

Webert Júnio Araújo

Universidade Federal de Minas Gerais
webertaraujo@ufmg.br

Gercina Angela de Lima

Universidade Federal de Minas Gerais
glima@eci.ufmg.br

EM BUSCA DE UMA METODOLOGIA PARA ENRIQUECIMENTO DE ONTOLOGIAS DE DOMÍNIO

RESUMO

Este resumo estruturado é parte de uma pesquisa de doutorado em andamento e foi apresentado no II Fórum de Pesquisas Discendentes da ECI/UFMG. O objetivo norteador é desenvolver uma metodologia com procedimentos detalhados para enriquecimento de todos os componentes das ontologias de domínio. A pesquisa caracteriza-se como aplicada, quanto à natureza, exploratória, quanto aos objetivos, e experimental no que concerne aos procedimentos metodológicos. As etapas propostas na metodologia desta investigação são sete: (1) pesquisa bibliográfica; (2) mapeamento da literatura sobre enriquecimento de ontologias; (3) estudo das normas e método oriundos da Biblioteconomia e Ciência da Informação para atualização e manutenção de Sistemas de Organização do Conhecimento; (4) estudo das metodologias de desenvolvimento de ontologias; (5) análise qualitativa e comparativa do conteúdo dos documentos levantados nas etapas 2, 3 e 4; (6) desenvolvimento da metodologia para enriquecimento de ontologias de domínio; (7) avaliação da metodologia desenvolvida. A pesquisa ainda possui resultados para serem apresentados e discutidos.

Palavras-Chave: Enriquecimento de ontologias, atualização de ontologias, ontologias de domínio.

ABSTRACT

This structured abstract is part of an ongoing doctoral research and was presented in the II Student Research Forum at ECI/UFMG. The guiding goal is to develop a methodology with detailed procedures for enriching all components of domain ontologies. The research is characterized as applied, in nature, exploratory in objectives and experimental as regards to methodological procedures. The proposed steps in the methodology of this investigation are seven: (1) bibliographic research; (2) mapping of the literature on ontology enrichment; (3) study of the norms and methods derived from Library and Information Science to update and maintain Knowledge Organization Systems; (4) study of ontology development methodologies; (5) qualitative and comparative analysis of the content of the documents raised in steps 2, 3 and 4; (6) development of the methodology for enriching domain ontologies; (7) evaluation of the developed methodology. There are no results yet to be presented and discussed.

Keywords: Ontology enrichment, ontology maintenance, domain ontology.

Universidade Federal de Minas Gerais

Correspondência/Contato
Av. Antônio Carlos, 6627
Pampulha: 31270-901
BELO HORIZONTE - MG

Escola de Ciência da Informação da UFMG

1. INTRODUÇÃO

As ontologias são essenciais para o contexto da Web Semântica devido às suas características de representação do conhecimento de modo formal, que possibilitam a interpretação do que está representado por computadores. Junto a isso, as ontologias permitem a integração semântica entre fontes heterogêneas. Ora, a Web Semântica visa justamente que os dados façam sentidos para as máquinas. Em suma, as ontologias auxiliam para que o sonho da Web Semântica se torne cada vez mais real.

O desenvolvimento de ontologias não é tarefa simples e passa por várias etapas. A área responsável por todos os processos associados ao desenvolvimento e manutenção de ontologias é a Engenharia Ontológica. Todos os processos da Engenharia Ontológica podem ser segmentados em três grupos: pré-desenvolvimento, desenvolvimento e pós-desenvolvimento. No pré-desenvolvimento são previstos os processos e atividades referentes ao planejamento e preparação para o desenvolvimento das ontologias, nessa fase são definidos os objetivos da ontologia, seus usuários, questões que ela irá responder, entre outros. No desenvolvimento estão os processos e atividades que lidam com a implementação da ontologia, assim são criados os conceitos, relacionamentos, instâncias, axiomas, etc. E por fim, no pós desenvolvimento estão os processos e atividades responsáveis pela manutenção, atualização e avaliação da ontologia, ou seja, depois de desenvolvida, a ontologia precisa ser verificada para saber se está de acordo com o que foi planejado no pré-desenvolvimento.

