

Enriquecimento de ontologias de domínio: uma revisão sistemática da literatura

Webert Júnio Araújo

Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, MG, Brasil;
webert@cefetmg.br; <https://orcid.org/0000-0002-9020-3711>

Gercina Ângela de Lima

Universidade Federal de Minas Gerais, Escola de Ciência da Informação, Belo Horizonte, MG, Brasil;
limagercina@gmail.com; <https://orcid.org/0000-0003-0735-3856>

Resumo: Este artigo apresenta uma revisão sistemática da literatura sobre a temática enriquecimento de ontologias de domínio com o propósito de sintetizar as principais discussões sobre tal tema. Para tanto, a metodologia respalda-se nas diretrizes para revisão sistemática em conjunto com o uso do software Start para gestão e organização dos 29 estudos selecionados em cinco bases de dados. Os dados coletados permitiram identificar algumas características dos estudos publicados sobre enriquecimento de ontologias, tais como: fontes de informação utilizadas, tipos de enriquecimento mais aplicados, tipos de técnicas de extração de informações utilizadas, bem como o grau de detalhamento dos métodos para enriquecimento de ontologias de domínio. Conclui-se que o artigo atingiu o objetivo e que as ontologias de domínio precisam acompanhar a dinamicidade do fenômeno que representa, sendo o processo de enriquecimento uma direção para isso.

Palavras-chave: ontologia de domínio; atualização de ontologia; enriquecimento de ontologia; revisão sistemática da literatura

1 Introdução

Este artigo discorre sobre parte dos resultados da tese de doutorado intitulada “Proposta metodológica para enriquecimento de ontologias de domínio” (ARAÚJO, 2021) e visa apresentar uma revisão sistemática de literatura sobre a temática Enriquecimento de Ontologias de domínio que engloba o período temporal de cerca de duas décadas (2001 a 2019).

As ontologias de domínio são construídas com objetivos determinados por meio da captura de conhecimento de especialistas de domínio, assim como da extração de conteúdo de fontes de informação que explicitam o conhecimento do domínio. Porém, o conhecimento é algo que muda com as novas descobertas, com mudanças de paradigmas ou até mesmo com refutações de teorias consolidadas.

Diante da provisoriedade do conhecimento humano, e como as ontologias são construídas com base nas abstrações que os especialistas fazem da realidade, os desenvolvedores de ontologias precisam garantir que esses instrumentos passem por processo de atualização sempre que houver necessidade, isto é, sempre que as ontologias não estiverem representando o conhecimento para o qual se propõem representar. Em outras palavras, atividades de atualização precisam fazer parte do ciclo de vida das ontologias, visto que o conhecimento sobre a realidade e a realidade não são estanques, são dinâmicos. Conforme pontuado por Tennis (2012) em um artigo sobre análise de domínio, a nossa visão de um domínio muda e os domínios mudam.

Sendo assim, vislumbra-se na temática Enriquecimento de Ontologia uma solução para as ontologias já desenvolvidas, mas que ainda se encontram incompletas e desatualizadas. O Enriquecimento de Ontologias (EO) trata-se de um processo associado a uma área mais ampla “atualização de ontologia”, e tem o propósito de expandir uma ontologia existente de acordo com seus componentes (classes, relações, axiomas).

O Enriquecimento de Ontologias pode aumentar a capacidade das ontologias em distintos aspectos, por exemplo: a ontologia pode representar um domínio de forma mais específica com a inclusão de conceitos característicos do domínio, pode melhorar a inferência automática de conhecimento com a adição de axiomas, pode ser utilizada no contexto da anotação semântica com a inclusão de sinônimos (realizações alternativas, como siglas, por exemplo) para um conceito. Ademais, enriquecimento é um processo menos desgastante (no que tange a tempo, recurso pessoal e econômico) do que construir uma ontologia do ponto inicial. Portanto, ontologias enriquecidas têm o potencial ampliado para a resolução de problemas em várias áreas das atividades humanas.

Tendo em vista tal cenário, este artigo tem por objetivo sintetizar as principais discussões sobre o enriquecimento de ontologias com o respaldo das diretrizes de revisão sistemática de Kitchenham (2004), em conjunto com o uso do software Start¹ para gestão e organização dos estudos.

Nas próximas seções, descreve-se, primeiramente, os fundamentos do enriquecimento de ontologia; depois apresenta-se a metodologia para a realização da revisão sistemática da literatura; em seguida, são feitas as análises e discussões dos resultados; e por fim, as considerações finais.

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, dado que o termo EO ainda não está sedimentado na literatura da Ciência da Informação. Dessa forma, apresenta-se a seguir algumas definições do termo e estabelece-se a abordagem adotada nesta investigação.

2.1 Definições de enriquecimento de ontologias

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. Dessa forma, o processo de EO só faz sentido se a ontologia já existe no domínio.

Já 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.

Guerram e Mellal (2018) dizem que EO consiste em adicionar automaticamente novos conceitos e relações a uma ontologia de domínio construída previamente. Eles ainda destacam o fato de essas novas inclusões serem colocadas em locais relevantes (ou certos) na ontologia.

As definições apresentadas por Petasis *et al.* (2011), Faatz e Steinmetz (2002) e Guerram e Mellal (2018) se completam, mesmo que algumas não

mencionem alguns componentes das ontologias nos quais o enriquecimento pode acontecer. Porém, todas as definições abordam EO em uma mesma perspectiva, a de expandir uma ontologia já existente, seja por meio de novos conceitos, de relações, propriedades ou axiomas. Com relação à inclusão de novas instâncias, existe uma divergência na literatura, haja vista que a atividade de inclusão de novas instâncias é conhecida como “população ou povoamento de ontologia”. Assim, neste estudo, seguiremos o viés que distingue Enriquecimento de Ontologia de População de Ontologia.

Na nossa perspectiva, em complemento às definições apresentadas no início desta seção, EO é um processo manual ou semiautomático, que dá suporte à área de manutenção e atualização de ontologia e é responsável pela expansão e especialização da estrutura da ontologia, por meio da inclusão de novos conceitos e seus símbolos linguísticos alternativos (sinônimos, por exemplo), relações (taxonômicas e não taxonômicas), suas propriedades restritivas e axiomas; que devem ser incluídos nos locais corretos na ontologia. Seguindo a analogia da “camada de bolo” presente em estudos sobre Web Semântica, o enriquecimento seria os preenchimentos que podem ser feitos no bolo (organizar as fatias) e os complementos que podem ser adicionados a ele sem a necessidade de refazê-lo, ou seja, a inclusão de confetes, chantilly, pasta americana, entre outros acessórios que tornam o bolo mais atrativo e útil para outros contextos.

2.2 Tipos e técnicas de enriquecimento de ontologias

Em busca de uma categorização dos métodos de EO, quanto ao grau de automatização, pode-se caracterizá-los em automáticos, semiautomáticos e manuais (geralmente realizados com o auxílio de um especialista de domínio). Na literatura são poucos os métodos totalmente automáticos, sendo mais comuns os métodos semiautomáticos. Ainda nesse assunto, Petasis *et al.* (2011, p. 142, tradução nossa) dizem que “um sistema automatizado é obviamente desejável, mas nem sempre é possível, especialmente com tarefas relacionadas ao enriquecimento ontológico”.

Outra possibilidade de categorizar os métodos é conforme o tipo de fonte de onde se extraem as informações para enriquecimento. Essas fontes podem ser

divididas em estruturadas, semiestruturadas e não estruturadas. Nesta pesquisa, optou-se pela categorização dos métodos consoante o tipo de componente da ontologia que será enriquecido.

