único de Monitoramento de Intervenções Governamentais no Campo Habitacional - SUMHAB (SILVA et al, 2005, apud MORI, 2009). Esse trabalho, descrito no próximo item foi fundamentado na dissertação de mestrado de Mori, publicada em 2009, que identificamos como uma manifestação do estado da prática de ontologias.

Figura 6 – Estrutura temática de monitoramento de intervenções governamentais no campo habitacional



Fonte: SILVA (2005, apud MORI 2009)

Figura 7 – Visão geral do processo de criação de ontologia de Mori (2009)



ONTOLOGIA NO ESTADO DA PRÁTICA

Mori (2009) observou que o SUMHAB define uma relação hierárquica direta dos módulos, dimensões e subdimensões modeladas como um disco temático, apresentado na figura 6, para classificação e consolidação dos indicadores de monitoramento das políticas públicas em habitação no Brasil. A ontologia nesse caso possibilitou a reunião terminológica sobre o mesmo domínio, permitindo a explicitação formal e consensual do conceito entre os envolvidos.

O projeto de Mori (2009) permitiu, em primeira fase, a alimentação de uma base de conhecimento com a recepção de termos no SiRILiCO (Duque, 2005) que indexaram termos através de uma ontologia leve. Na construção da ontologia, essa base de conhecimento foi comparada ao repositório de termos da área e validada sob o aspecto semântico. Essa base de conhecimento foi considerada por Mori como uma motivadora para utilização dos sistemas de monitoramento da área, uma vez que o processo de gestão se dá de forma mais aderente as necessidades dos gestores no uso da linguagem técnica da área de habitação.

A modelagem do conhecimento sobre políticas públicas de habitação, também denominada incremento da ontologia, é a etapa onde os termos de um texto, processo ou sistema de monitoramento são identificados e inseridos na base, tornando-a um repositório ontológico (MORI, 2009). Na pesquisa, esta etapa foi realizada utilizando-se de palavras-chaves, termos e conceitos da área, construindo assim uma estrutura temática do sistema de monitoramento de políticas públicas, reforça o autor. A figura 8 demonstra uma visão geral do trabalho, e esclarece as várias etapas de modelagem do conhecimento proposta na formação da ontologia.

Mori descreve que a construção da ontologia é dividida em três momentos, que se iniciam com a **Classificação** dos termos tendo como fonte os textos do domínio em estudo, o momento da **Busca**, que se revela com a utilização da base de conhecimento

para apoiar a recuperação da informação para monitoramento de políticas públicas, e, por fim, o momento da **Retroalimentação**, que ocorre quando a busca é infrutífera, e se permite ao usuário definir o conceito que não encontrou na base de conhecimento. Assim o usuário poderá inclusive criar seu indicador e conceituá-lo.

A segunda etapa é definida por Mori (2009) como uso da ontologia criada a partir da base de conhecimento. Nesta etapa, os indicadores são validados contra a base de conhecimento, sendo possível agregar valor ao processo de comunicação inerente ao monitoramento, em especial para gestão de políticas públicas de proporção continental, como as que ocorrem no Brasil.

Mori (2009) conclui sua pesquisa com o entendimento de que “ontologias como meio de compartilhamento de conhecimento são tecnologias importantes... para redução do impacto da variação terminológica entre os vários gestores” de políticas públicas, além de consolidar um modelo de disseminação de conhecimento.

O GAP exige, em função de sua experimentação, a população de lições aprendidas, que construíram um entendimento interdisciplinar de planejamento da gestão nas diversas verticais de mercado. Para se manter extensivo foi necessário popular o modelo com um conjunto de novas informações, adquiridas durante o uso do modelo, o que ensejou a necessidade de ferramentas capazes de consolidar a ontologia. A construção e manutenção semiautomática do modelo é parte essencial para o desenvolvimento do GAP, e como tal será mais bem esclarecida nos próximos tópicos.

PRODUÇÃO SEMIAUTOMÁTICA DA ONTOLOGIA DE GESTÃO

O uso da ontologia, como uma concepção formal, pode explicitar e compartilhar um domínio evitando problemas de ineficiência de interpretação, sobrecarga de ruídos, ambiguidades e caos na comunicação e no desenvolvimento de tarefas. Duque e lobin (2004) ressaltam que o

uso de analisadores automáticos semânticos pode extrair a estrutura conceitual, descrever frases e usar a semântica entre termos e relacionamentos para estabelecer conexões entre eles na busca de conteúdo sobre determinado domínio.

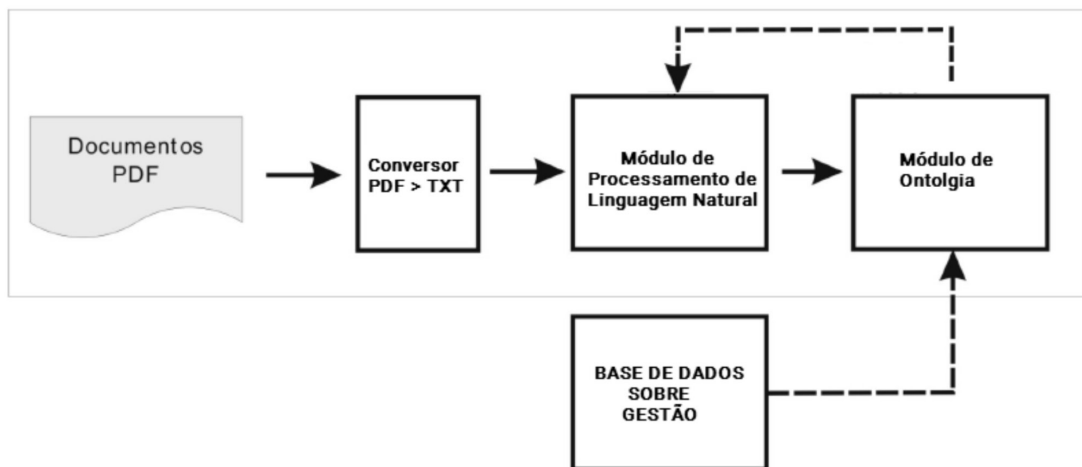
Entretanto, mesmo com o uso da tecnologia, a melhoria obtida por uma ontologia sempre revela dois problemas: ambiguidade e inconsistências. A ambiguidade dos termos foi abordada como um dos principais problemas na recuperação de informações em banco de dados (HAN et al., 2005). Os autores constataram que “por causa de variações dos termos, termos idênticos, termos com erros ortográficos, ou pseudônimos, a ambiguidade e a inconsistência foram detectadas em suas pesquisas”.

A solução foi abordada de várias maneiras e sempre foi relacionada à criação de arquivos de autoridade. Auld, *apud* French et al (1997), enfatizou que esse tipo de estratégia é denominado “trabalho de autoridade” e necessita de procedimentos computacionais. Eles destacam a abordagem de Taylor (1997), cujo primeiro princípio de controle de autoridade está preocupado com todas as variantes de um termo, as quais são explicitadas em um único formulário, compreendidas por usuários como a forma única para o termo, aquele que aglutina todas as possibilidades de interpretação do termo. Esta “forma única” foi denominada por Taylor (1997) “nome canônico”.

O processo de construção da forma única resolve o problema da ambiguidade, contudo seu procedimento não é automatizado, vez que exige avaliar as diversas interpretações para formatar a forma única do termo em um domínio. Assim, caracterizamos a identificação do nome canônico como semiautomática, e trabalhamos para construção da ontologia com o modelo desenvolvido por Duque (2005), denominado SiRILiCO, que já contempla a criação da forma única dos termos do GAP, buscando definir termo canônico da gestão de alta performance, base fundamental para construção e manutenção da ontologia do GAP.

O modelo desenvolvido por Duque (2005), conforme figura 6, consiste na conversão de arquivos PDF encontrados na internet, desenvolvidos em linguagem natural, captados no domínio da gestão de alta performance, que são processados para criação da ontologia do GAP. O processamento ocorre nos submodelos sintático e semântico, que conta com 3 (três) sistemas: Syntactic Parser, GeraOnto e Protegé.

Figura 8 – Modelo do SiRILiCO



Fonte: Duque (2005)

Nossa proposição com este trabalho é utilizar o SiRILiCO (DUQUE, 2005) para construir as ontologias de trabalho especializadas para organização em análise, que seja comparada com a ontologia de gestão do *framework* GAP, derivando em um plano de trabalho especializado para a organização. Assim, devemos observar que os documentos em PDF são necessariamente documentos da organização em estudo, e que a base de dados sobre gestão é a base de dados do GAP, suportada por sua ontologia aqui descrita.