Frente aos três grupos da Engenharia Ontológica, nessa pesquisa o foco será no pós-desenvolvimento, ou seja, na fase de manutenção e atualização de ontologia. Essa fase é composta de vários processos, a saber: reengenharia, reestruturação, evolução, versionamento, customização, população, modificação, enriquecimento, etc. Dessa forma, de modo mais específico, a investigação dessa pesquisa se empenhará no processo de Enriquecimento de Ontologia.

O processo de Enriquecimento trata da expansão e especialização da estrutura da ontologia, através da inclusão de novos conceitos e seus símbolos linguísticos alternativos (sinônimos, por exemplo), relações (taxonômicas e não-taxonômicas) e suas propriedades restritivas e axiomas; que devem ser incluídos nos locais corretos na ontologia. Visa cooperar justamente para que as ontologias representem o domínio de modo mais completo e fidedigno.

Ontologias são construídas com base no conhecimento que determinado grupo de especialistas tem de um domínio, assim como as fontes de informação que expli-

citam esse conhecimento. Porém o conhecimento científico é algo que muda com as novas descobertas ou até mesmo com refutações de teorias consolidadas. Um dos principais objetivos do conhecimento científico é entender os fenômenos do mundo real. Portanto, através do método científico e do rigor científico, empenha-se em descrever a realidade ou como funcionam os fenômenos no mundo. Diante disso, os desenvolvedores de ontologias precisam garantir que esses instrumentos passem por um processo de manutenção e atualização sempre que houver necessidade, ou seja, sempre que as ontologias não estiverem representando o conhecimento que reflita a realidade do domínio modelado.

Para que toda a capacidade das ontologias seja aproveitada, tais como possibilidade de funcionar como um acordo semântico e terminológico em um campo, inferência de novos conhecimentos, aplicação na recuperação da informação, etc, é preciso que a ontologia seja a mais completa possível, com todos os conceitos associados ao domínio de desenvolvimento, definições completas para todos os conceitos e relacionamentos, estabelecimento das propriedades concernentes aos conceitos, atribuição de axiomas, que ajudam a restringir as interpretações de determinado conceito e permitem a inferência automática de conhecimento.

Algumas das principais metodologias para desenvolvimento de ontologias (Método 101 de NOY e MCGUINESS, 2001; Sensus de SWARTOUT, 1996) não prevêem o processo de manutenção de ontologias. Como se depois de construída a ontologia não precisasse passar por mais nenhuma etapa. Essas limitações de algumas metodologias impactam diretamente na reutilização desses instrumentos. O processo de Manutenção de Ontologias está previsto nas metodologias (Methontology de FERNÁNDEZ et al, 1997; On-to-Knowledge de SURE, STAAB e STUDER, 2003; NeON de SUÁREZ-FIGUEROA, 2010), porém, na maioria delas não está bem descrito como essa manutenção deve ocorrer. Isso se deve principalmente ao grau de maturidade da área de Engenharia de Ontologia que ainda é baixo se comparado com a Engenharia de Software. (SUÁREZ-FIGUEROA, 2010). A metodologia NeOn é uma das mais avançadas no que refere ao desenvolvimento de ontologia, pois prevê nove cenários possíveis para o desenvolvimento de ontologias.

A temática manutenção/atualização de ontologias foi ignorada por um tempo pelos pesquisadores, que estavam mais preocupados com a construção de ontologias. A manutenção de ontologias só ganhou notoriedade quando se percebeu que as ontologias não são sistemas estáticos e que precisavam ser atualizadas, posto que se baseiam em conhecimento. Conforme afirmam Smith e Klagges (2008, p. 24), “[...] mesmo

uma ontologia baseada em princípios filosóficos sempre será um edifício parcial e imperfeito, sujeito a correção e aprimoramento, de modo a atender a novas necessidades científicas.”