Quanto ao tipo de informação adquirida para enriquecer as ontologias, pode-se dividir os métodos em:

- a) enriquecimento lexical - refere-se à identificação de termos, sinônimos e realizações alternativas para determinado conceito, que nas ontologias são inseridos como propriedades de anotação;
- b) enriquecimento conceitual - trata da descoberta de novos conceitos/entidades, inseridos nas ontologias como classes e subclasses;
- c) enriquecimento de relações - refere-se à identificação de relacionamentos taxonômicos e outros relacionamentos semânticos (ou não taxonômicos) para inclusão na ontologia;
- d) enriquecimento de axiomas - trata da inclusão de novos axiomas e/ou restrições na estrutura da ontologia com fins de geração de novos fatos ou inferência de conhecimento.

No que concerne às técnicas utilizadas para EO, elas se baseiam principalmente no uso de Processamento de Linguagem Natural (PLN), métodos estatísticos e algoritmos de *Machine learning*.

3 Metodologia

Essa pesquisa possui caráter exploratório, descritivo e apresenta uma revisão sistemática de literatura (RSL). A RSL é uma metodologia rigorosa para procurar evidências. Possui questões claramente feitas para identificar e avaliar as pesquisas sobre determinado assunto, para, posteriormente, organizar e analisar os dados dos estudos recuperados que serão incluídos na revisão. Este método se caracteriza por garantir à pesquisa científica maior confiabilidade, posto que possibilita uma pesquisa detalhada da literatura e avaliação crítica de estudos individuais.

Para a execução da revisão em questão, adaptaram-se as diretrizes para revisão sistemática de Kitchenham (2004) e utilizou-se o software Start para gestão e organização dos estudos. Sendo assim, as etapas da revisão foram

divididas em duas grandes fases: (A) Planejamento da revisão e (B) Condução da revisão. A fase de Planejamento da revisão é composta de duas etapas: (1) identificação da necessidade da revisão e (2) desenvolvimento do protocolo da revisão. E a fase de Condução da revisão é composta das etapas²: (1) identificação de estudos; (2) seleção dos estudos; (3) extração dos dados e monitoramento; e (4) síntese dos dados.

Na etapa A.1 - identificação da necessidade de revisão, realizou-se uma análise exploratória da literatura sobre enriquecimento de ontologias e, diante do grande número de estudos, percebeu-se a necessidade da revisão para coletar de forma organizada os estudos publicados sobre a aplicação de métodos de enriquecimento em ontologias de domínio.

Na etapa A.2 - desenvolvimento do protocolo da revisão, desenvolveram-se as (1) questões de pesquisa da revisão, (2) seleção das bases de dados, (3) criação das *strings* de busca e (4) criação dos critérios de inclusão e exclusão. Vale ressaltar que, antes do preenchimento do protocolo, realizou-se uma pesquisa exploratória no Google Acadêmico (GA) com o intuito de identificar as bases de dados nas quais havia uma representatividade maior de estudos sobre a temática, bem como identificar as palavras-chave mais utilizadas para representar esses estudos.

As questões de pesquisa que a revisão de literatura objetiva responder são: (1) quais são os estudos que aplicaram métodos para enriquecimento de ontologias de domínio publicados entre 2000 e 2019? (2) quais as características dos métodos para enriquecimento de ontologias de domínio? (3) quais as lacunas existentes nesses métodos? (4) quais são os aspectos positivos dos métodos existentes? O período temporal foi definido com base na pesquisa exploratória no Google Acadêmico, que revelou que os primeiros estudos surgiram a partir dos anos 2000.

As bases de dados selecionadas para levantamentos dos estudos foram as seguintes: ACM Digital Library, Google Acadêmico (GA), IEEE, Scopus e Web of Science (WoS). Também foram realizadas consultas pré-testes nas bases Library and Information Science Abstracts (LISA) e Information Science & Technology Abstracts (ISTA). No entanto, a quantidade de registros recuperados

foi ínfima, e os mesmos já estavam contemplados nas bases de dados selecionadas anteriormente. Sendo assim, essas duas bases não foram incluídas na seleção.

Após a seleção das bases de dados, passou-se à elaboração das *strings* de busca, que continham os termos mais relevantes para recuperação dos estudos. Observa-se, portanto, que cada base de dados demandou uma *string* que se adequasse aos parâmetros específicos daquela base. Contudo, a diretriz da *string* se manteve, ou seja, buscar por estudos que continham as palavras “ontologia” e “enriquecimento”, bem como suas variações no campo “título”. No Quadro 1, apresentam-se as *strings* de busca separadas por base de dados.

Quadro 1 - *Strings* de busca por base de dados

Base de dados	String
ACM	Publication title (ontolog*) AND publication title (enrich*)
GA	Allintitle: ontology enrichment
IEEE	Document title "ontolog*" AND document title "enrich*"
Scopus	title(ontolog* AND enrich*)
WoS	TI=(ontolog* AND enrich*)

Fonte: Araújo (2021).

Na última atividade do desenvolvimento do protocolo, definiram-se os critérios de exclusão (E) e inclusão (I). A definição dos critérios foi feita com base no que se almeja com a revisão, ou seja, no seu objetivo, que se refere à identificação de estudos que aplicaram o processo de enriquecimento em ontologias de domínio. A seguir, no Quadro 2, estão os quatro critérios de inclusão e os sete critérios de exclusão.

Quadro 2 - Critérios de inclusão e exclusão adotados para seleção dos documentos

Critérios de inclusão (I)	Critérios de exclusão (E)
Título possui relação com a temática enriquecimento de ontologia de domínio;	Título não possui relação com a temática enriquecimento de ontologias de domínio;
Resumo possui relação com a temática enriquecimento de ontologia de domínio;	Resumo não possui relação com a temática enriquecimento de ontologia de domínio;
Serão incluídos trabalhos publicados e disponíveis integralmente nas bases de dados científicas buscadas;	Serão excluídos trabalhos que não apresentem resumo/ <i>abstract</i> ;
Serão incluídos trabalhos que tenham aplicado um método de enriquecimento em ontologia de domínio;	Serão excluídos <i>short papers</i> e trabalhos disponibilizados em bases <i>pre-print</i> ;
-	Serão excluídos documentos que não estejam em português ou inglês;
-	Serão excluídos trabalhos que não estejam disponíveis em texto completo;
-	Serão excluídos os trabalhos que não apliquem o processo de enriquecimento em ontologias de domínio.

Fonte: Araújo (2021).

Na fase B - Condução da revisão, etapa B.1 - identificação de estudos, realizaram-se pesquisas nas cinco bases de dados selecionadas e chegou-se ao total de 880 estudos, a partir da aplicação do protocolo desenvolvido na fase anterior.

Na etapa B.2 - seleção dos estudos, com o uso do software Start, realizou-se a análise/filtragem dos 880 registros recuperados nas bases de dados ACM, Google Acadêmico, IEEE, Scopus e Web of Science. A primeira atividade executada foi a exclusão de 454 duplicatas. Em seguida, procedeu-se à análise do título³ dos 427 registros restantes. Assim, foram realizadas a leitura dos títulos desses estudos e a classificação segundo os seguintes critérios de inclusão e exclusão: (Inclusão) título possui relação com a temática enriquecimento de ontologia de domínio; (Exclusão) título não possui relação com a temática enriquecimento de ontologias de domínio.