Com o uso do SiRILiCO, devemos procedemos do seguinte modo para estruturar a ontologia de comparação da organização para proceder a integração com a ontologia do *framework* GAP, na busca de um modelo de gestão para organização, que contemple as disciplinas necessárias, as correlações interdisciplinares e as ferramentas que devem ser utilizadas para projetar a solução de gestão da organização.

- » análise da base de dados da organização, dos documentos relacionados ao planejamento e ao processo de gestão, bem como de todos os documentos que tratem de processos de negócio e necessidades de informação;
- » tratamento das informações de forma semiautomática para identificar a arquitetura informacional multimodal necessária ao atendimento das inquietudes levantadas e identificadas como de relevância;
- » validação do modelo informacional para atendimento multimodal de todos os requisitos levantados, considerando a possibilidade de ajustes em função da redução de redundâncias e ambiguidades identificadas junto ao corpo gestor da organização;
- » estruturação de rotinas de tratamento e carga para construção da base de dados que represente o modelo homologado. Nessa etapa podemos utilizar a infraestrutura de TIC da organização para consumir webservice que permita manter a integridade referencial dos termos extraídos da

linguagem natural dos textos e bases de dados da organização;

- » identificação do nome canônico de todos os termos relacionados ao processo de gestão da organização, seguindo as etapas:
 - o converter todos os arquivos pdf em arquivos txt;
 - o enviar os textos (desde a introdução até antes das referências) para um analisador sintático (Syntactic, Parser - VISL), que execute automaticamente a análise e gere uma árvore sintática com todas as TAGs sintáticas;
- » comparação sintática e semântica da forma única, resultado do conjunto de termos canônicos do SiRILiCO para extração das disciplinas, ferramentas e interdisciplinaridade do GAP aderente aos requisitos da organização consolidados em linguagem natural da fontes envolvidas;
- » desenvolvimento da ontologia de atendimento específico aos requisitos organizacionais do caso em tratamento. Essa ontologia deve ser populada considerando todo o trabalho de identificação da relevância, de remoção dos dois erros primários da ontologia, e da efetiva validação realizada pelo corpo diretivo da organização;
- » com base na ontologia especializada nos requisitos organizacionais, desenvolvemos um plano de trabalho para implementar o modelo de gestão na organização, que deve estar suportado por acompanhamento de todos os indicadores determinados e específicos para cada etapa do plano de trabalho.
- » acompanhar a implantação do modelo de gestão a partir do framework do GAP, buscando captar e validar todos os indicadores, e prospectando resultados futuros em relação às etapas realizadas para permitir melhor decisão, junto aos patrocinadores, quanto ao ritmo de implantação das ondas de desenvolvimento do plano de trabalho.

CONSTRUCTOS PARA O GAP

Para o desenvolvimento da base de conhecimento sobre gestão na organização trabalhada com o GAP, será necessário proceder conforme a figura 8, observando a necessidade de desenvolvimento de cada uma das etapas, vez que o conhecimento estabelecido na organização é a fonte inicial de trabalho para um novo plano GAP de maturação da gestão.

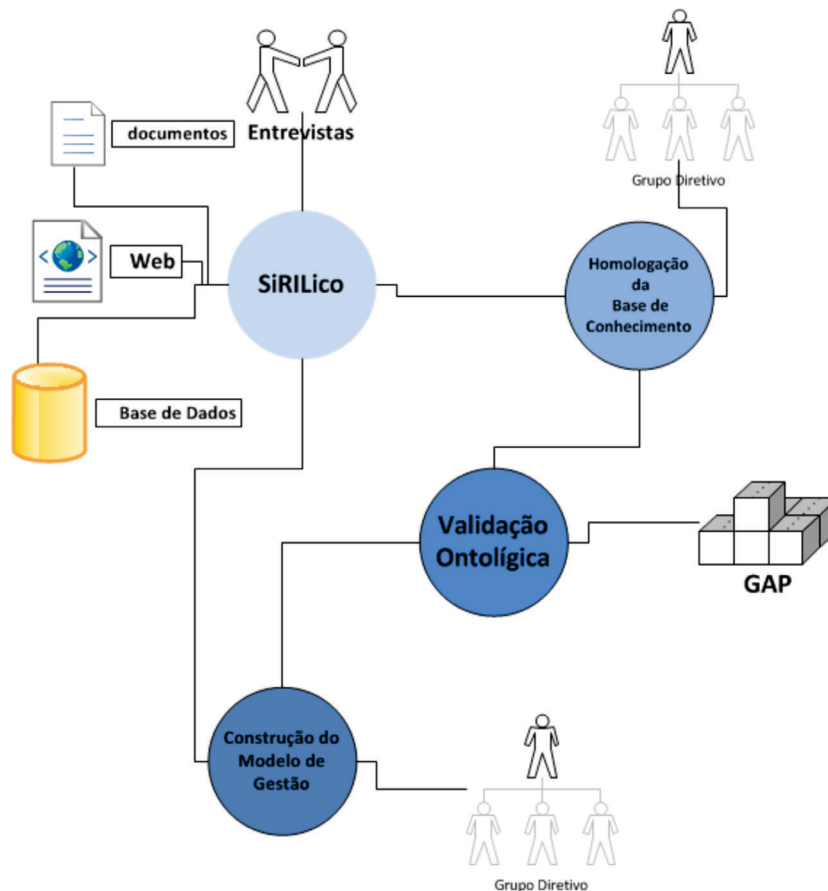
Nesse sentido, seja qual for o tipo de organização, privada ou pública, é a partir dos textos relativos ao processo de gestão, dos documentos relacionados aos processos, bem como de fontes relacionadas ao negócio da organização que partimos para a fase de construção da ontologia leve de gestão. Esse trabalho é consolidado a partir de entrevistas junto ao corpo diretivo da organização, captura de termos em documentos na Web e em bases de dados dos sistemas corporativos.

Todas as fontes são consolidadas com o SiRILiCO, formando uma base de conhecimento sobre o modelo de gestão da organização em estudo.

Na fase de homologação os gestores patrocinadores e do corpo diretivo participam de oficinas de homologação da terminologia, buscando sempre o consenso e a definição clara de conceitos importantes para o negócio da organização.

A base de conhecimento é validada junto ao GAP, momento em que todas as disciplinas se interrelacionam e formam um domínio sobre gestão que será ponto de origem para todo o planejamento de gestão da organização. Aqui são estabelecidos todos os indicadores de controle, tanto para implantação do plano, como para identificação de necessidades de ajustes, controle de rota, necessidades de treinamento, ou ainda mudança de processo de negócio.

Figura 8 – Processo de população do GAP com SiRILiCo



O modelo de gestão derivado do processo anteriormente descrito é apresentado ao corpo diretivo, e são estabelecidos os planos de implantação, bem como as ondas de impacto de gestão necessárias ao ajuste do modelo de gestão com o menor impacto para o desenvolvimento dos negócios. Esse trabalho final exige monitoramento constante, sendo importante apontar as necessidades de tomada de decisão ao corpo diretivo, permitindo-lhe reduzir impactos negativos e riscos ao negócio.

CONCLUSÃO

O uso de uma ferramenta semiautomática para geração de ontologia foi identificado como um item importante para compor o *framework* do GAP, em especial porque agiliza o processo de planejamento para desenvolvimento da alta performance, e porque permite uma comparação referencial entre modelos de mesma natureza, a saber a ontologia do GAP e a ontologia de gestão da organização em estudo.

A prática desse modelo certamente vai permitir aprimorar o resultado da aplicação do GAP na construção de solução de alta performance em gestão, fazendo com que seus resultados possam permitir, no futuro, a análise das ontologias registradas para identificar via sistemas autônomos quais as informações podem ser generalizadas na estruturação de novos axiomas, classes e relacionamentos para o desenvolvimento da performance de gestão em qualquer tipo de vertical do mercado.

Para pesquisas futuras, devemos considerar que a base dados das ontologias poderá ser estruturada como um modelo automático de geração de novas informações sobre gestão, que associadas às inovações nesse domínio poderão robustecer o *framework* do GAP, transformando-o em um modelo de arquitetura da informação capaz de dar respostas às futuras necessidades do emprego de capital nas economias da era da informação.

