

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/327652216>

Avaliação e enriquecimento semântico de ontologia com base na comparação a um corpus: um estudo de caso no domínio da agricultura

Article in *Revista Ibero-Americana de Ciência da Informação* · December 2018

DOI: 10.26512/rici.v11.n3.2018.10433

CITATION

1

READS

118

2 authors:



Gercina Angela de Lima

Federal University of Minas Gerais

142 PUBLICATIONS 560 CITATIONS

[SEE PROFILE](#)



Webert Araújo

Federal University of Minas Gerais

12 PUBLICATIONS 4 CITATIONS

[SEE PROFILE](#)

Avaliação e enriquecimento semântico de ontologia com base na comparação a um *corpus*: um estudo de caso no domínio da agricultura

Webert Júnio Araújo

Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, MG, Brasil
araujowebert0@gmail.com

Gercina Ângela de Lima

Universidade Federal de Minas Gerais, Escola de Ciência da Informação, Belo Horizonte, MG,
Brasil
glima@eci.ufmg.br

DOI: <https://doi.org/10.26512/rici.v11.n3.2018.10433>

Recebido/Recibido/Received: 2018-04-02

Aceitado/Aceptado/Accepted: 2018-07-03

Resumo: Este estudo é resultado da dissertação de mestrado que apresenta uma pesquisa sobre avaliação de ontologia. O objetivo norteador foi avaliar uma ontologia de domínio por meio de uma proposta de avaliação baseada na comparação a um corpus de texto e a partir da avaliação propor o enriquecimento semântico. A ontologia que se propôs avaliar neste trabalho foi desenvolvida no âmbito da EMBRAPA Campinas, denomina-se OntoAgroHidro e representa o impacto das mudanças climáticas na agricultura e nos recursos hídricos. A metodologia da pesquisa se respaldou em uma proposta de avaliação de ontologias desenvolvida por Brewster *et al.* (2004), que foi adaptada para o contexto do estudo e possui quatro etapas: 1) definição das questões de competência; 2) seleção da coleção de documentos que compõe o corpus para avaliação da ontologia; 3) seleção das expressões relacionadas aos conceitos das questões de competência; 4) avaliação e inclusão das expressões na ontologia. Os resultados demonstraram que a ontologia avaliada apresentou um percentual de aproximadamente 70% dos conceitos do corpus selecionado para avaliação. Os resultados também indicaram particularidades na representação de conceitos pela OntoAgroHidro. Conclui que a avaliação de ontologia aplicada foi satisfatória, pois permitiu identificar vários pontos de melhoria na ontologia avaliada.

Palavras-chave: Avaliação de Ontologia. Enriquecimento Semântico. Ontologias. Organização do Conhecimento.

Ontology evaluation and terminological enrichment based on comparison to a corpus: a case study in the agricultural domain

Abstract: This study is the result of the dissertation which presents a research of ontology evaluation. The guiding objective was to evaluate an ontology domain through an evaluation proposal based on comparison to a corpus. The ontology we proposed to evaluate in this work was developed by the EMBRAPA Campinas, it is called OntoAgroHidro and represents the impact of climate changes on agriculture and water resources. The research methodology was endorsed in a proposal for ontologies evaluation developed by Brewster *et al.* (2004), which was adapted to the context of this study and has four steps: 1) definition of competency questions; 2) selection of the collection of documents that make up the corpus for the ontology evaluation; 3) selection of expressions related to the concepts of competency questions; 4) evaluation and inclusion of the terms in the ontology. The results showed that the ontology evaluated presents a percentage of approximately 70% of the text corpus concepts selected for evaluation. The results also indicate particularities in the representation of concepts by

OntoAgroHidro. It concluded that the applied ontology evaluation was satisfactory, because it was possible to identify several areas for improvement in the evaluated ontology.

Keywords: Knowledge Management. Ontology evaluation. Ontologies. Semantic enrichment.