Realizada a análise do título, foram excluídos 260 estudos, restando 167 para a próxima filtragem, análise do resumo. Nessa atividade, realizou-se a leitura do resumo dos 167 estudos e classificação de acordo com os seguintes critérios de inclusão e exclusão: (I) resumo possui relação com a temática enriquecimento de ontologia de domínio; (E) serão excluídos trabalhos que não apresentem resumo/*abstract*; (E) resumo não possui relação com a temática enriquecimento de ontologia de domínio. Após a análise do resumo, foram excluídos 83 estudos, restando, assim, 84 estudos para a próxima análise.

A análise seguinte refere-se à leitura do texto completo dos 84 estudos restantes. Eles foram analisados com base nos seguintes critérios: (I) serão incluídos trabalhos publicados e disponíveis integralmente nas bases de dados científicas buscadas; (E) serão excluídos trabalhos que não estejam disponíveis em texto completo; (E) serão excluídos trabalhos que não estejam em português ou inglês; (E) serão excluídos *short papers* e trabalhos disponibilizados em bases *pre-print*. Assim, após essa análise, 36 estudos foram excluídos, conforme os critérios de exclusão mencionados; restaram 48 estudos para a última triagem.

Os textos completos dos 48 estudos restantes foram analisados e classificados de acordo com os seguintes critérios: (I) serão incluídos trabalhos que tenham aplicado um método de enriquecimento em ontologia de domínio; (E) serão excluídos os trabalhos que não aplicaram o processo de enriquecimento em ontologias de domínio. Após essa análise, restaram 29 estudos, considerados o *corpus* para a revisão de literatura e dos quais foram extraídos os dados que caracterizam os estudos.

Na etapa B.3 - extração dos dados e monitoramento, com o uso das funcionalidades do software Start, desenvolveram-se formulários de extração de dados que apresentavam os seguintes campos: fonte de informação, tipo de enriquecimento, técnica aplicada, detalhamento das etapas, principais lacunas e aspectos favoráveis dos estudos. Sendo assim, esses campos serviram de guia para a extração dos dados dos estudos selecionados. Ademais, extraíram-se os dados quantitativos para caracterização dos estudos.

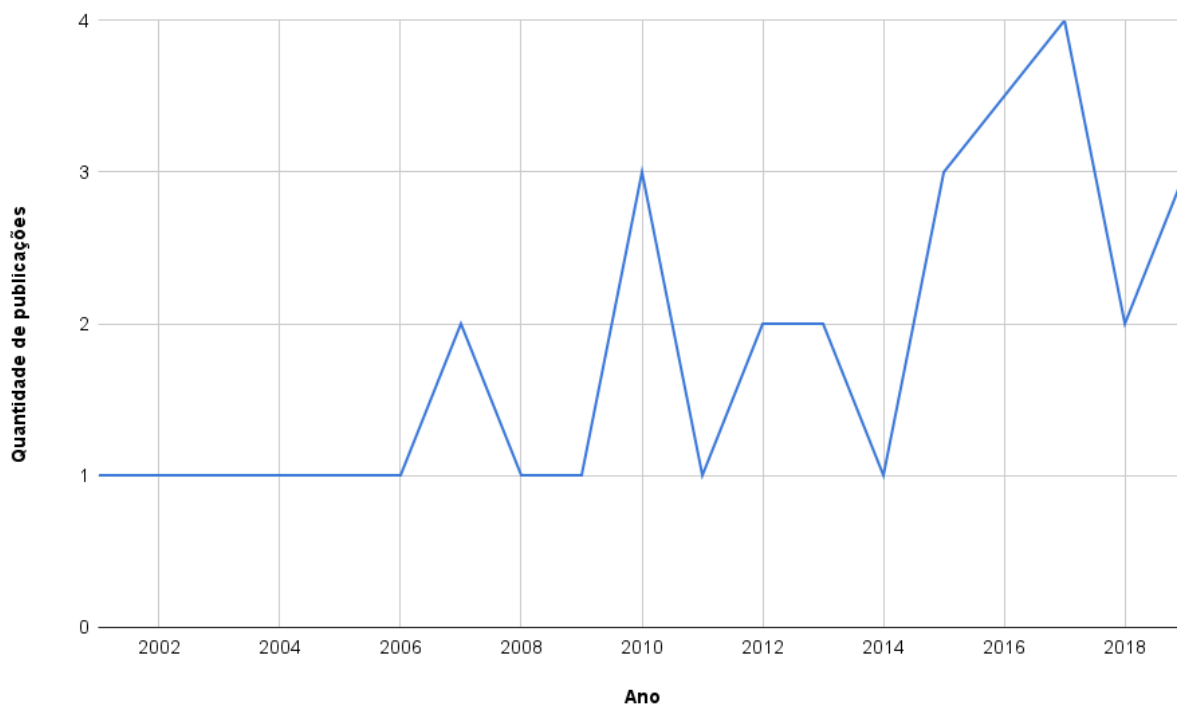
A etapa B.4 - síntese dos dados, refere-se à organização e escrita dos dados que foram extraídos na etapa anterior. Na próxima subseção apresentam-se os principais dados coletados na revisão de literatura.

4 Resultados e discussão

Nesta seção, apresentam-se dados quantitativos e qualitativos dos 29 estudos selecionados conforme parâmetros apresentados na seção anterior. Discorre-se sobre os seguintes aspectos: número de publicações por ano; tipologia documental; país de origem dos autores; fonte de informação utilizada; tipos de enriquecimento; tipos de técnicas aplicadas nos estudos; informações sobre o detalhamento dos métodos; limitações e aspectos favoráveis dos estudos.

No que se refere ao número de publicações por ano, percebe-se que existe uma certa equivalência e distribuição das publicações entre os anos. Os anos 2010, 2015 e 2019 possuem três publicações (conforme Gráfico 1). Somente o ano de 2017 teve uma maior ocorrência, com quatro publicações.

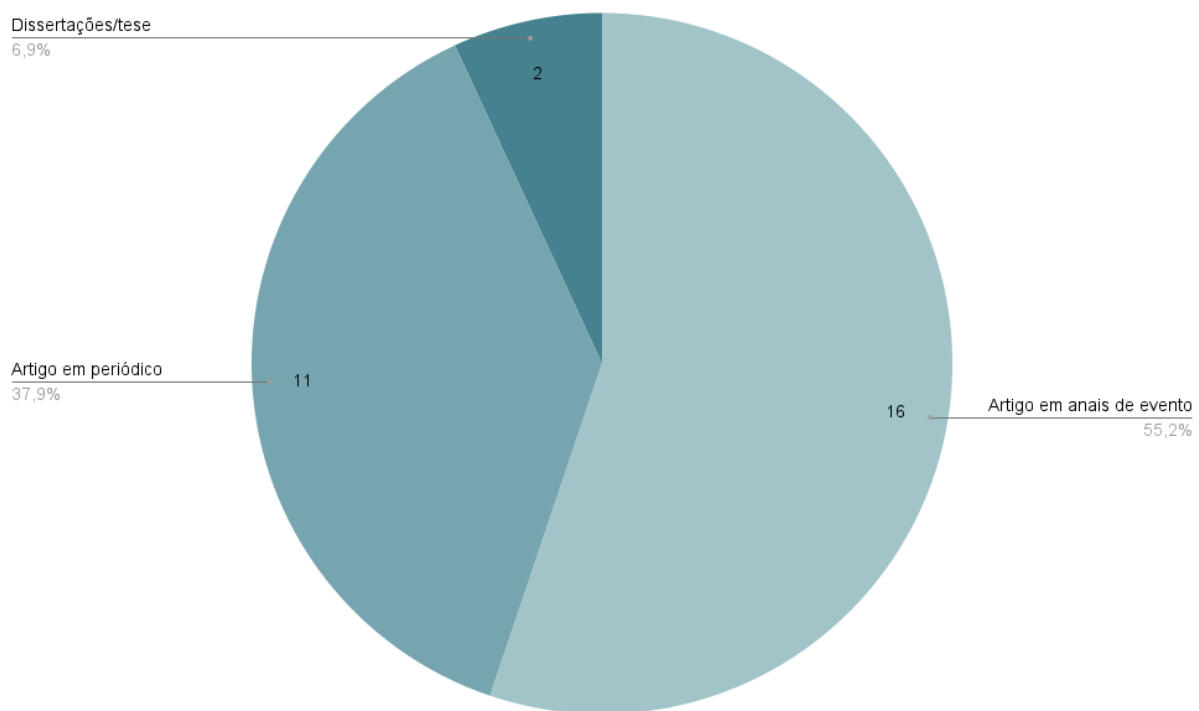
Gráfico 1 - Quantidade de publicações por ano