REFERÊNCIAS

- ARAÚJO, Filipe da Silva. *Comportamentos informacionais e identificação de ativos: um estudo de caso em uma organização militar*. 2011. Monografia (Especialização)– Universidade de Brasília, 2011.
- CARTAXO, M. A. *A contribuição da arquitetura da informação para gestão do conhecimento*. 2016. Tese (Doutorado)- UNB, Brasília, 2016.
- CRESWELL, J. W. *Projeto de pesquisa: métodos qualitativo, quantitativo e misto*. 2. ed. Porto Alegre: Artmed, 2007.
- ECHEVERRIA, R. *Ontologia del lenguaje*. 4. ed. Santiago do Chile: Dolmen Ediciones, 1997.
- GOMÉZ-PÉREZ, A. *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. Madrid, Espanha: Springer, 1967.
- GREMBA, J.; MYERS, C. The IDEALSM model: a practical guide for improvement. *Bridge*, n. 3, 1997.
- GRUBER, T. R. Towards principles for the design of ontologies used for knowledge sharing. In: INTERNACIONAL WORKSHOP ON FORMAL ONTOLOGY IN CONCEPTUAL ANALYSIS AND KNOWLEDGE REPRESENTATION, 1993, Itália. *Proceedings...* Itália: [s.n.], 1993.
- KOTLER, P. *Administração de marketing: análise, planejamento, implementação e controle*. 4. ed. São Paulo: Atlas, 1998.
- MCFEELEY, B. *IDEAL: a user's guide for software process improvement*. Pittsburg: PA, 1996. Technical Report CMU/SEI-96-HB-001.
- MORI, A.; DUQUE, C. G., SILVA, M. F. S. *Ontologias na modelagem de conhecimentos de políticas públicas habitacionais*. [S.l.: s.n.], 2009.
- ZAWISLAK, P. A. Gestão de tecnológica para empresas de tecnologia estabilizada. In: SIMPÓSIO DE GESTÃO DA INOVAÇÃO TECNOLÓGICA, 19., 1996, São Paulo. *Anais...* 1996.
- ZUPPOLINI, Breno Andrade. *Aspectos formais e ontológicos da filosofia da ciência de Aristóteles*. 2014. 159 f. Dissertação (Mestrado)- Universidade Estadual de Campinas, Instituto de Filosofia e Ciências Humanas, Campinas, 2014.

Ontological approach to the normative dimension of organizations: an application of Documents Acts Ontology

Maurício B. Almeida

PhD., Associate Professor of the Graduate Program in Knowledge Organization & Knowledge, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Lattes: <http://lattes.cnpq.br/5218069708058487>

E-mail: priv.mba@gmail.com

Mathias Brochhausen

PhD., Associate Professor of University of Arkansas for Medical Science (UAMS), Little Rock, Arkansas, United States

<https://uams-triprofiles.uams.edu/profiles/display/126735>

E-mail: MBrochhausen@uams.edu

Fabiana Bigão Silva

PhD Candidate, Graduate Program in Knowledge Organization & Knowledge, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.

Lattes: <http://lattes.cnpq.br/5144721555999622>

E-mail: fbigao@gmail.com

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

ABSTRACT

Modeling and knowledge representation are indispensable activities for developing information systems in the scope of modern corporations. In order to accomplish these activities, one should analyze a corporation both in terms of its physical structure and in terms its rules structure. The former we call here the descriptive dimension and the later prescriptive dimension. In this paper, we briefly discuss the descriptive dimension, and focus on the prescriptive dimension to explain the rights and obligations that corporations have to manage. After presenting background theories, we analyze the corporation through ontological principles taking advantage of theories of the so-called social ontology, namely, social acts, speech acts and document acts theory. The relevance of developing such analysis rests on the possibility of outlining a strong understanding of corporations by characterizing the nature of rights and duties obligations connected to corporative processes using Basic Formal Ontology (BFO) and Documents Acts (D-acts) Ontology. In doing this, we introduce a formal framework suitable to be applied in information systems working in the context of modern technologies like the Semantic Web.

Keywords: Corporations. Knowledge representation. Ontology. Social ontology. Document Acts.

Abordagem ontológica à dimensão normativa das corporações: uma aplicação da ontologia Atos dos Documentos

RESUMO

Modelagem e representação do conhecimento são atividades indispensáveis para o desenvolvimento de sistemas de informação nas corporações modernas. Para executar essas atividades, é preciso analisar a corporação tanto em termos de sua estrutura física quanto em termos de sua estrutura de regras. O primeiro caso é aqui denominado dimensão descritiva, e o segundo caso dimensão prescritiva. Neste artigo, aborda-se brevemente a dimensão descritiva, enfatizando a dimensão prescritiva para explicar os direitos e as obrigações que as corporações devem gerenciar. Após apresentar um background de teorias necessárias para o entendimento, analisa-se a corporação através de princípios ontológicos, a partir das teorias da assim chamada ontologia do social, em particular, atos sociais, atos da fala e atos dos documentos. A relevância em desenvolver tal análise diz respeito à possibilidade de melhor entendimento das corporações ao se caracterizar a natureza dos direitos e das obrigações ligadas aos processos corporativos usando a Basic Formal Ontology (BFO) e a Document Acts (D-acts) Ontology. Assim, introduz-se um esquema formal adequado para aplicação em sistemas de informação no escopo de modernas tecnologias como a Web Semântica.

Palavras-chave: Corporações. Representação do conhecimento. Ontologia. Ontologia do social. Atos dos Documentos.

Enfoque ontológico a la dimensión normativa de las corporaciones: una aplicación de la ontología de Actos de Documentos

RESUMEN

Modelaje y representación del conocimiento son actividades indispensables para el desarrollo de sistemas de información en las corporaciones modernas. Para ejecutar esas actividades, es preciso analizar la corporación tanto en términos de su estructura física como también en términos de su estructura de reglas. El primer caso es aquí denominado dimensión descriptiva y el segundo caso dimensión prescriptiva. En ese artículo, se aborda brevemente la dimensión descriptiva, enfatizando la dimensión prescriptiva para explicar los derechos y las obligaciones que las corporaciones deben gestionar. Después de presentar un background de teorías necesarias para el entendimiento, se analiza la corporación a través de principios ontológicos, a partir de las teorías de la así llamada ontología de lo social, en particular, actos sociales, actos del habla y actos de los documentos. La relevancia en desarrollar tal análisis se refiere a la posibilidad de un mejor entendimiento de las corporaciones al caracterizarse la naturaleza de los derechos y de las obligaciones ligadas a los procesos corporativos usando a Basic Formal Ontology (BFO) y la Document Acts (D-acts) Ontology. De esa forma, se introduce un esquema formal adecuado para aplicación en sistemas de información en el ámbito de modernas tecnologías como la Web Semántica.

Palabras clave: Corporaciones. Representación del conocimiento. Ontología. Ontología social. Actos de Documentos.

INTRODUCTION

In order to overcome issues of modeling and knowledge representation, ontologies have received more and more attention for the purpose of developing well-founded models for information systems (SMITH, 2003; GUIZZARDI, 2005; FONSECA, 2011).

Ontology is a term originated in Philosophy and it is now also used to denote a structure in form of an inverted tree, composed by entities and respective relations, used for purposes of knowledge representation. In the ontology development community, in particular in the healthcare arena, there still exists a debate about the nature of organizations, such as corporations. The team developing the Ontology for Biomedical Investigations (BANDROWSKI et al., 2016) defines corporation as:

An entity that can bear roles, has members, and has a set of organization rules. Members of organizations are either organizations themselves or individual people. Members can bear specific organization member roles that are determined in the organization rules. The organization rules also determine how decisions are made on behalf of the organization by the organization members.¹

The creators of OBI comment:

It [the definition] leaves open if an organization is a material entity or a dependent continuant, as no consensus was reached on that. [...] It is debatable what the organization itself is (some kind of dependent continuant or an aggregate of people). [...] favorite is still to define organization as a kind of 'legal entity' [...].²

This paper presents results of an on-going investigation of how corporations should be representing following the ontological principles of the Basic Formal Ontology (BFO) (SMITH, 2015).

In order to reach our purposes, we present a background section that consists of two parts: first, we present a historical overview describing several theories concerning the nature of corporation; second, we present background information about

BFO and its extensions such as the Document-acts Ontology (d-acts) (BROCHHAUSEN; ALMEIDA; SLAUGHTER, 2013), which are the starting points of our ontological analysis. After the background section, we briefly describe the descriptive dimension of corporations, addressing their structure and their units.

The focus of this paper is the representation of normative lattice that holds the corporation together. We are analyzing the corporation from the point of view of their claims, duties and obligations that exist within itself and which are support its existence. To address the normative perspective, we use pre-existing social ontology approaches, including theories of social acts (REINACH, 2012), speech acts (SEARLE, 1969) and document acts (SMITH, 2012).

The relevance of this analysis rests on the possibility of improving the understanding of corporations by a better understanding the activities that take place within them that are based on the normative backbone of the organization. The resulting framework can be applied in the context of Semantic Web Technologies to implement computer-parsable representations that can contribute to a higher level of automation in managing information about corporations.