As questões que norteiam o presente estudo são: (1) Como desenvolver uma metodologia para enriquecimento de todos os componentes das ontologias de domínio? (2) O desenvolvimento de uma metodologia para enriquecimento de ontologias de domínio pode contribuir para solucionar os problemas relacionados às ontologias incompletas?

2. ENRIQUECIMENTO DE ONTOLOGIAS

Enriquecimento de Ontologia (EO) é um assunto emergente na área de Engenharia Ontológica, portanto, ainda existe pouco embasamento teórico sobre esse processo. Há também uma certa confusão terminológica, uma vez que o termo EO ainda não está totalmente sedimentado.

Na literatura há um uso indiscriminado de termos para definir a atividade de enriquecimento de ontologia ou atividades que têm relação estreita. Alguns dos termos são: expansão de ontologia (ontology expansion¹), ampliação de ontologia (ontology enlarge), completude da ontologia (ontology completeness), melhoria de ontologia (ontology improvement), extensão de ontologia (ontology extension²), refinamento de ontologia (ontology refinement), especialização de ontologia (ontology specialization³). Vale ressaltar que alguns termos podem ser usados como sinônimos de enriquecimento (expansão, extensão, alargamento), já outros, tratam de processos diferentes (evolução e refinamento, por exemplo). No âmbito deste estudo, utilizaremos o termo “Enriquecimento de Ontologia” em detrimento dos outros, pois é o termo mais comumente utilizado na literatura.

Petasis et al. (2011), define EO como a atividade de expandir uma ontologia com conceitos, relações e regras adicionais, inserindo-os na posição correta na ontologia. O autor diz ainda que é a atividade responsável por ampliar o conhecimento prévio em determinado domínio.

¹ Expansão de ontologia refere-se à atividade de facilitar a modificação da ontologia preservando a sua consistência. (Suárez-Figueroa, 2010).

² Extensão de Ontologia é uma atividade do Enriquecimento de Ontologia responsável por alongar a ontologia em largura. (Suárez-Figueroa, 2010).

³ Especialização de Ontologia é uma atividade do processo de Enriquecimento Ontológico que visa estender a ontologia em profundidade, ou seja, especificar os conceitos da ontologia.

Faatz e Steinmetz (2002) definem EO como a operação baseada em um conjunto de fórmulas que tem o propósito de incluir novos conceitos, propriedades e novas instâncias em ontologias existentes.

Na nossa perspectiva, em complemento às definições apresentadas, EO é um processo semiautomático que dá suporte à área de Manutenção de Ontologia e é responsável pela expansão e especialização da estrutura da ontologia, através da inclusão de novos conceitos e seus símbolos linguísticos alternativos (sinônimos, por exemplo), relações (taxonômicas e não-taxonômicas) e suas propriedades restritivas e axiomas; que devem ser incluídos nos locais corretos na ontologia.

Suárez-Figueroa (2010) associa duas atividades ao processo de Enriquecimento de Ontologia, que são: Extensão da Ontologia é a atividade de alongar a ontologia em largura (horizontalmente), ou seja, trata-se de ampliar a estrutura da ontologia com a inserção de conceitos amplos, que podem configurar uma nova classe ou subclasse na ontologia. Já Especialização da Ontologia é a atividade de estender a ontologia em profundidade (verticalmente), ou seja, inclusão de conceitos mais específicos.

Em busca de uma categorização dos métodos existentes para EO, pode-se caracterizá-los em automáticos, semiautomáticos e manuais, quanto ao grau de automatização ou a execução dos processos sem a interferência de um humano, geralmente um especialista de domínio. Na literatura não existe um método totalmente automático, sendo mais comuns os métodos semiautomáticos. Ainda nesse assunto, Petasis et al. (2011, p.142) diz que “um sistema totalmente automatizado é obviamente desejável, mas nem sempre é possível, especialmente com tarefas relacionadas ao enriquecimento ontológico.”