Fonte: Araújo (2021).

Os dados referentes à tipologia documental, apresentados no Gráfico 2, revelam que 37,9% das publicações são originárias de artigos de periódicos. No entanto, a maioria das publicações (52,2%) é proveniente de anais de eventos científicos, enquanto 6,9% referem-se a teses e dissertações. A explicação para esses dados pode ter relação com o fato de a temática enriquecimento de ontologia ainda ser um assunto emergente, que está em desenvolvimento. Assim, é mais comum a publicação dessas temáticas em anais de eventos.

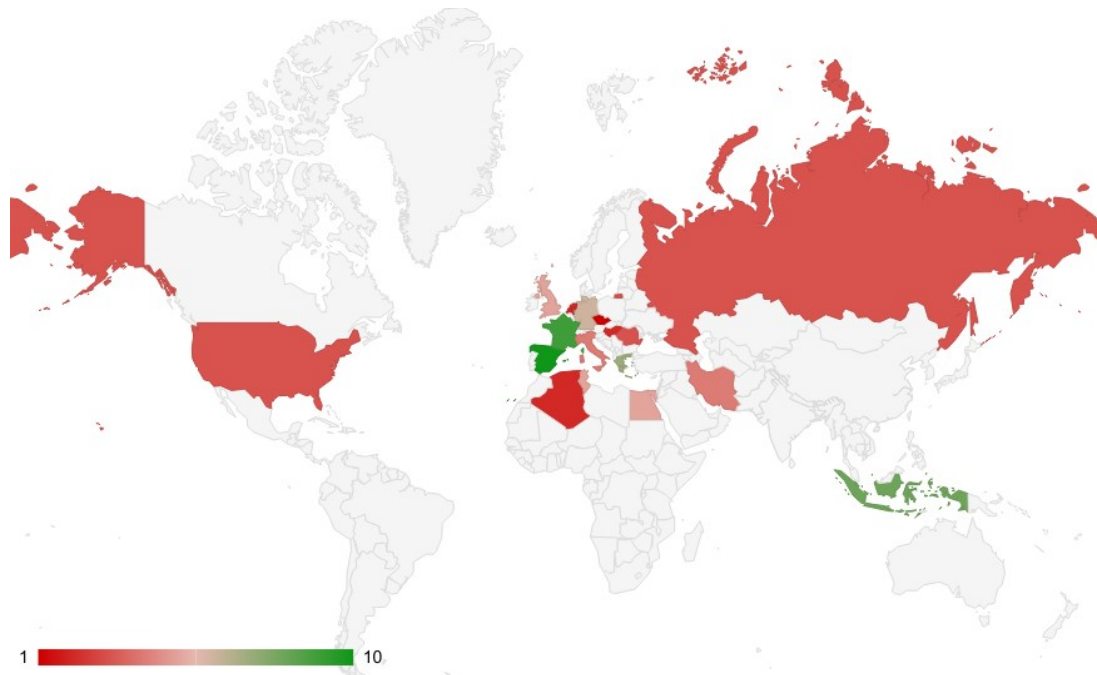
Gráfico 2 - Quantitativo de publicação por tipologia documental



Fonte: Araújo (2021).

Com relação ao país de origem dos autores, há uma distribuição em 17 países. No entanto, não foi selecionado nenhum artigo de pesquisadores brasileiros, considerando os parâmetros adotados na presente pesquisa, o que reforça a importância do presente estudo. A Figura 1 mostra uma concentração de publicações nos países europeus, sendo que a Espanha é o país que mais possui autores (dez) que publicaram sobre enriquecimento, seguido da França, com nove autores. A Indonésia possui oito autores; e a República Tcheca, apenas um autor.

Figura 1 - Origem dos autores dos documentos selecionados



Legenda: Espanha (10); França (9); Indonésia (8); Grécia (7); Alemanha (6); Egito (5); Reino Unido (5); Tunísia (5); Irã (4); Itália (4); Estados Unidos (3); Romênia (3); Rússia (3); Argélia (2); Holanda (2); Hungria (2); República Tcheca (1)

Fonte: Araújo (2021).

Apesar de a literatura sobre métodos para aquisição de conhecimento de forma automática e semiautomática (MAEDCHE, 2002; SHAMSFARD; BARFOUROUSH, 2004; BUITELAAR; CIMIANO; MAGNINI, 2005; CIMIANO; VÖLKER, 2005) ser profícua em estudos (também conhecidos como métodos de Aprendizagem de Ontologia), nesta pesquisa não se considerou esse tipo de trabalho, haja vista que esses estudos focam em adquirir conhecimento para o desenvolvimento de uma ontologia do início ou apenas focam na tarefa de extrair conhecimento, mas não discorrem propriamente sobre o enriquecimento da ontologia. O propósito desta revisão de literatura foi a identificação de métodos e metodologias de enriquecimento de ontologias que foram realmente aplicados, não apenas estudos que focam na técnica de aquisição de conhecimento.

Os primeiros estudos sobre Enriquecimento de Ontologias surgiram a partir dos anos 2000, quando se identificou a necessidade de atualizar, corrigir ou aperfeiçoar as ontologias já desenvolvidas. Assim foram concebidas várias atividades vinculadas à manutenção de ontologias.

Mediante a análise da literatura, percebe-se um grande volume de propostas e métodos para enriquecimento de ontologias provenientes, principalmente, de estudiosos da área de Ciência da Computação. A seguir, dedica-se a descrever as características principais dos estudos.