Evaluación ontológica y enriquecimiento terminológico basado en la comparación con un corpus: un estudio de caso en el ámbito agrícola

Resumen: Este estudio es el resultado de la disertación que presenta una investigación de evaluación ontológica. El objetivo rector fue evaluar un dominio de ontología a través de una propuesta de evaluación basada en la comparación con un corpus. La ontología que proponemos evaluar en este trabajo fue desarrollada por EMBRAPA Campinas, se llama OntoAgroHidro y representa el impacto de los cambios climáticos en la agricultura y los recursos hídricos. La metodología de investigación fue respaldada en una propuesta de evaluación de ontologías desarrollada por Brewster et al. (2004), que se adaptó al contexto de este estudio y tiene cuatro pasos: 1) definición de preguntas de competencia; 2) selección de la colección de documentos que componen el corpus para la evaluación ontológica; 3) selección de expresiones relacionadas con los conceptos de preguntas de competencia; 4) evaluación e inclusión de los términos en la ontología. Los resultados mostraron que la ontología evaluada presenta un porcentaje de aproximadamente el 70% de los conceptos de corpus de texto seleccionados para la evaluación. Los resultados también indican particularidades en la representación de conceptos por OntoAgroHidro. Concluyó que la evaluación ontológica aplicada fue satisfactoria, ya que fue posible identificar varias áreas de mejora en la ontología evaluada.

Palabras clave: Enriquecimiento semántico. Evaluación Ontológica. Ontologías. Organización del conocimiento.

1 Introdução

A organização do conhecimento é relevante para vários domínios do saber, visto que, para facilitar a comunicação entre a comunidade científica de um determinado domínio, é necessário representá-lo para eliminar possíveis ambiguidades e torná-lo disponível de forma consensual, por meio de seus termos e relacionamentos. É nesse sentido que a Biblioteconomia e Ciência da Informação (BCI) tem-se destacado, por desenvolver e pesquisar técnicas de organização e representação, que são instrumentos aproveitados em várias áreas de estudo. Entre esses instrumentos, estão as ontologias.

Ontologias podem ser consideradas, entre outras abordagens, sistemas de organização do conhecimento que possibilitam a representação de coisas existentes no mundo real. Elas permitem representar determinada realidade de um domínio do conhecimento, que seja compartilhada pelos especialistas desse domínio, isto é, que revelem um consenso nas escolhas das categorias, termos e relações que representam esse domínio na ontologia. Em razão desse consenso é que as ontologias se mostram profícuas na representação e compartilhamento de estruturas do conhecimento de diferentes domínios.

Embora sejam ferramentas que possuem relevante potencial na Representação do Conhecimento (RC), as ontologias precisam ser bem modeladas para que se aproximem ao máximo da realidade que pretendem representar. Geralmente, isso acontece sem o rigor necessário e elas acabam não representando bem o domínio que se dispõe a representar, seja

por falta de conhecimento sobre modelagem da equipe de desenvolvimento, seja por falta de uso de uma metodologia pelos desenvolvedores. Assim, elas acabam sendo mais artesanais do que científicas (JONES, BENCH-CAPON, VISSER, 1998 *apud* ALMEIDA, 2009). Diante desse quadro, uma temática que se mostra bastante útil nesse contexto é a avaliação de ontologias.

As propostas já existentes para avaliação de ontologias propõem analisar o potencial da ontologia na representação da realidade, contribuindo com a eliminação de inconsistências. Buscam saber se a ontologia está representando bem o domínio a que se propõe representar e se os conceitos e as relações escolhidas para representar são os melhores. Dessa forma, podem contribuir para avaliar uma ontologia já existente, ou mesmo propor melhorias para sua modelagem.

Nesse contexto, o presente trabalho propõe avaliar uma ontologia de domínio desenvolvida pela Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Campinas, denominada OntoAgroHidro, que representa os conceitos relacionados às mudanças climáticas, aos recursos hídricos e à agricultura. O objetivo é avaliar o desempenho da ontologia na representação dos conceitos do domínio e enriquecer semanticamente a ontologia a partir da avaliação.

2 Avaliação de Ontologia

O estudo da avaliação de ontologias se deve à necessidade de verificar a sua eficiência na representação do conhecimento e na realização de outras tarefas para as quais foram desenvolvidas, uma vez que, com o desenvolvimento de ontologias em várias áreas do conhecimento e muitas vezes sem nenhum respaldo metodológico, logo surgiram ontologias ineficientes, que não representavam de maneira adequada o conhecimento a que se propunham representar.