No que tange ao tipo de conhecimento adquirido para enriquecer as ontologias, pode-se dividir os métodos em:

- enriquecimento léxico: refere-se à identificação de termos, sinônimos e realizações alternativas para determinado conceito;
- enriquecimento conceitual: trata da descoberta de novos conceitos/entidades;
- enriquecimento de relações: identificação de relacionamentos taxonômicos e outros relacionamentos semânticos (ou não taxonômicos) para inclusão na ontologia;
- enriquecimento de axiomas: inclusão de novos axiomas e/ou

restrições na estrutura da ontologia com fins de geração de novos fatos ou inferência de conhecimento.

3. METODOLOGIA

Neste capítulo são apresentadas as características da pesquisa e os procedimentos metodológicos que embasam essa pesquisa.

3.1. Caracterização da pesquisa

Gil (2009) apresenta algumas orientações e características dos diversos tipos de pesquisa. Seguindo o viés do autor, essa pesquisa se caracteriza como aplicada quanto à sua natureza, pois objetiva gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos, ou seja, tem o propósito de desenvolver uma metodologia para enriquecimento de ontologias de domínio para resolver o problema de ontologias desatualizadas, incompletas e que não representam o domínio adequadamente. No que tange aos objetivos, se caracteriza como uma pesquisa exploratória, pois tem a finalidade de apresentar informações sobre o assunto e entendê-lo melhor. Por fim, com relação aos procedimentos metodológicos, se qualifica como uma pesquisa experimental.

3.2. Procedimentos metodológicos

Para o desenvolvimento dessa pesquisa propõem-se as sete etapas apresentadas a seguir. Destaca-se que por se tratar de uma pesquisa em andamento, algumas etapas (etapas um e dois) já foram realizadas e outras ainda estão em planejamento.

Etapas 1. Pesquisa bibliográfica: com o intuito de se familiarizar com o assunto da pesquisa, nesta etapa foi realizada um estudo sobre as temáticas: Sistema de Organização do Conhecimento, ontologia, engenharia ontológica, aprendizagem de ontologia e enriquecimento de ontologia. Foram utilizadas como fonte de pesquisa as bases de dados e periódicos nacionais e internacionais das áreas de Ciência da Informação e Ciência da Computação (visto que grande parte da literatura relacionada às ontologias está nessa área), bibliotecas digitais de teses e dissertações e livros. Entre as bases pesquisadas estão: Portal de Periódicos da Capes, Google Acadêmico, Base de dados de Periódicos em Ciência da Informação (BRAPCI) Biblioteca Digital de Teses e Dissertações (BDTD) do Instituto Brasileiro de Informação em Ciência e Tecnologia (

IBICT), Portal ACM, CiteSeer; além de anais de eventos relevantes na área de ontologias, tais como: Ontobrás, Formal Ontology in Information Systems (FOIS), International Conference on Biomedical Ontology (ICBO).

Etapa 2. Mapeamento por meio de revisão narrativa da literatura das propostas, métodos e metodologias para enriquecimento de ontologias: foi realizado um mapeamento das várias propostas individuais para enriquecimento de ontologias de domínio com o propósito de saber como está o estado da arte da área. Nesta pesquisa, optou-se por realizar uma revisão de literatura do tipo narrativa. Assim, pesquisou-se em bases de dados da Ciência da Informação e Ciência da Computação, trabalhos que abordassem o Enriquecimento de Ontologias. Para tanto, utilizou-se as seguintes expressões (em inglês e português): *Enriquecimento de Ontologia, Enriquecimento Ontológico, Expansão de Ontologia, Extensão de Ontologia, Especialização de Ontologia, Refinamento de Ontologia, Ampliação de Ontologia, Completude de Ontologia, Melhoria de Ontologia*. Além disso, aplicou-se estratégias de busca em bases de dados, como o uso de truncamentos, operadores booleanos, pesquisa avançada e filtros específicos de cada base de dados. O período temporal abordado foi de 1990 a 2018. Optou-se por iniciar a partir de 1990, pois foi a partir dessa década que os estudos sobre ontologias se tornaram mais comuns da Ciência da Computação e Ciência da Informação.