BACKGROUND: THE NATURE OF CORPORATIONS

In this section, we present a brief summary of the theory of corporations from the history of Law and Economics. These two fields are generally considered the authoritative sources for defining corporations. Considering these theories is useful to understand the ontological commitment of each one of them.

THE NATURE OF THE CORPORATION IN LAW THEORIES

Corporations are a product of Roman law. Ancient Rome had already elaborated the notion of corporation including most of the legal attributes one can find in modern corporations.

¹ Retrieved August 23, 2017 from: http://www.ontobee.org/ontology/OBI?iri=http://purl.obolibrary.org/obo/OBI_0000245

² Retrieved August 23, 2017 from: http://www.ontobee.org/ontology/OBI?iri=http://purl.obolibrary.org/obo/OBI_0000245

Thus, a substantial understanding of the nature of corporations had already existed when the first English business corporations were chartered in the 15th century (WILLISTON, 1888).

Within the English legal system, a corporation had the following core attributes: i) it is a legal unit with its own legal rights and responsibilities; ii) it is distinct from the individuals who are members constituting it over time; iii) it is a creation of law and may achieve legal status by an act of the state. These core attributes, plus some additions, were applied in the United States assigning the corporation with the following legal attributes: i) it can contract, sue and be sued; ii) it can acquire and dispose of property; iii) it has its own seal by which it could act as a body distinct from its members; iv) its shares are transferable; v) its membership may change without affecting its perpetual existence; vi) it cannot commit treason; and, vii) it cannot serve as a trustee (BLACKSTONE, 2016).

The classical formulation of this set of attributes has been referred to as the “artificial person” view of corporations, which is merely one amidst several theories that explain the notion of corporate personality (DEWEY, 1926; KOESSLER, 1949; BLUMBERG, 1993; BLUMBERG, GEORGAKOPOULOS, STRASSER, GOUVIN, 2004; MALCOLM, 1994; PETRIN, 2014). In this view, which was considered to be the orthodox view at least until the beginning of the 20th Century, corporations are intangible legal entities that have personality, but a personality different from the personality of their members (MACHEN, 1911).

To a great extent, the Anglo-American world has maintained that corporations must be treated as persons. The extension of rights and obligations from a natural person to a corporation actually arose from the interpretation of the word “person”, for example, in the American Constitution (DEWEY, 1926). However, in countries like France, Germany and Italy there was great debate that yielded a set of heterogeneous theories, for example (MACHEN, 1911): *Fiction Theory*, *Concession Theory*, *Group Personality Theory* or *Realist*

Sociological Theory, *The Bracket Theory* or *Symbolist Theory*, *Purpose Theory* or *Theory of Zweckvermögen*, *Hohfeld's Theory*, and *Kelsen's Theory*. We present a brief description of the tenets of each theory according to MACHEN (1911), DEWEY (1926) and KOESSLER (1949).

The *Fiction Theory* – introduced by *Savigny* – discusses who would be the real owner of a property considering that property, according to the law, may belong to a corporation. As a corporation is not a natural person, the solution is to create a sort of *fictitious person*, which is the owner of the corporate property. The corporation was considered a creation of law having no existence apart from its individual members, and whose acts are attributed to the corporate entity (MACHEN, 1911; KOESSLER, 1949).

Contrary to that the *Concession Theory* – advanced by *Savigny*, *Dacey* and *Salmond* – asserts that corporations do have not legal personality. While Fiction Theory ultimately is based on the philosophical position suggesting that a corporation is a thing of the intellect, the Concession Theory focuses alone on the source of its legal power. In fact concession theory is indifferent to questions of reality, and states that a juridical person is merely a concession by a state or a legal system (DEWEY, 1926).

Real Entity Theory, also called *Group Theory* or *Realist Sociological Theory* – introduced by *Althusius* and *Gierke* – was focused on sociological facts. According to their supporters, collective groups (as corporations) have a real mind, a real will and a real ability to action. Thus, the existence of a corporation is not based on any fiction or in any merely mental entity. A corporation actually is a social organism, which is different from biological organisms in regards to the fact that it does not have the power to reproduce (MACHEN, 1911; DEWEY, 1926).

Associational Theory, also *Bracket Theory* or *Symbolist Theory* – propounded by *von Jhering* and others – suggests that the juristic corporate personality is only a symbol employed to facilitate the work of what Machen (1911) called “corporate bodies”.

On the one hand, this theory follows the Fiction Theory in maintaining that the existence of a corporation is a fiction, a metaphor; on the other hand, it states that the corporate personality is not created by the state because it actually does not exist. In other words, the corporation is solely an abbreviated form to represent people that are member of that kind of collective (MACHEN, 1911).

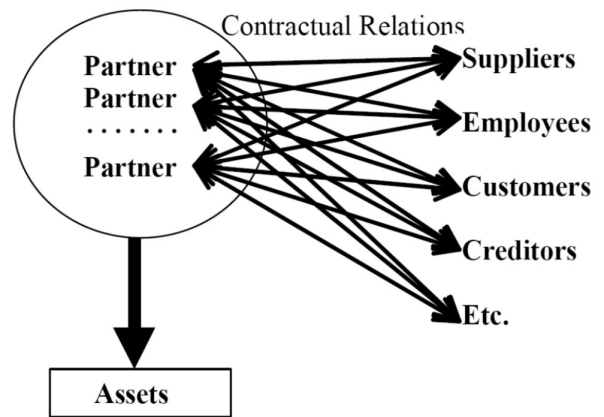
Purpose Theory, also *Theory of Zweckvermögen* – created by *Bekker* and *von Brinz* – is a variant of the Fiction Theory that explains ownership in charitable corporations (MACHEN, 1911). It considers a corporation as a fictitious entity, but focusing on the purpose of those who manage the property instead of focusing on the ownership of property by an object (KOESSLER, 1949).

THE NATURE OF THE CORPORATION IN ECONOMIC THEORIES

The field of Economics has a special interest in defining corporations, since these entities involve many aspects of the society and play a central role in economic analysis. In this section, we present a brief description of the tenets of individual economic theories about corporations.

Iwai (2001) states that within the economic context, in the simplest perspective, individuals (natural persons) are subjects of property rights and things are objects of property rights, that is, individuals own things and things are owned by individuals. However, there are cases in which a group of individuals act as partners and owns the assets jointly. These partnerships then enter in several contractual relations with other parties, like employees, suppliers, customers, etc., in order to perform activities and reach their goals. In this scenario, whenever there is either a new partner admitted or a partner withdrawing from the partnership, the original contractual relations no longer portray the actual situation. Thus, the number of contractual relation grows at same time that corporation grows, and transactions costs can be high (Fig.1). (IWAI, 2001)

Figure 1 – Contractual relations between the corporation and outside parts



Source: Iwai (2001)

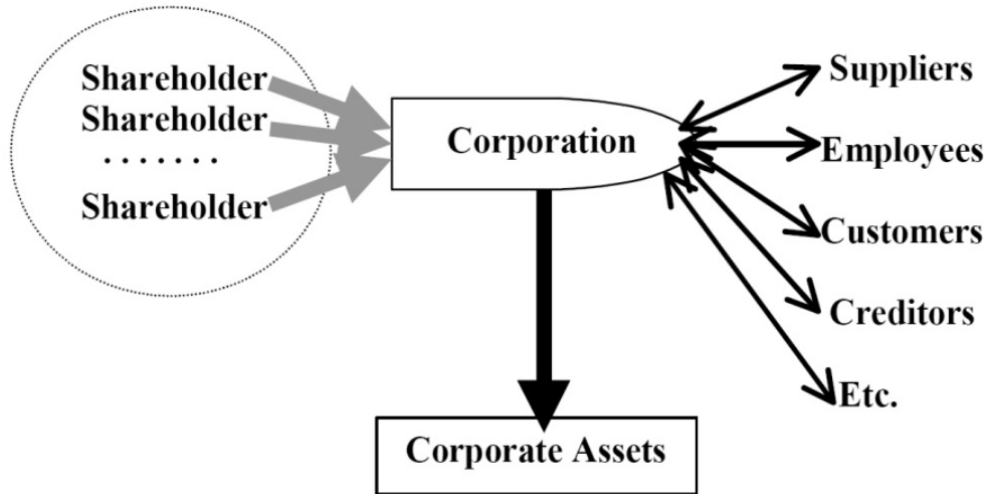
The corporation is a solution for the existence of these multiple contractual relations. When a group of individuals creates a corporation, it also creates an “additional individual” that has the same legal capacity to own real assets as the partners themselves have (Figure 2). Outside parties enter into a contract with this additional individual, independently of the actual partners or their number (IWAI, 2001).