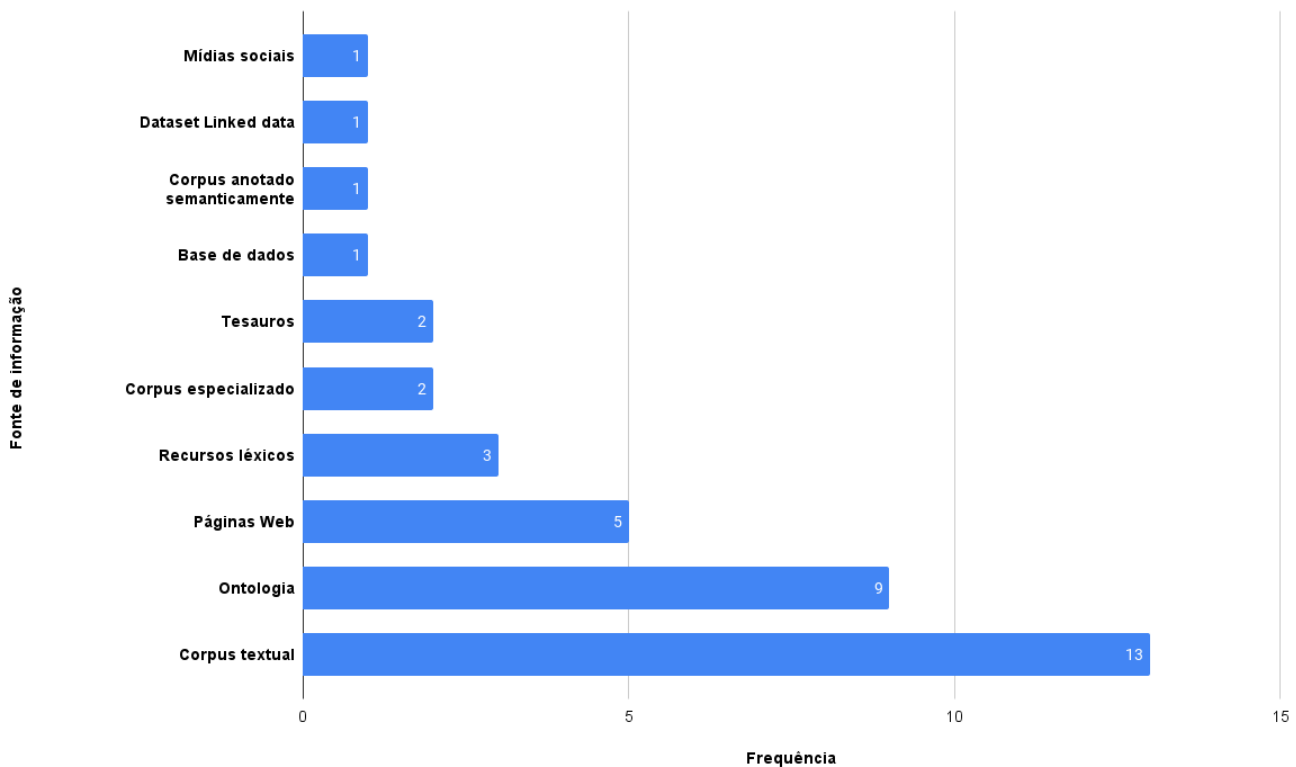
Percebe-se, com base nas características apresentadas de cada um dos 29 estudos selecionados nesta revisão de literatura, que existe uma grande heterogeneidade de métodos para enriquecimento de ontologias. Todos eles apresentam uma característica própria e não existe nenhum método igual. No entanto, também existem algumas similaridades entre os métodos: todos partem de uma ontologia inicial, todos se baseiam em alguma fonte para extração de informações para enriquecimento (ainda que os tipos de fontes de informações escolhidas difiram); todos propõem pelo menos o enriquecimento de um componente da ontologia; todos eles se baseiam em uma ou mais técnicas para extração da informação. Esses são os elementos básicos do processo de enriquecimento, posto que não existe enriquecimento sem uma ontologia já desenvolvida; também é preciso extrair as informações de alguma fonte (ainda que a fonte de informação seja a própria ontologia). Por conseguinte, utiliza-se, no mínimo, uma técnica para extração da informação e, por fim, é preciso enriquecer ao menos um tipo de componente da ontologia.

Para coletar dados referentes aos 29 estudos selecionados e descritos nesta revisão de literatura, foram tabulados dados das seguintes categorias: (1) fonte de informação, (2) tipo de enriquecimento (com as subcategorias lexical, conceitual, de relações taxonômicas, de relações não taxonômicas e de axiomas) e técnica(s) aplicada(s), (3) detalhamento das etapas (com os campos, sim, não e parcialmente) e (4) principais lacunas dos métodos e aspectos favoráveis.

O Gráfico 3 - Frequência das fontes de informação utilizadas para enriquecimento apresenta a sistematização dos métodos quanto à fonte de informação utilizada para extração das informações. Percebe-se que a fonte de informação mais comum para extração da informação para enriquecimento são os *corpora* textuais. Isso tem relação com o fato de ser o tipo de fonte de mais fácil acesso, com maior disponibilidade em vários domínios do conhecimento. Ademais, os documentos textuais são importantes fontes de onde podem-se

extrair aspectos linguísticos que podem ser transformados em aspectos semânticos na estrutura da ontologia por meio dos relacionamentos entre os conceitos. O Gráfico 3, mostra, também, que as ontologias são muito utilizadas para extrair informações passíveis de contribuir com o enriquecimento.

Gráfico 3 - Frequência das fontes de informação utilizadas para enriquecimento



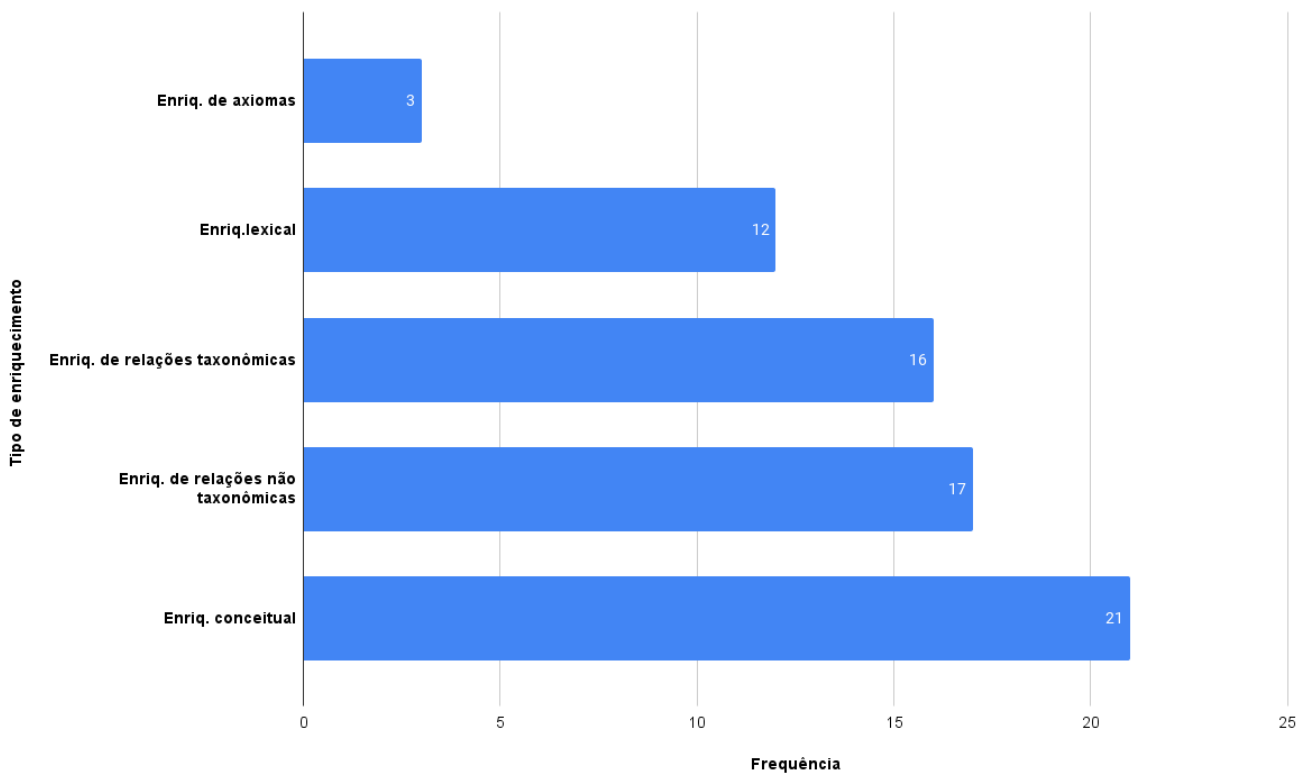
Fonte: Araújo (2021).