Mesmo as metodologias e métodos para a construção de ontologias mais citadas na literatura técnica, como a TOVE (Toronto Virtual Enterprise Ontology Project), Methontology, NeON, Método 101, Método Sensus, não apresentam um consenso com relação às etapas para construção de ontologias, apresentando diferentes etapas sendo direcionadas a diferentes propósitos e aplicações. Portanto, após leituras realizadas sobre esse tópico, constatou-se que ainda falta um modelo/padrão unificado para construção de ontologias, embora existam iniciativas de metodologias para desenvolvimento de ontologias (MENDONÇA, 2015). Então, com tal deficiência, a temática avaliação de ontologias vem como uma solução, pois tem o propósito de analisar a ontologia com o intuito de identificar seu desempenho na representação do conhecimento.

Vrandecic (2010) aponta a importância de se avaliar ontologias, dada a relevância dessas no contexto da web semântica. De acordo com o autor, erros e omissões em ontologias podem levar a aplicações que não aproveitam todo o potencial das ontologias. Ademais, boas ontologias garantem o reuso do conhecimento e melhora a cooperação entre especialistas de domínios.

Gómez-Pérez (2001) apresenta cinco critérios pelos quais uma ontologia pode ser avaliada. São eles: (1) coerência, que está relacionado a saber se é possível obter conclusões contraditórias a partir de definições explícitas na ontologia; (2) completude, que visa identificar se a ontologia não possui definições incompletas; (3) concisão, que contempla evitar que haja definições desnecessárias e inúteis; (4) capacidade de expansão, que se refere ao esforço exigido em adicionar novas definições ou mais conhecimento para as suas definições, sem alterar o conjunto de propriedades que já está definido; e (5) sensibilidade, que se refere a como pequenas mudanças na definição podem alterar o conjunto de propriedades que já está bem definido. Esse autor deixa claro o objetivo da avaliação de ontologias ao afirmar que:

a avaliação de ontologias refere-se à construção correta do conteúdo da ontologia, isto é, garantir que a sua definição corresponda corretamente aos requisitos da ontologia e às questões de competência ou refletir corretamente o mundo real. O objetivo é comprovar a conformidade do modelo de mundo com o mundo modelado formalmente. A avaliação de ontologia inclui: cada definição e axioma individual; coleção de definições e axiomas que são declarados explicitamente na ontologia; definições que são importadas de outras ontologias; definições que podem ser inferidas utilizando outras definições e axiomas (GOMEZ-PEREZ, 2001, p. 393, tradução nossa¹).

A literatura mostra que não há um consenso com relação à classificação das diversas propostas para avaliação de ontologias. Pesquisadores como Porzel e Malaka (2004) classificam as propostas como aquelas de cunho quantitativo e qualitativo; outros, como Sabou *et al.* (2006) preferem classificá-las de acordo com o método de avaliação, automático ou manual. Almeida (2009) compila o trabalho de Brank, Grobelnik e Mladenić (2006) que apontaram outro tipo de classificação, que vai além das classificações dicotômicas, separando as propostas de avaliação de ontologias em quatro classes. São elas:

¹The evaluation of ontologies refers to the correct building of the content of the ontology, that is, ensuring that its definitions (a definition is written in natural language and in a formal language) correctly implement ontology requirements and competency questions or perform correctly in the real world. The goal is to prove compliance of the world model (if it exists and is known) with the world modeled formally. Ontology evaluation includes: each individual definition and axiom; collection of definitions and axioms that are stated explicitly in the ontology; definitions that are imported from other ontologies; definitions that can be inferred using other definitions and axioms (GOMEZ-PEREZ, 2001, p. 393).

- Proposta de avaliação com base na comparação da ontologia a um “modelo padrão” ou *Gold Standard*. Ex.: Maedche e Staab (2002), Brank, Grobelnik e Mladení c (2006);
- Proposta de avaliação com base no uso da ontologia em uma aplicação/tarefa e avaliação dos resultados. Ex.: Fernández, Cantador e Castells (2006), Clarke et al (2013);
- Proposta de avaliação que envolve comparações com fontes de dados não estruturados sobre um domínio que será coberto pela ontologia. Ex.: Brewster et al (2004), Hlomani e Stacey (2014);
- Proposta de avaliação em que um especialista (humano) tenta estimar o quão bem a ontologia responde a um conjunto predefinido de critérios, padrões, requerimentos. Ex.: Lozano-Tello e Gómez-Pérez (2004), Almeida (2009).