According to Iwai (2001), in seeking to understand the nature of corporation in this context, one needs to understand the ownership relation. Here an additional potential quandary arises. Corporations seem play a dual role with respect to the ownership relation: they can act as individuals, that is, as legal subjects; they can act as thing, that is, as legal objects. The corporation (as an individual) owns its assets and it is owned (as thing) by shareholders. In the former case, it acts legally as a person; in the latter, it acts legally as a thing (IWAI, 2001).

BACKGROUND: ONTOLOGY OF SOCIAL ENTITIES

In his section be provide background information on pre-existing ontology resources. In the first sub-section 3.1 we introduce basics of ontologies in computer and information science, BFO as our top-level ontology and some complementary principles; and in the second sub-section 3.2 we discuss essentials of the social ontology.

Figure 2 – Corporation as a device to simplify outside relations



Source: Iwai (2001)

INTRODUCING OF ONTOLOGIES AND BASIC FORMAL ONTOLOGY (BFO)

Ontology is a term with several different meanings in different scientific fields, such as Philosophy, Computer Science, and Library & Information Science. In *Philosophy*, ontology is a branch of Metaphysics, which includes the study of existence, identity, change over time, dependency, qualities, and so forth (SMITH, 2003). The term “social ontology” refers to the ontology of human social devices, for example, money, property, governments, nations, marriages and so forth (SEARLE, 2010).

In *computer and information science*, two uses for the term ontology are most important: i) the use of ontological principles to understand and represent reality as support to modeling activities (WAND; WEBER, 1990); ii) the representation of a knowledge domain through a formal language to be processed by automatic reasoners (STAAB; STUDER, 2004).

Top-level ontologies are artefacts of the second kind. They are category systems for knowledge representation that have been developed to meet specific goals concerning to information modeling, automatic reasoning and information retrieval (HOEHNDORF, 2010). Examples of top-level ontologies are the *Descriptive Ontology for Linguistic*

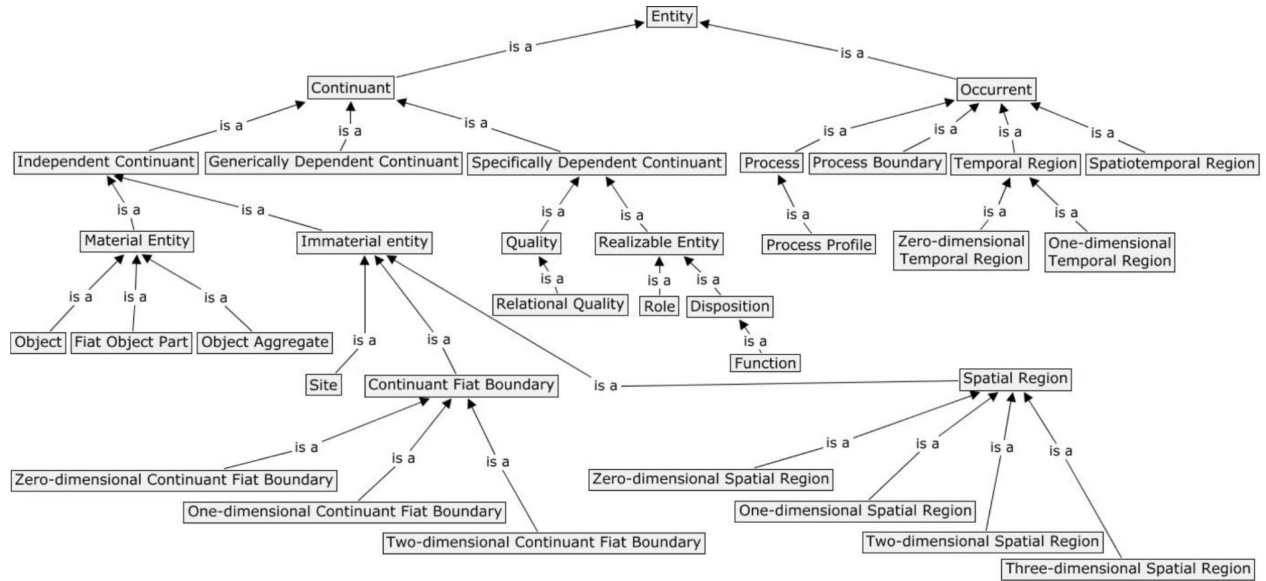
and Cognitive Engineering (DOLCE) ((GANGEMI et al, 2002); and Basic Formal Ontology (BFO) (GRENON; SMITH; GOLDBERG, 2002).

In this paper we focus on using BFO to model social entities relevant to corporations. BFO is following the principles of *ontological realism* (SMITH; CEUSTERS; 2010) and has shown a lot of promise and uptake lately. As a top-level ontology, BFO intends both to represent the most generic categories and to provide means of categorizing entities in a knowledge representation (SPEAR, 2006). BFO has a large acceptance in domains like medicine, biology, bioinformatics and related fields, law, geography, to mention a few.

BFO consists of a taxonomy that provides a categorization of the entities existing in the world and it thus, provides an account of reality (Fig.3). Consequently, the root node of BFO’s taxonomy is “entity”.

The second level introduces the major distinction underlying BFO: the distinction of entities into *continuants* and *occurrents*. Continuants are those entities which exist self-identical through time and which undergo changes. Occurrents only exist in their successive temporal parts (GRENON & SMITH, 2004). The full framework of BFO is describe in detail in Arp, Smith; Spear (2015).

Figure 3 – BFO top-level ontology, its levels and categories



Source: Smith et al. (2015)

In addition to top-level ontologies, a variety of formal frameworks is available for creating and testing ontological representations (SMITH, 2003).

Mereology use two sets of principles to explain the relations between parts and wholes: i) *principles of decomposition*, which take one from a whole to its parts; and ii) *principles of composition*, which take one from the parts to the whole (SIMONS, 1987). A relevant property of mereology is *extensionality*, which ensures that: i) entities are completely defined by their parts; and ii) no composite objects with the same proper parts can be distinguished (VARZI,1996). The principle of extensionality is important when explaining the descriptive dimension of a corporation.

The Theory of Granular partitions deals with granular partitions, which are cognitive devices that people employ to label or sort activities. Examples are lists, hierarchies, classifications and so forth (BITTNER; SMITH, 2008). Within the context of granular partitions, an object is anything that exists and that can be recognized by some unit of partition. Objects can be either *bona fide objects* or *fiat objects*: while the former exists independently of human demarcating activities, the latter exists only because

of the very same activities. Indeed, partition units can recognize fiat objects from limits based on human cognition, and fiat objects are created through the projection of partitions in a portion of reality (SMITH, 2001).

ESSENTIALS OF SOCIAL ONTOLOGY: SOCIAL ACTS, DOCUMENT ACTS AND THEIR IMPLEMENTATION IN ONTOLOGIES

In order to explain the normative dimension of analysis, we need to introduce key components of social ontology, namely: *social acts*; *speech acts*, *document-acts* and Document Act Ontology (SEARLE, 1969; 1976; SEARLE, 1983; SEARLE, 1984; SEARLE, 2002; SMITH 2012; BROCHHAUSEN, ALMEIDA, SLAUGHTER, 2013). Historically, the ontological analysis of legal acts and legal entities has paved the way to social ontologies.

An ontological theory explaining social acts, spoken or written, was first advanced by *Reinach* (SMITH, 2012) in his work focused on the phenomenological analysis of civil law. In his analysis, Reinach starts out from *spontaneous acts*: the set of experiences a person may have, in which, in addition, the very

same person has an active participation. Indeed, some experiences require the existence of a subject of reference and Reinach called them *non-self-directable*. One specific subtype of those acts are acts that in addition to referring to another subject, have to be perceived by the very same subject. Acts that need to be perceived are then called *social acts*, actually, acts with intentional roots (MULLIGAN, 1987). Both *speech acts* and *document acts* are subtypes of social acts.

The *Theory of Speech Acts* was originally proposed by Austin. Speech Acts Theory proposes that the elementary units to understand human communication are speech acts (MULLIGAN, 1987). A speech act is the basic unit of meaning, constituted by three connected dimensions: *locutionary acts*, *illocutionary acts* and *perlocutionary acts*. We do not need to plan going into more detail here. Further examples can be found in Morris (2007).

Searle consolidated Austin's approach to a general Theory of Speech Acts by providing a theoretical framework that gathered together the three dimensions involved in speech acts. Searle distinguishes "brute force facts", basically, the facts of the natural science; from the so-called "institutional facts", which are created through language. Thus, there is a difference between observer-independent features, for example mass and gravitational attraction; and observer-dependent features, for example money and property, which are examples of institutions in Searle's sense. Searle describes and analysis some of those entities – powers, functions, acts, events, states, properties, and relations – that do not belong to the realm of brute physical reality, but rather to the realm of institutions. Through these entities, people are able to bring into existence entities of the social reality (SMITH, 2003).