Diante das estratégias de busca abordadas no parágrafo anterior juntamente com a eliminação de duplicidades e de trabalhos que não tinham relação com o escopo desta pesquisa (após análise do título e do resumo), chegou-se ao total de 35 trabalhos. Assim, realizou-se a leitura desses trabalhos e eliminou-se 20 trabalhos com base nos seguintes critérios: (1) trabalhos que não apresentavam o enriquecimento de ontologias e sim de outros Sistemas de Organização do Conhecimento ou de bases de dados (como DBPedia); (2) trabalhos que abordavam outro processo e não o enriquecimento (como aprendizagem de ontologia, evolução); (3) trabalhos que tratavam apenas da técnica de aquisição de conhecimento e não resultavam no enriquecimento de ontologias. Dessa forma, considerou-se 15 trabalhos nesta etapa.

Etapa 3. Estudo das normas e métodos oriundos da Biblioteconomia e Ciência da Informação para atualização e manutenção de Sistemas de Organização do Conhecimento: pretende-se identificar e estudar as principais normas e métodos oriundos da BCI para manutenção e atualização de Sistemas de Organização do Conhecimento, pois acredita-se que eles podem fornecer subsídios válidos para a proposta metodológica que se almeja nesta pesquisa.

Etapa 4. Estudo das metodologias de desenvolvimento de ontologias: objetiva-se identificar e estudar as metodologias de desenvolvimento que abordam o assunto manutenção e/ou atualização de ontologias com o intuito de aproveitar procedimentos que possam ser aplicados na metodologia para enriquecimento de ontologias que está sendo desenvolvida.

Etapa 5. Análise qualitativa e comparativa do conteúdo dos documentos levantados nas Etapas 2, 3 e 4: pretende-se realizar a análise de conteúdo (qualitativa e comparativa) desses estudos, afim de reconhecer as principais características relevantes dos métodos existentes e que possam ser aplicadas na metodologias que se pretende desenvolver. Assim como identificar as principais falhas e lacunas que esses métodos apresentam para que possam ser superados no produto final dessa pesquisa (uma metodologia para enriquecimento de ontologias de domínio).

Etapa 6. Desenvolvimento da metodologia para enriquecimento de ontologias de domínio: com base nos insumos gerados na Etapa 5, pretende-se desenvolver uma metodologia detalhada para enriquecimento de ontologias de domínio. Diante dos estudos e análises realizadas até o momento, é possível esboçar um plano de ação inicial para desenvolvimento da Metodologia para Enriquecimento de Ontologias, apresentado a seguir no Quadro 2.

QUADRO 1. Plano de ação inicial para desenvolvimento da Metodologia de Enriquecimento de Ontologias