Percebeu-se, nos estudos, que o uso de ontologias como fonte de informação é muito comum no domínio biomédico, mais especificamente com relação às ontologias que fazem parte do Consórcio *Open Biomedical Ontologies* (OBO). Essas ontologias possuem características comuns e se fundamentam em uma mesma ontologia de alto nível, a *Basic Formal Ontology* (BFO), facilitando a aplicação de algumas técnicas de extração da informação.

O Gráfico 4 - Frequência dos tipos de enriquecimentos realizados revela que o enriquecimento conceitual é o tipo mais comum nos estudos analisados. Isso tem relação com o fato de esse tipo de enriquecimento exigir técnicas menos complexas e mais comuns na literatura; além disso, é mais simples a identificação

de elementos linguísticos e a transformação desses elementos em conceitos (classes e subclasses na ontologia), se compararmos com o enriquecimento de outros componentes, que exigem técnicas mais aprimoradas. Já o enriquecimento de axiomas é o que menos aparece nos estudos, com apenas três estudos que propõem o enriquecimento de tal componente. Argumenta-se que as razões para isso também têm relação com a complexidade, posto que o enriquecimento desse tipo de componente é dificilmente realizado de forma automática, visto que exige mais capacidade de análise, necessitando de um especialista para adicionar os axiomas.

Gráfico 4 - Frequência dos tipos de enriquecimentos realizados

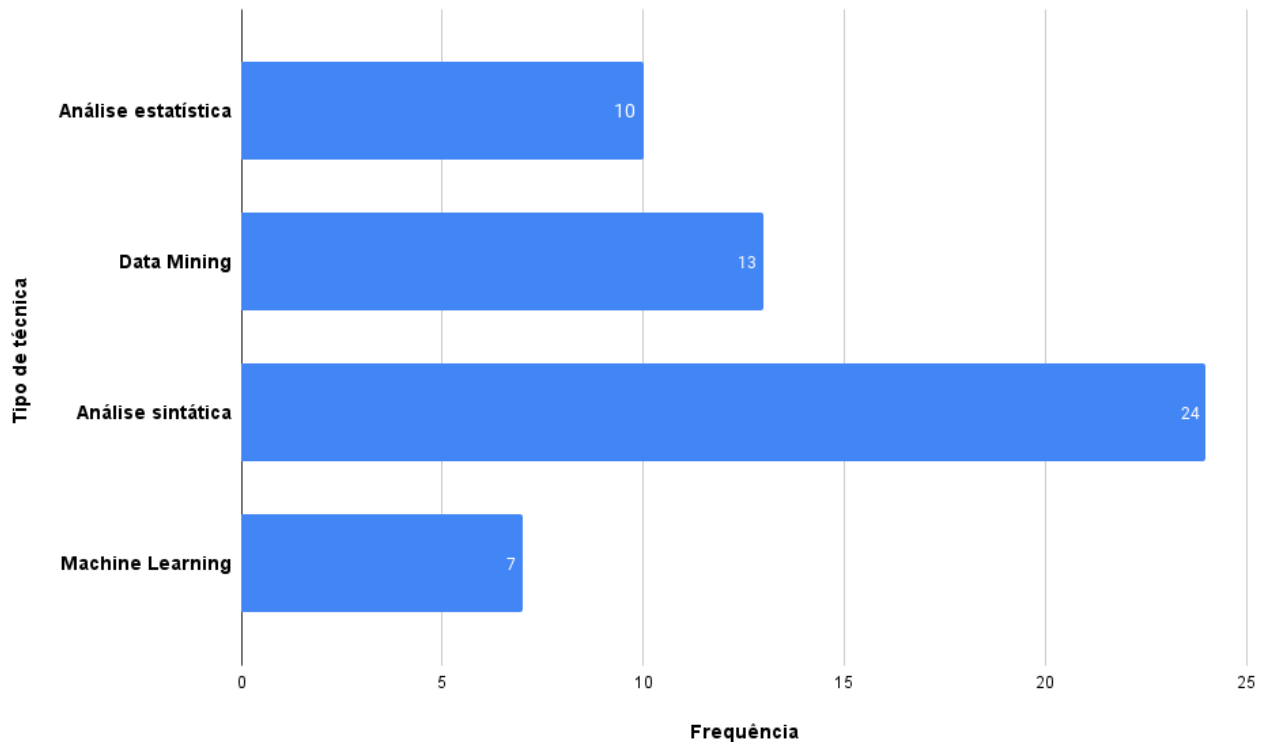


Fonte: Araújo (2021).

O Gráfico 5 - Frequência dos tipos de técnicas utilizadas para extração da informação revela que existe uma grande diversidade de técnicas. Portanto, para facilitar a identificação, podemos classificar essas técnicas em: análise sintática, análises estatísticas, *Data mining* e *Machine learning*. Percebe-se que as técnicas do tipo análise sintática são as mais recorrentes. Isso se deve à existência de

inúmeras técnicas ligadas à análise sintática (ou técnicas de Processamento de Linguagem Natural), tais como tokenização, *parsing*, *part-of-speech tagging*, lematização, etc.

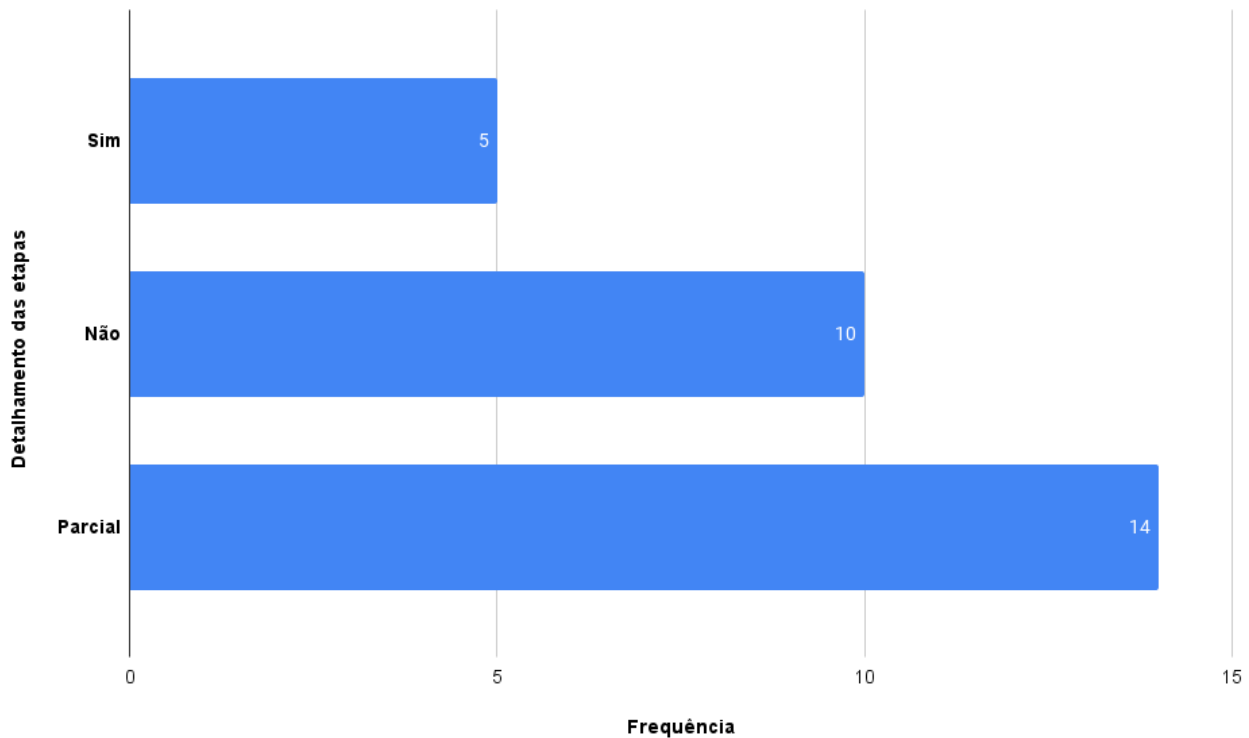
Gráfico 5 - Frequência dos tipos de técnicas utilizadas para extração da informação