One issue regarding the speech acts is its evanescence, a result of its inherent orality. This makes the possible range of effect of a speech act temporally and regionally constrained. To overcome this limitation Smith extended speech act theory to the *Theory of Document Acts* (SMITH, 2012). Indeed, a speech act exists only in the moment of

its performance; documents, on the other hand, are continuant entities able to persist in time while absorbing modifications through its history.

A theory of document acts explains how people and organizations make use of documents to bring about social entities, such as social roles. Documents here do not serve only to document information, through their involvement in a document act they also have social and institutional (ethical, legal) powers, contributing to bring about a variety effects, which have made possible new types of social institutions. Examples of this kind of documents are contracts, statutes of incorporation, deeds, patents, statements of accounts, marriage licenses, stock certificates, diplomas, registrations of baptism, insurance certificates, and so forth. In addition, the theory document of acts involves what humans do with documents, for example: signing, stamping, depositing in registries, filling in, registering, conveying, validating, attaching, to mention a few. Documents, in this context, play an essential role in social interactions, and in human daily activities (SMITH, 2012; SMITH, 2014).

Document Acts Theory is operationalized through the aforementioned Document Act Ontology (d-acts). This ontology reuses pre-existing ontologies developed on the basis of the *Open Biological and Biomedical Ontologies (OBO) Foundry principles* (SMITH et al., 2007). The basic representations of information artifacts is reused from the *Information Artifact Ontology (IAO)*³. Besides IAO, we imported selected classes and object properties from *Ontology of Biomedical Investigations*⁴ (BANDROWSKI (2016), *National Center for Biotechnology Information (NCBI) Taxonomy*⁵ (NCBI, 2016) and the *Ontology of Medically Related Social Entities (OMRSE)*⁶ (HICKS; HANNA; WELSH; HOGAN, 2016).

³ Retrieved July 4, 2012 <https://bioportal.bioontology.org/ontologies/IAO>

⁴ Retrieved July 4, 2012, from <http://www.berkeleybop.org/ontologies/oboall/obi/obi.owl>

⁵ Retrieved July 4, 2012, from http://www.berkeleybop.org/ontologies/oboall/ncbi_taxonomy/ncbi_taxonomy.owl

⁶ Retrieved July 4, 2012, from <http://www.berkeleybop.org/ontologies/oboall/omrse/omrse.owl>

ONTOLOGICAL ANALYSIS OF CORPORATIONS

After we presented the legal and historical context (section called “Background: the nature of corporation”) and some background regarding ontologies and social ontologies (section called “Background: ontology of social entities”), we are now ready to introduce the ontological analysis of corporations. The next sub-section briefly explains the descriptive dimension, and the last subsection of this section describes the prescriptive dimension.

THE DESCRIPTIVE DIMENSION

In this section, we briefly discuss the descriptive dimension of the corporation, since our focus rests on the prescriptive dimension. We describe how the corporation can be divided in units and subunits in the scope of BFO. In the remaining of this section, we present a brief description of the tenets of each theory according to Smith (2016) and Vizenor (2006).

Corporations have some relevant features: i) they do not depend on other entities except constitutive entities; ii) they remain numerically one and the same over time; iii) they are composed by their members. Thus, corporations have some features in common with BFO’s independent continuants, which are characterized by the presence of all their parts at any time, and by the fact that their existence does not depend on any other discrete entity. In order to understand a corporation and its divisions, we need to use the partitive principles (presented before in the background section). So, what kind of whole a corporation would be? *Summative wholes* are exhaustively defined by their constituent parts. According to the principle of mereological extensionality, objects with the same parts are equal. However, this notion does not conform to the intuitive notion that corporations preserve identity even undergoing changes in membership. *Integral wholes*, on the other hand, have strong structural connections among their parts, differently of corporations, which are not maximally connected like, for example, organisms. Finally, *aggregates* are not defined in terms of their own elements.

Otherwise, they have detached parts that integral wholes do not have. As corporations have members linked together (persons, units, for example), one can acknowledge that corporations are kinds of aggregates: they are material entities consisting exactly of a plurality of objects, and these objects are member-parts of them all times at which they exist (SMITH, 2016).

In order to divide a corporation in units and subunits, we make use of partitive theories (presented before in the background section). Mereology deals with material entities and does not seem to be the best framework to explain corporations and its units. It is hard to believe that the relation between a corporation and its members is a part-whole relation, insofar as in mereology, the part-whole relation is transitive. Accordingly, one might say inconsistent statements, for example: if John is part of a corporation, any part of John as for example his eyes or his mouth, are also part of the corporation. In addition, mereology cannot account the fact that a corporation preserves its identity over time even when it loses or gain members.

The Granular Partitions theory, on the contrary, employ cognitive devices to show how people divide the world. It relies on the distinction between bona fide objects and fiat objects, as mentioned before: bona fide objects exist independently of human subdividing activity; and fiat objects exist only because of the very same subdividing activity. Using the Granular Partitions theory – namely, Theory A – one can define units, sub-units (or cells and sub-cells), as well as the relation between unit/subunit (or cell/sub-cell). Likewise, using the Granular Partitions theory – namely, Theory B – one can define the projection relation that holds from a cell to the reality; and the location relation that holds from an object to a cell (SMITH, 2016; VIZENOR, 2006).

Finally, we assume here that, in the scope of BFO, a corporation is an independent continuant, a kind of aggregate. However, a colony of ants or herd of antelopes can be classified as an aggregate. What are the differences between a corporation as an aggregate and the other examples? What distinguishes a corporation from other kinds of aggregates? To answer this question, we need to approach the normative dimension of corporations.

THE PRESCRIPTIVE DIMENSION

If corporate units (descriptive dimension) are *fiat* objects, one can ask: how does a *fiat* object come to existence? A *fiat* object comes to existence through the human cognition, which can create a mental partition, for example, delimiting a corporate unit. Considering that corporations are long-lasting entities, how a fiat object is sustained in existence? (SMITH, 2016). Here, we are approaching the realm of the social ontology: to explain the verbal form, we use the *Theory of Speech Acts* (AUSTIN, 1975; SEARLE, 2010); to explain the written form we use the *Theory of Document Acts* (SMITH, 2012).

The Document Acts theory, which is crucial to represent traceable duties, obligations and responsibilities, for example within a corporation, can be related to BFO through the *d-acts* (BROCHHAUSEN; ALMEIDA; SLAUGHTER, 2013). *d-acts* incorporates the kinds of acts present in the document act theory, namely: *social acts*, *deontic declarations*, *document acts*, and *standing declarations*. We start our explanation from the notion of *social act* and continue its subtypes, in *d-acts*⁷.

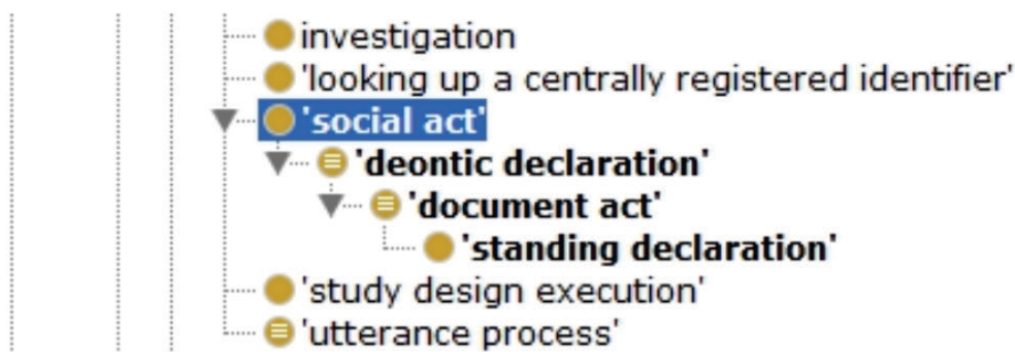
A *social act* is a process that is carried out by a conscious being or an aggregate of conscious beings. It is spontaneous, directed towards another conscious being or another aggregate of conscious beings, and that needs to be perceived. The term “spontaneous” is used in the following meaning of being self-generated and it does not stand in contrast to being planned.

Examples of social acts are: Colonel Klink giving Sergeant Schultz an order; Jake promising Jill to take her to the junior prom.