Além da classificação das diferentes propostas de avaliação de ontologias, há autores que dividem cada avaliação em níveis. Os principais são: o léxico, que mede a qualidade das palavras; o taxonômico, que considera as conexões hierárquicas entre os conceitos que utilizam a relação ‘é-um’; o semântico, que aborda as outras relações entre os conceitos; o sintático, que considera os requisitos sintáticos da linguagem formal usada para descrever a ontologia; o contextual ou de aplicação, que considera o contexto da ontologia; o estrutural, que leva em consideração os princípios e critérios envolvidos na construção da ontologia (FERNÁNDEZ, CANTADOR, CASTELLS, 2006). Um dos argumentos para essa divisão em níveis é a diminuição da complexidade na avaliação, visto que avaliar cada nível da ontologia separadamente pode ser mais simples do que avaliar a ontologia por completo.

Desde o surgimento do tema, várias propostas de avaliação de ontologias foram desenvolvidas, principalmente por pesquisadores ligados à Ciência da Computação. A maioria dessas propostas foram construídas com base em ontologias específicas e seguindo os objetivos dessas ontologias. Ademais, algumas propostas são apenas modelos teóricos, ou seja, não foram realizados testes empíricos para validação das propostas. Dessa forma, o que se identifica é uma falta de padrão metodológico para avaliar ontologias, visto que as propostas existentes estão dispersas e não possuem um modelo comum no qual se fundamentar.

Diante das várias propostas para avaliação de ontologias existentes na literatura, esta pesquisa se baseou em uma proposta desenvolvida por Brewster *et al.* (2004), que pode ser classificada como aquelas que envolvem comparações com fontes de dados não estruturados sobre um domínio que será coberto pela ontologia.

Brewster *et al.* (2004) propõem uma abordagem para avaliação de ontologias por meio da comparação da ontologia com uma coleção de textos de determinado domínio. O objetivo é medir a proximidade entre a ontologia e o domínio que ela se propõe a representar. Os

autores justificam que a literatura se tem mostrado a fonte de dados mais efetiva para a construção de várias ontologias. Eles dizem ainda que esse tipo de avaliação pode ser feito por meio da extração de termos em dada coleção de textos e simplesmente contagem dos termos que aparecem na ontologia e na coleção. Porém, em seus estudos, eles propõem o uso de uma arquitetura mais sofisticada que se baseia em três passos: 1) identificar palavras-chave/termos na coleção de textos; 2) expandir a consulta, ou seja, procurar termos semanticamente próximos, como os sinônimos, por exemplo (os autores utilizam o Word Net² para fazer isso); 3) fazer o mapeamento dos termos identificados na ontologia que está sendo avaliada. Por fim, ao comparar a coleção com a ontologia, contam-se quantos termos na ontologia estão entre os termos da coleção.

3 Metodologia

O universo desta pesquisa é a ontologia da Rede AgroHidro da EMBRAPA, denominada OntoAgroHidro (OAH). A OntoAgroHidro foi desenvolvida por especialistas do domínio dentro de um projeto da Rede Agro-Hidro da EMBRAPA, denominado “Impactos da agricultura e das mudanças climáticas nos recursos hídricos: diagnose e propostas de adaptação e mitigação em bacias hidrográficas nos Biomas brasileiros” com o propósito de auxiliar na realização do projeto. O processo de construção da OAH se baseou em etapas de algumas metodologias existentes, porém, não foi feito o uso sistemático de nenhuma metodologia específica. O propósito da OAH é a representação do conhecimento sobre os impactos das mudanças climáticas e da agricultura nos recursos hídricos.

As etapas para realização da pesquisa se baseiam em um trabalho desenvolvido por Brewster *et al.* (2004). Porém, para sua utilização nesta pesquisa, foi necessário fazer adaptações, visto que a proposta de avaliação dos autores não apresenta detalhamentos para a sua execução.