Fonte: Araújo (2021).

Com relação ao detalhamento das etapas de cada proposta, o Gráfico 6 revela que a maioria (48,3%) delas apresenta um detalhamento parcial sobre as etapas do método proposto. Destaca-se também que uma grande parcela (34,5%) das propostas não detalha como o método de enriquecimento foi aplicado, impossibilitando, assim, a sua replicação.

Gráfico 6 - Detalhamento das etapas do método nos estudos



Fonte: Araújo (2021).

A partir da revisão de literatura, perceberam-se algumas limitações dos métodos propostos para enriquecimento de ontologias. Entre as principais, estão:

- a) técnicas que exigem a supervisão de um especialista - as técnicas empregadas no método exigem um linguista computacional para sua aplicação;
- b) métodos muito específicos para o domínio biomédico - as técnicas e as fontes utilizadas no método não podem ser generalizadas para outros domínios, visto que foram construídas especificamente para aplicação no domínio biomédico;
- c) resultado insatisfatório do enriquecimento totalmente automático - métodos que não possuem a intervenção de um especialista tendem a não produzir resultados satisfatórios;
- d) falta de previsão do enriquecimento de todos os componentes - é importante que um método possibilite o enriquecimento de todos os componentes das ontologias de domínio;

- e) ausência de detalhamento dos passos para execução do método - as formas como as atividades devem ser desenvolvidas precisam ser claras e objetivas para facilitarem o trabalho do ontologista;
- f) dificuldade de acesso às fontes de informação por serem muito específicas - as fontes de informação não podem ser de difícil acesso. O método deve oferecer opções de fontes de informação;
- g) validação em um cenário limitado - a atividade de validação ou avaliação do método de enriquecimento deve possibilitar avaliar todos os aspectos do método.

No entanto, outros dados que emergiram da análise dos documentos referem-se aos aspectos favoráveis, que se destacam como critérios almejados por um método completo. Entre os aspectos favoráveis, aparecem:

- a) independência de software ou ferramenta específica - como as tecnologias da informação e comunicação evoluem rapidamente, softwares podem ser descontinuados ou podem se tornar obsoletos;
- b) independência de domínio - uma característica importante que os métodos devem possuir é a possibilidade de serem aplicados em diferentes domínios do conhecimento para ocorrência da generalização do método;
- c) uso de fonte de informação de fácil acesso - as fontes de onde são extraídas as informações para enriquecimento das ontologias devem ser de fácil acesso em qualquer domínio do conhecimento, visto que se trata da matéria-prima para realização do processo de enriquecimento;
- d) etapas do método bem descritas - o método deve apresentar de forma detalhada o passo a passo para a sua execução, possibilitando, assim, a replicação;
- e) enriquecimento da ontologia por completo - o método deve possibilitar o enriquecimento de todo o conteúdo da ontologia e não apenas de uma parcela;
- f) aplicação do enriquecimento de axiomas - o método precisa propor enriquecimento de axiomas, proposta de enriquecimento raro na literatura.

A literatura apresenta alguns estudos autodenominados propostas de Enriquecimento de Ontologias, porém o enriquecimento é feito em bases léxicas,

como o WordNet (AGIRRE *et al.*, 2000; GHARIB *et al.*, 2012); em *datasets* de Linked Data, como a DBpedia (TÖPPER; KNUTH; SACK, 2012); ou propõe o enriquecimento com instâncias (VALARAKOS *et al.*, 2003), que foge da proposta deste trabalho, haja vista que aqui se considera a tarefa de adicionar instâncias a uma ontologia como População de Ontologias.

Percebe-se que existem várias iniciativas para Enriquecimento de Ontologias de domínio, porém, elas ainda se encontram dispersas e focadas na solução de problemas específicos. A solução para isso é o desenvolvimento de um método ou metodologia genéricos (aplicável a qualquer ontologia de domínio e para todos os tipos de enriquecimentos) para enriquecimento de ontologias. Nesse contexto, acredita-se que esta investigação produziu alguns insumos em tal direção.

5 Considerações finais

Neste artigo, discorreu-se sobre o Enriquecimento de Ontologia, temática pertinente para o ciclo de vida das ontologias de domínio e que ainda é um assunto que não está sedimentado na literatura. Sendo, portanto necessário a realização de estudos que contribuam para o entendimento e caracterização da temática. Para tanto, realizou-se uma revisão sistemática de literatura utilizando as diretrizes de Kitchenham (2004) e o software Start para gestão e organização dos estudos.

Nesse intuito, a revisão sistemática da literatura partiu da busca de estudos sobre métodos de enriquecimento aplicados em ontologias de domínio, obtendo-se, primeiramente, 880 registros. No entanto, após aplicar os critérios de exclusão e inclusão na seleção dos documentos, resultaram-se em 29 estudos, no período de 19 anos, para compor o portfólio da revisão de literatura. Ressalta-se que os primeiros estudos sobre Enriquecimento de Ontologias surgiram a partir dos anos 2000, quando se identificou a necessidade de atualizar, corrigir ou aperfeiçoar as ontologias já desenvolvidas.

A análise quantitativa desses estudos evidenciou que a maioria das publicações são provenientes de anais de eventos científicos, com uma maior concentração de autores europeus, principalmente, da Espanha e França, sendo a

maioria dos pesquisadores da área de Ciência da Computação. Contudo, não houve nenhum estudo de pesquisadores brasileiros entre os 29 selecionados.

A análise qualitativa dos dados constatou a existência de distintos métodos de Enriquecimento de Ontologias com características específicas, algumas limitantes e outras favoráveis, que apontam as necessidades de melhorias, bem como os caminhos que estão funcionando no processo de enriquecimento. Ficou evidente que a fonte de informação mais comum para extração da informação para enriquecimento são os *corpora* textuais, seguida do uso das próprias ontologias, principalmente no domínio biomédico, sendo o enriquecimento conceitual o mais frequente. Constatou-se, também, que existe uma grande diversidade de técnicas para extração da informação, podendo ser por meio do uso da análise sintática, análises estatísticas, *Data mining* e *Machine learning*. Dessas, as técnicas do tipo análise sintática são as mais recorrentes.