A *social act* is a process that is carried out by a conscious being or an aggregate of conscious beings. It is spontaneous, directed towards another conscious being or another aggregate of conscious beings, and that needs to be perceived. The term “spontaneous” is used in the following meaning of being self-generated and it does not stand in contrast to being planned. Examples of social acts are: Colonel Klink giving Sergeant Schultz an order; Jake promising Jill to take her to the junior prom.

A *deontic declaration* is a social act that brings about, transfers or revokes a socio-legal entity. Declarations do not depend on words spoken or written, but sometimes are they merely actions, for instance the signing of a document. Deontic declarations express what ought to be, for example, a promise is a declaration of how things ought to be. They can either create or remove social entities in reality: a promise creates both an obligation of one person and a claim of another person; but a declaration can also revoke a social entity, such as an order or an obligation. Examples of deontic declarations are: my consenting verbally to buy a used TV set for \$ 500; John Robbie’s taking of Mrs. Steven’s jewels; Jane Doe’s revoking of informed consent over the phone; John Doe signing an employment contract.

Figure 4 – Main classes of D-acts ontology



Source: D-acts ontology

⁷ <http://purl.obolibrary.org/obo/iao/d-acts.owl>

A *document act* is a social act creating, revoking or transferring a socio-legal entity by validating (signing, stamping, publishing) a document. Examples of document acts are: my filling in an immigration form; a judge’s signing and stamping a court order. Finally, a *standing declaration* is a document act that postulates a socio-legal fact for a specified group of people or organizations. An example is in the case of the promise from the Chief Cashier of the Bank of England ‘to pay the bearer on demand the sum of £5’ that is printed on each £5 note.

Figure 5 presents an example of using d-acts Ontology. In such example, we represent a promotion process in a corporation: Maria has received a promotion, but her hospital privileges and her obligation to communicate with the department supervisor remains unchanged.

DISCUSSION

The prescriptive (normative) dimension of analysis presented before still present the necessity of further research and discussion. At this moment, we raise some speculations that admittedly need to be properly verified in future works. One speculation regards the relevance of corporative documents and document acts for charactering corporations and kinds of corporations.

We believe that one can characterize the kind of corporation through both the documents that serve as inputs to document acts and the kinds of the very same document acts. We survey some theoretical evidences, coming from philosophers and researchers, for our premise that corporations are distinguished by documents they produce and use (ALMEIDA, 2016).

Figure 5 – Example of using D-acts

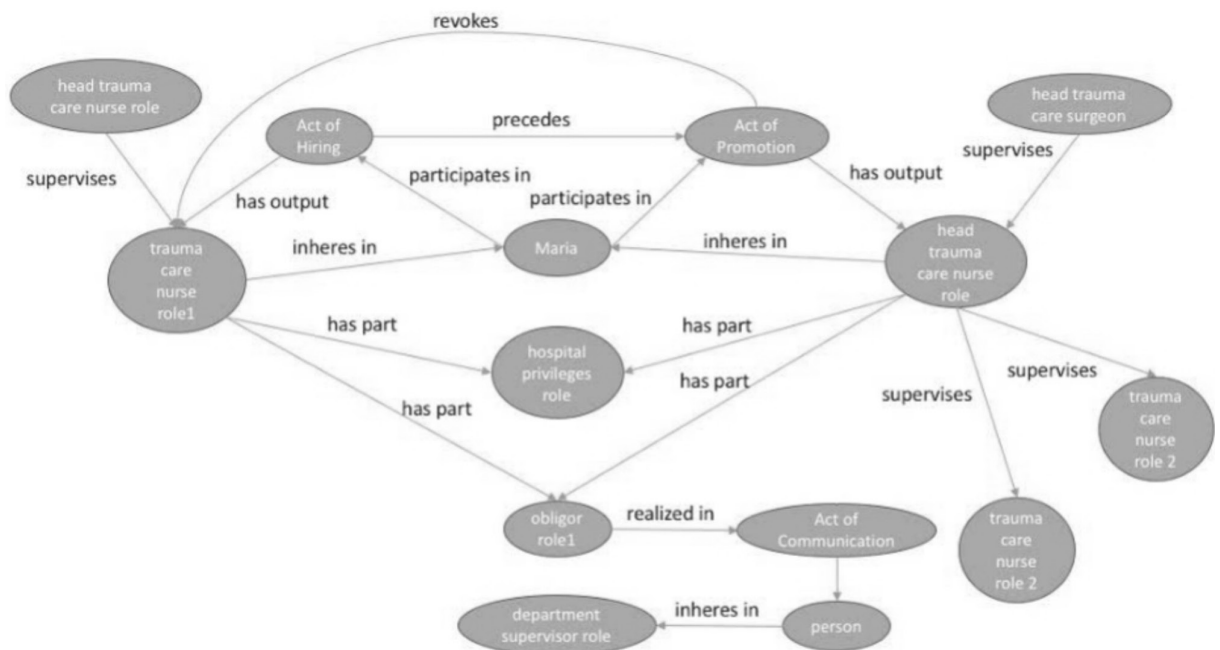


Image credit: J. Neil Otte, University at Buffalo (SUNY)

Practices and resources regarded documents reveal how one does organization of information within a corporation. The way one structures documents imposes certain kinds of administration to employees and eventually create indexes of how the corporation is managed. Indeed, documents are both the means and the results of a continuous process of notation, summarization, and information dissemination that aims to construct a depiction of what happens in the corporate environment (LEDEMA, 2013).

The ubiquity of documents in our society is also described in a theory called *documentality*, according to which “there is nothing social outside the text (FERRARIS, 2012)”. According to this theory, since nothing social exists outside the text, society is based on the registrations in documents, and this very act of registering is the condition for creating social objects. Thus, documents constitute the fundamental ingredient of the social world (FERRARIS, 2012).

Indeed, documents are important to explain society, and social relations, for example, those ones produced and manipulated within a corporation. As society became more and more complex, “the mnemonic powers of individuals have been extended prosthetically through documents in ways which have given rise to a variety of novel artifacts of social reality” (SMITH, 2014, p. 5).

So, we also say that one could identify the kind of corporation through the kinds of document acts performed. Our speculation is that, in identifying the kinds of D-acts that a document contains or generates may provide clues to formally distinguish a corporation from other ones.

FINAL REMARKS

In this paper, we presented the preliminary foundations of an ontological representation for corporations. One might say that we briefly approach the ontological status of corporations both in describing historical theories regarding the nature of corporations to identify their ontological commitments, and in dividing our task in two dimensions, namely, descriptive and prescriptive. However, to fully define the ontological status of corporation is a complex task beyond the purposes of the paper.

We choose to focus the prescriptive (normative) dimension of the corporation and carried out our research using well-founded theories – formal ontology, speech acts, document acts – in the hope to offer a solid framework to analyze corporations.

As our final remarks, we would like to emphasize the connection among the different representations existing throughout our ontological analysis for purposes of formalization and application in the modern information systems. One can realize that both the reality and different levels of representations of it – partitions, speech acts, d-acts – can be mapped to BFO, a transcategorial formal top-level ontology. The relevance of such analysis rests on the need of a better understanding of corporations, as well as the advantages of a formal framework to be applied in information systems working in the context Semantic Web.

As future works, we plan to overcome some limitations of the present stage of our on-going research. We need to consider the myriad of kinds of existent corporations: third sector, churches, clubs, commercial associations, public companies and so forth. Here, for the sake of simplicity, we privilege to choose the standard commercial corporations with two or more partners. Also, something that can be approached in future papers is the classification or organization of the deontic powers involved in claims and obligations that d-acts represents. Finally, we need to evaluate how modern theories were influenced by old theories about the nature of the corporation.