O motivo da escolha da proposta de Brewster *et al.* (2004) se deve ao fato de ser uma proposta com etapas de simples aplicação e que atende aos propósitos da presente pesquisa. Além disso, a aplicação de uma proposta baseada na comparação a uma fonte de dados permite a avaliação do nível léxico, vocabulário, conceitos e dados, o que já é suficiente para o escopo desta pesquisa.

² Word Net é uma base de dados léxica para o Inglês (<http://wordnetweb.princeton.edu/perl/webwn>)

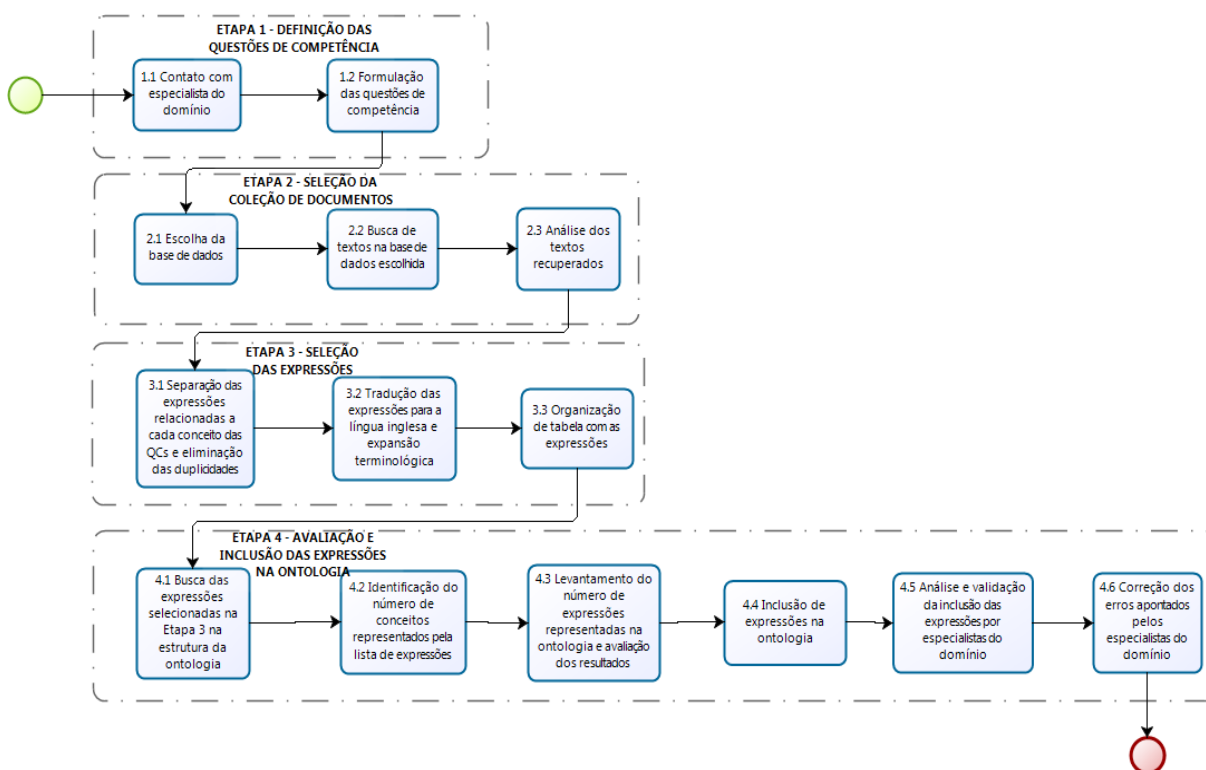
3.1 Percurso metodológico

As etapas realizadas para a avaliação da OntoAgroHidro dividem-se em quatro; são elas:

- Etapa 1 – Definição das questões de competência;
- Etapa 2 – Seleção da coleção de documentos que compõe o corpus de texto para avaliação da ontologia;
- Etapa 3 – Seleção das expressões relacionadas aos conceitos das questões de competência;
- Etapa 4 – Avaliação e inclusão das expressões na OntoAgroHidro.