#	Etapa	Subetapa	Resultado
1	Verificar se há necessidade de enriquecimento	a) Analisar os objetivos da ontologia, usuários e questões de competência se houver	Documento de especificação
2	Diagnóstico da ontologia com base nos 41 pitfalls de POVEDA-VILLALÓN (2016)	a) Estudar os 41 pitfalls b) Identificar aqueles que possuem relação com o processo de enriquecimento de ontologias c) Aplicar os pitfalls relevantes	Relatório com detalhes de pontos da ontologia que precisam de enriquecimento
3	Identificação das fontes de dados para realização do enriquecimento	a) Explicar como pesquisar fontes	Relatório com fontes potenciais para extração de dados
4	Definição do (s) método (s) de aquisição e extração do conhecimento	b) Escolher os métodos e técnicas para extração do conhecimento	
5	Extração do conteúdo candidato para enriquecimento	a) Solicitar análise de um especialista do domínio do conteúdo candidato.	Relatório detalhado de cada conteúdo e o feedback do especialista.
6	Enriquecimento dos componentes da ontologia a partir do conteúdo extraído	a) Analisar o conteúdo extraído na etapa anterior; b) Incluir o conteúdo na ontologia	Ontologia enriquecida.
7	Avaliação e validação do conteúdo do enriquecimento	a) Verificar a consistência da ontologia após a inclusão do conteúdo.	Relatório com o resultado da avaliação.
8	Documentação da Metodologia	a) Descrever o que foi realizado em cada passo da metodologia.	Documento com todas as etapas da metodologia detalhadas

Fonte: elaborado pelo autor.

Etapa 7. Avaliação da metodologia desenvolvida: com o intuito de validar a metodologia, aplicar-se-á a metodologia em duas ontologias de domínios distintos. Realizar-se-á o enriquecimento dessas ontologia de domínio em todos os seus níveis. A justificativa para testar a metodologia em domínios distintos se deve ao objetivo principal dessa investigação, ou seja, desenvolver uma metodologia que não seja dependente de um domínio específico. A partir desse processo será feito um

diagnóstico da metodologia, ou seja, quais os pontos fortes, fracos e em quais aspectos a metodologia precisaria ser melhorada.

4. CONSIDERAÇÕES INICIAIS

Por se tratar de um resumo estruturado de uma pesquisa em andamento, ainda não existem resultados a serem apresentados e discutidos. Portanto, como considerações iniciais, espera-se atingir o objetivo principal dessa investigação, a saber, desenvolvimento de uma metodologia detalhada para enriquecimento de ontologias domínio; e que a metodologia desenvolvida possa ser útil para ontologistas e especialistas de domínio na tarefa de realizar manutenção em ontologias.

REFERÊNCIAS

FAATZ, A.; STEINMETZ, R. **Ontology Enrichment with Texts from the WWW**. Darmstadt, Germany: [s.n.], 2002.

FERNÁNDEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. **Methontology: from ontological art towards ontological engineering**. AAAI Technical Report SS, 1997.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2009.

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.

PETASIS, G. et al.. **Ontology population and enrichment: State of the art**, " In Knowledge-driven multimedia information extraction and ontology evolution, Springer-Verlag, January 2011, pp. 134-166.

POVEDA-VILLALÓN, M. **Ontology Evaluation: a pitfall-based approach to ontology diagnosis**. Tese (Doutorado), Madrid: Escuela Técnica Superior Ingenieros Informáticos, 2016.

SMITH, B.; KLAGGES, B. **Bioinformatics and Philosophy**. In: MUNN, K.; SMITH, B. **Applied ontology: an introduction**. Frankfurt: Ontos Verlag, 2008, p. 21-38.

SUÁREZ-FIGUEROA, M. C. **NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse**. Tese (Doutorado), Madrid: Facultad de Informática da Universidad Politécnica de Madrid, 2010.

SURE, Y.; STAAB, S.; e STUDER, R. **On-To-Knowledge Methodology (OTKM)**. 2003. Disponível em: http://www.sfu.ca/~mhatala/iat881/papers/2003_ontohandbook_otkmethodology.pdf. Acesso em 27 ago. 2018.

SWARTOUT, B. et al. **Toward distributed use of large-scale ontologies**. 1996. Disponível em: http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html. Acesso em: 13 ago. 2018.

Webert Júnio Araújo

Bibliotecário do Centro Federal de Educação Tecnológica de Minas Gerais e Doutorando do Programa de Pós Graduação em Gestão e Organização do Conhecimento da ECI/UFMG.

Gercina Angela de Lima

Professora Associada no Departamento de Organização e Tratamento da Informação na ECI/UFMG.