Por meio da análise dos estudos foi possível evidenciar alguns aspectos favoráveis nos métodos de enriquecimento: (1) independência de software ou ferramenta específica, (2) independência de domínio, (3) uso de fonte de informação de fácil acesso, (4) etapas do método bem descritas, (5) enriquecimento da ontologia por completo, (6) aplicação do enriquecimento de axiomas.

No entanto, perceberam-se também algumas limitações dos métodos propostos para enriquecimento de ontologias, a saber: (1) técnicas que exigem a supervisão de um especialista, (2) métodos muito específicos para o domínio biomédico, (3) resultado insatisfatório do enriquecimento totalmente automático, (4) falta de previsão do enriquecimento de todos os componentes, (5) ausência de detalhamento dos passos para execução do método, (6) dificuldade de acesso às fontes de informação por serem muito específicas, (7) validação em um cenário limitado.

As ontologias de domínio estão suscetíveis a limitações por serem construídas com base em conceitualizações ou abstrações que determinado grupo de especialistas faz de um fenômeno do mundo físico ou social. Primeiro, porque a forma de conceber o mundo está sujeita a alterações. Segundo, porque o próprio mundo físico e social se transforma e evolui, caracterizando-se como um sistema

dinâmico. Diante disso, as ontologias de domínio precisam acompanhar a dinamicidade do fenômeno que representa, e o processo de enriquecimento é uma direção para isso.

Referências

AGIRRE, E. *et al.* Enriching very large ontologies using the WWW. **arXiv**: preprint cs/0010026, New York, 2000. Disponível em: <https://doi.org/10.48550/arXiv.cs/0010026>. Acesso em: 25 jun. 2021.

ARAÚJO, W. J. **Proposta metodológica para enriquecimento de ontologias de domínio**. 2021. Tese (Doutorado em Gestão e Organização do Conhecimento) - Programa de Pós-Graduação em Gestão e Organização do Conhecimento, Escola de Ciência da Informação da Universidade Federal de Minas Gerais, Belo horizonte, 2021.

BUITELAAR, P.; CIMIANO, P.; MAGNINI, B. **Ontology learning from text: methods, evaluation and applications**. Amsterdam: IOS Press. 2005.

CIMIANO, P.; VÖLKER, J. Text2Onto: a framework for ontology learning and data-driven change discovery. *In*: INTERNATIONAL CONFERENCE ON NATURAL LANGUAGE PROCESSING AND INFORMATION SYSTEMS, 10., 2005, Alicante. **Processing** [...]. Berlin: Springer, 2005. p. 227-238. Disponível em: https://doi.org/10.1007/11428817_21. Acesso em: 25 jun. 2021.

FAATZ, A.; STEINMETZ, R. **Ontology enrichment with texts from the WWW**. *In*: PROCEEDINGS OF ECML-SEMANTIC WEB MINING 2002, Helsinki. **Processing** [...]. Darmstadt: Technische Universität/KOM, 2002.

GHARIB, T. F. *et al.* Enriching ontology concepts based on texts from WWW and corpus. **Journal of Universal Computer Science**, Austria, v. 18, n. 16, p. 2234-2251, 2012.

GUERRAM, T.; MELLAL, N. A domain independent approach for ontology semantic enrichment. *In*: INTERNATIONAL CONFERENCE ON NATURAL LANGUAGE PROCESSING, 7., 2018, Ljubljana. **Conference** [...]. Ljubljana: Association for Computational Linguistics, 2018. p. 13-19.

KITCHENHAM, B. Procedures for performing systematic reviews. **Technical Report Keele University**, United Kingdom, jul. 2004.

MAEDCHE, A. **Ontology learning for the semantic web**. New York: Springer, 2002.

PETASIS, G. *et al.* Ontology populati on and enrichment: state of the art. *In*: **Knowledge-drive multimedia information extraction and ontology evolution**. Berlin: Springer-Verlag, 2011. p. 134-166.

SHAMSFARD, M; BARFOROUSH, A.A. Learning ontologies from natural language texts. **Human-Computer Studies**, United Kingdom, v. 60, n. 1, p. 17-63, 2004.

TENNIS, J. T. Com o que uma análise de domínio se parece no tocante a sua forma, função e gênero? **Brazilian Journal of Information Science**, Marília, v. 6, n. 1, p. 3-15, 2012. Disponível em: <https://doi.org/10.36311/1981-1640.2012.v6n1.02.p3>. Acesso em: 4 jan. 2020.

TÖPPER, G.; KNUTH, M.; SACK, H. DBpedia ontology enrichment for inconsistency detection. *In*: INTERNATIONAL CONFERENCE ON SEMANTICS, 8., 2012, Austria. **Proceedings** [...]. New York: ACM, 2012.

VALARAKOS, A. G. *et al.* A methodology for enriching a multi-lingual domain ontology using machine learning. *In*: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE IN RECENT ADVANCES IN NLP (RANLP), 2003, Borovest. **Proceedings** [...]. Bulgaria: Bulgarian Academy of Sciences, 2003. p. 490-494.

Domain ontologies enrichment: a systematic literature review

Abstract: This article presents a systematic review of the literature on the topic of domain ontologies enrichment for synthesize the main discussions. To this end, we use guidelines for systematic review, together with the use of Start software to manage and organize the 29 studies selected in five databases. The data collected allowed us to identify some features of published studies on ontology enrichment, such as sources of information used, types of enrichment most applied, types of information extraction techniques used, as well as the degree of detail of the methods for domain ontologies enrichment. We conclude that this investigation achieved its objective and that domain ontologies need to follow the dynamics of the phenomenon it represents, with the enrichment process being a direction for this.

Keywords: domain ontology; ontology update; ontology enrichment; systematic literature review

Recebido: 13/10/2022

Aceito: 28/03/2023

Declaração de autoria

Concepção e elaboração do estudo: Webert Júnio Araújo, Gercina Ângela de Lima.

Coleta de dados: Webert Júnio Araújo, Gercina Ângela de Lima.

Análise e interpretação dos dados: Webert Júnio Araújo, Gercina Ângela de Lima.

Redação: Webert Júnio Araújo, Gercina Ângela de Lima.

Revisão crítica do manuscrito: Webert Júnio Araújo, Gercina Ângela de Lima.

Como citar

ARAÚJO, Webert Júnio; LIMA, Gercina Angela de. Enriquecimento de ontologias de domínio: uma revisão sistemática da literatura. **Em Questão**, Porto Alegre, v. 29, e-127825, 2023. DOI: <https://doi.org/10.1590/1808-5245.29.127825>



¹ O software Start (*State of the Art through Systematic Review*) foi desenvolvido por pesquisadores do Laboratório de Pesquisa em Engenharia de Software da Universidade Federal de São Carlos (UFSCAR). Esse software permite a gestão de todas as etapas da revisão sistemática da literatura, desde o planejamento, seleção de estudos, até a etapa de redação do relatório da revisão.

² Na proposta original de Kichenham (2004), são previstas cinco etapas na fase de Condução da revisão. Nesta tese, não foi aplicada a etapa de Avaliação da qualidade dos estudos, pois se chegou a um número razoável de estudos, não havendo assim a necessidade de filtragem dos mesmos.

³ Para aqueles estudos que continham títulos muito genéricos ou não eram suficientemente claros, realizou-se a leitura do resumo.