REFERENCES

- ALCHIAN, A.; DEMSETZ, H. Production, information costs, and economic organization. *American Economic Review*, n. 62, p.777-795, 1972.
- ALMEIDA, M. B. *An ontological study of healthcare corporations and their social entities*. 2016. Disponível em: <<http://icbo.cgrb.oregonstate.edu/node/347>>. Acesso em: 20 nov. 2016.
- ARP, R.; SMITH, B.; SPEAR, A. D. *Building ontologies with basic formal ontology*. Cambridge: MIT Press, 2015. 208 p.
- AUSTIN J. L. *How to do things with words*. Cambridge: Harvard University Press, 1975.
- BANDROWSKI, A. et al. *The ontology for biomedical investigations*. 2016. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0154556>>. Acesso em: 10 ago. 2017.
- BLACKSTONE, W. *Commentaries on the laws of England: rights of persons*. Oxford: Oxford University Press, 2016.
- BITTNER T.; SMITH B. The theory of granular partitions. In: MUNN, V.; SMITH, S. (Ed.). *Introduction to applied ontology*. Berlin: Onto-Verlag; 2008. P. 125-158.
- BLUMBERG P. I. et al. *Blumberg on corporate groups*. 2004. _____, *The multinational challenge to corporation law: the search for a new corporate personality*. Oxford: Oxford University Press, 1993.
- BROCHHAUSEN M.; ALMEIDA M. A.; SLAUGHTER L. Towards a formal representation of document acts and the resulting legal entities. In: JOHANSSONIAN Investigations. Frankfurt: Ontos, 2013. P. 120-39.
- CHANDLER, A. Organizational capabilities and the economic history of the industrial enterprise. *Journal of Economic Perspectives*, v. 6, p.79-100, 1992.
- COASE, R. H. The nature of the firm. *Economica*, n. 4, p. 386-405, 1937.
- DEWEY, D. *The historic background of corporate legal personality*. 1926. Disponível em: <http://rci.rutgers.edu/~tripmcc/phil/dewey-historic_background_of_corporate_legal_personality.pdf>. Acesso em: 02 ago. 2017.
- EASTERBROOK, H.; FISCHER, D. *The economic structure of corporate law*. Cambridge: Harvard University Press, 1991.
- FERRARIS, M. *Documentality: why it is necessary to leave traces*. Fordham: Fordham University Press, 2012.
- FONSECA, F. The double role of ontologies in information science research. *Journal of the American Society for Information Science and Technology*, v. 58, n. 6. p.786-793, 2007.
- GANGEMI, A. et al. *Sweetening ontologies with DOLCE*. 2002. Disponível em: <<http://www.loa.istc.cnr.it/Papers/DOLCE-EKAW.pdf>>. Acesso em: 12 ago. 2014.
- GRENON, P.; SMITH, B. SNAP and SPAN: towards dynamic spatial ontology. *Spatial Cognition and Computation*, v. 4, n. 1, p. 69-103, 2004.
- _____; GOLDBERG, L. J. Biodynamic ontology: applying BFO in the biomedical domain. In: PISANELLI, D. M. (Ed.). *Ontologies in Medicine*. Amsterdam: IOS Press, 2004.
- GUIZZARDI, G. *Ontological foundations for structural conceptual models*. Twente: University of Twente, 2005.
- HICKS, A. et al. *The ontology of medically related social entities: recent development*. 2016. Disponível em: <https://www.researchgate.net/publication/305271770_The_ontology_of_medically_related_social_entities_Recent_developments>. Acesso em: 10 ago. 2017.
- HOEHNDORF, R. *What is an upper level ontology?*. 2010. Disponível em: <<http://ontogenesis.knowledgeblog.org/740>>. Acesso em: 10 jan. 2017.
- IWAI, K. *What is corporation?: the corporate personality controversy and comparative corporate governance*. 2001. Disponível em: <https://www.researchgate.net/publication/24135014_What_is_Corporation_---The_Corporate_Personality_Controversy_and_Comparative_Corporate_Governance>. Acesso em: 12 mar. 2015.
- JENSEN, M.; MECKLING, W. The theory of firm: managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, v. 3, p. 305-360, 1976.
- KOESSLER, M. *The person in imagination or persona ficta of the corporation*, 1949. Disponível em: <<http://digitalcommons.law.lsu.edu/cgi/viewcontent.cgi?article=1615&context=lalrev>>. Acesso em: 15 maio 2017.
- LEDEMA, R. *Discourers of post-bureaucratic organization*. Amsterdam: John Benjamins Publishing Company. 2003.
- MACHEN, A. W. Corporate personality. *Harvard Law Review*, v. 24, n. 4, p. 253-67, 1911.
- MALCOLM, J. *Do corporations have rights?*. Disponível em: <<https://www.malcolm.id.au/honours/index.html#contents>>. Acesso em: 15 jan. 2017.
- MORRIS, M. *An introduction to the philosophy of language*. Cambridge: Cambridge University Press, 2007.
- MULLIGAN, S. *Speech acts and sachverhalt: reinach and the foundations of realist phenomenology*. Amsterdam: Springer Netherlands, 1987.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION - NCBI. Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, n. 44, p. 7-19, 2016. Disponível em: <<http://doi.org/10.1093/nar/gkv1290>>. Acesso em: 14 nov. 2017.

NELSON, R.; WINTER, S. G. *An evolutionary theory of economic change*. Cambridge: Harvard University Press, 1977.

PELIKAN, P. Evolution, economic competence and the market for corporate control. *Journal of Economic Behavior and Organization*, v.12, p. 279-303, 1989.

PENROSE, E. *The theory of the growth of the firm*. Cambridge: Oxford Scholarship Online, 1959.

PETRIN, M. *Reconceptualizing the theory of the firm: from nature to function*. 2017. Disponível em: <[http://www.pennstatelawreview.org/118/1/1-Petrin%20\(PS%20final\).pdf](http://www.pennstatelawreview.org/118/1/1-Petrin%20(PS%20final).pdf)>. Acesso em: 14 nov. 2017.

PRAHALAD, C. K.; HAMEL, G. The core competence of the corporation. *Harvard Business Review*, v. 68, p.79-91, 1990.

REINACH, A. *The apriori foundations of the civil law*. Frankfurt: Ontos-Verlag, 2012.

SEARLE J. R. *Speech acts: An essay in the philosophy of language*. Cambridge University Press, 1969.

_____. A classification of illocutionary acts. *Language in Society*, v. 5, n. 1, p.1-23, 1976.

_____. *Intentionality: an essay in the philosophy of mind*. Cambridge: Cambridge University Press, 1983.

_____. *Minds, brains and science*. Harvard: Harvard University Press, 1984.

_____. Speech acts, mind, and social reality. In: GREWENDORF, G.; MEGGLE, G. (Ed.). *Speech acts, mind, and social reality: discussions with John R. Searle*. [S.l.: s.n.], 2002. P. 3-16.

_____. *Making the social world: the structure of human civilization*. Oxford: Oxford University Press, 2010.

SIMONS, P. *Parts: a study in ontology*. New York: Clarendon Press, 1987.

SMITH, B. Fiat objects. *Topoi*, v. 20, n. 2, p.131-148, 2001.

_____; CEUSTERS W. Ontological realism: a methodology for coordinated evolution of scientific ontologies. *Applied Ontology*, v. 5, p. 139-188, 2010.

_____. *Ontology and information systems*. 2003. Disponível em: <http://ontology.buffalo.edu/smith/articles/ontology_PIC.pdf>. Acesso em: 20 jan. 2012.

_____. et al. *BFO 2.0 specification and user's guide*. 2015. Disponível em: <http://ontology.buffalo.edu/bfo/Reference/BFO_March12.docx>. Acesso em: 02 jan. 2016.

_____. *John Searle: from speech acts to social reality*. Cambridge: Cambridge Press, 2003.

_____. How to do things with documents. *Rivista di Estetica*, n. 50, p. 179-198, 2012.

_____. *Documents-acts*. 2014. Disponível em: <<http://ontology.buffalo.edu/smith/articles/document-acts.pdf>>. Acesso em: 10 jan. 2017.

_____. *The ontology of the orngigram*. Disponível em: <<https://www.youtube.com/watch?v=bp3B2GkgaB8&feature=em-subdigest>>. Acesso em: 13 jan. 2017.

SPARKES, A. W. *Talking Philosophy: a wordbook*. London: Routledge, 1991.

SPEAR, A. D. *Ontology for the twenty first century: an introduction with recommendations* [Internet]. 2006. Disponível em: <<http://ifomis.uni-saarland.de/bfo/documents/manual.pdf>>. Acesso em: 14 jan. 2017.

STAAB S.; STUDER R. *Handbook on ontologies*. Berlin: Springer, 2004.

TEECE, D. J. Towards an economic theory of the multi-product firm. *Journal of Economic Behavior and Organization*, v. 3, p. 39-63, 1982.

VARZI, A. Parts, wholes, and part-whole relations: the prospects of mereotopology. *Data and Knowledge Engineering*, v. 20, p. 259-286, 1996.

VIZENOR L. *Corporate being: a study in realist ontology*. [S.l.]: University at Buffalo, 2006.

WAND, Y.; WEBER, R. Mario Bunge's ontology as a formal foundation for information systems concepts. In: WEINGARTNER, P.; DORN, J.W.G. (Ed.). *Studies on Mario Bunge's treatise*. Amsterdam: Rodopi, 1990.

WILLIAMSON, O. E. *The economic institutions of capitalism*. New York: Free Press, 1985.

WILLISTON, S. History of the law of business corporations before 1800. II. (Concluded). *Harvard Law Review*, v. 2, n. 4, p.149-166, 1888. DOI: 10.2307/1321930.

www.ibict.br



Ministério da
Ciência, Tecnologia
e Inovação