Na Figura 1 apresenta-se uma visão geral das etapas de avaliação e cada uma de suas subetapas. O detalhamento de cada uma das etapas encontra-se na seção 4 – Resultados: apresentação e discussão – no qual também são apresentados os resultados obtidos em cada etapa, ao mesmo tempo em que são descritos os procedimentos metodológicos.

Figura 1: Visão geral das etapas de avaliação da ontologia



Fonte: Elaborada pelos autores.

Legenda: Circulo verde representa o inicio das etapas e o circulo vermelho o término das etapas

4 Resultados: apresentação e discussão

Nesta seção, apresentam-se, detalhadamente, os procedimentos utilizados para a avaliação da ontologia OntoAgroHidro e os resultados alcançados. Essa forma de apresentar os procedimentos metodológicos e os resultados tem como objetivo facilitar o entendimento do emprego da metodologia na pesquisa.

4.1 Etapa 1 – Definição das questões de competência

A realização desta etapa utilizou os seguintes procedimentos: (1) contato com um especialista do domínio; (2) formulação das questões de competência.

Três questões de competência foram formuladas pelo especialista do domínio para a avaliação da OntoAgroHidro. Como as questões tratam de assuntos relacionados ao domínio representado pela ontologia, era esperado que a ontologia tivesse em sua estrutura o conhecimento necessário para responder a essas questões.

As três questões formuladas pelo especialista foram:

Questão 1) Os cultivos orgânicos ocasionam algum impacto na qualidade da água de nossas bacias?

Questão 2) Faltarão água nos reservatórios de abastecimento de água para a população paulista em 2016?

Questão 3) Como o desmatamento de nascentes afeta a quantidade e a qualidade de água de nossos rios?

A análise das três questões de competência formuladas pelo especialista resultou em 11 conceitos. São eles: 1) Cultivo orgânico; 2) Qualidade de água; 3) Bacia; 4) Água; 5) Reservatório de abastecimento de água; 6) População; 7) Tempo (referente ao ano 2016); 8) Desmatamento; 9) Nascente; 10) Quantidade de água; 11) Rio.

4.2 Etapa 2 – Seleção da coleção de documentos que compõe o corpus de texto para avaliação da ontologia

A realização desta etapa resultou nos seguintes procedimentos: (1) escolha da base de dados; (2) busca de textos na base de dados escolhida; (3) análise dos textos recuperados.

(1) Escolha da base de dados:

Como recomendação do especialista do domínio, escolheu-se a Base de Dados da Pesquisa Agropecuária (BDPA³), que é composta dos acervos das bibliotecas da EMBRAPA, constituída por documentos tais como artigos científicos, boletins, relatórios, livros, teses, trabalhos apresentados em eventos, entre outros tipos de documentos.

(2) Busca de textos na base de dados escolhida:

³ Disponível em: <<https://www.bdpa.cnptia.embrapa.br/consulta/>>. Acesso em: 02 fev. 2016.

Utilizando os conceitos principais presentes em cada questão de competência, realizaram-se as buscas na base de dados BDPA. O resultado dessa pesquisa, discriminado por questão de competência, está detalhado na Tabela 1 – Resultados da busca dos documentos na base BDPA.

Tabela 1: Resultados da busca dos documentos na base BDPA

Questão	Conceito	Expr. de busca	Quant. de reg. recup.	Obs.
Questão 1	Qualidade de água	"qualidade de água"	282	
	Cultivo orgânico	"cultivo orgânico"	926	
	Bacia	Bacia	1.951	Busca no campo: assunto.
Questão 2	Água	Água	12.334	Busca no campo: assunto.
	Reservatório de abastecimento de água	reservatório de abastecimento de água	49	Não foram utilizadas aspas, pois a revocação foi muito baixa.
	* ⁴ População (população paulista)	#	0	Não foram extraídos termos da BDPA, apenas dos tesouros.
	*Tempo (2016)	#	0	Não foram extraídos termos da BDPA, apenas dos tesouros
Questão 3	Desmatamento	desmatamento	1.307	
	Nascente	Nascente	19	Busca no campo "conteúdo"
	Quantidade de água	"quantidade de água"	284	
	Rio	Rio	9.844	Busca no campo: assunto

Fonte: Elaborada pelos autores.

Os registros recuperados foram analisados aplicando-se os seguintes critérios:

- i. Consideração apenas dos registros que apresentavam o documento em texto completo no formato PDF (*PortableDocumentFormat*);
- ii. Realização de leitura de um documento por vez, iniciando no primeiro registro recuperado que possuía o texto completo no formato PDF⁵;
- iii. Realizada a leitura de um documento, foram identificadas todas as expressões, que podem ser consideradas variações semânticas e/ou conceitos próximos que tivessem relação com o conceito da questão de competência buscado;
- iv. Efetuada a seleção das expressões de um documento, era feita a análise de outro documento em seguida. Procedeu-se uma análise contínua até que não aparecessem novas expressões relacionadas ao conceito da questão de competência. Assim, decidia-se que havia chegado ao limite de expressões relacionadas a determinado conceito da questão de competência. Dessa forma, repetia-se o processo com outro conceito da questão de competência;

⁴ Para os conceitos de 'população' e 'tempo' não foram buscados documentos na BDPA, pois são muito genéricos e, com base em um teste na BDPA, verificou-se que os documentos recuperados não eram relevantes; assim, as expressões relacionadas aos dois conceitos foram buscadas nos tesouros AGROVOC (Agricultura e Vocabulário) e NAL (National Agricultural Library). De onde foram extraídos os descritores relacionados aos dois conceitos.

⁵ Por se tratar de um estudo exploratório, esse processo de foi feito de forma manual, ou seja, leitura de cada documento selecionado. Porém, acreditamos que a automatização desse processo pode facilitar essa etapa.

A Tabela 2 mostra o número de documentos analisados detalhado por cada conceito das questões de competência. Foi analisado um total de 59 documentos.

Tabela 2: Quantidade de documentos analisados

Questão	Conceito	Quant. de docs. analisados
Questão 1	Qualidade de água	10
	Cultivo orgânico	8
	Bacia	6
Questão 2	Água	5
	Reserv. de abast. de água	7
	⁶ População (população paulista)	0
	Tempo (2016)	0
Questão 3	Desmatamento	7
	Nascente	5
	Quantidade de água	6
	Rio	5

Fonte: Elaborada pelos autores.

4.3 Etapa 3 – Seleção das expressões relacionadas aos conceitos das questões de competência

A realização desta etapa resultou nos seguintes procedimentos: (1) separação das expressões relacionadas a cada conceito das questões de competência e eliminação das duplicidades; (2) tradução das expressões para a língua inglesa e expansão terminológica; (3) organização de uma tabela com as expressões.

- (1) Separação das expressões relacionadas a cada conceito das questões de competência e eliminação das duplicidades;

Nesta fase, foram separadas as expressões relacionadas aos conceitos das questões de competência. Para cada conceito das questões de competência, listaram-se as expressões extraídas da coleção de documentos. Cabe frisar que a separação das expressões categorizadas de acordo com cada conceito da questão de competência é importante, pois permitirá avaliar de forma mais específica quais são os conceitos que a ontologia representa e quais são os conceitos que precisam ser representados.

⁶Para os conceitos de ‘população’ e ‘tempo’ não foram buscados documentos na BDPA, pois são muito genéricos e, com base em um teste na BDPA, verificou-se que os documentos recuperados não eram relevantes. Assim, as expressões relacionadas aos dois conceitos foram buscadas nos tesouros AGROVOC e NAL. De onde foram extraídos os descritores relacionados aos dois conceitos.

Em seguida, realizou-se uma análise nas expressões extraídas da coleção de documentos para eliminação das expressões repetidas que estavam relacionadas a um mesmo conceito das questões de competência.

(2) Tradução das expressões para a língua inglesa e expansão terminológica;

Traduziram-se as expressões para suas correspondentes na língua inglesa⁷, utilizando o tradutor *online* Google Tradutor e os tesouros AGROVOC⁸ e NAL⁹. Além disso, para suprir ausências de expressões relevantes, que não foram levantadas na análise à coleção de documentos, utilizaram-se os tesouros AGROVOC e NAL com o intuito de complementar a representação dos conceitos presentes nas questões de competência (expansão terminológica). Para isso, as expressões em inglês¹⁰ foram buscadas nas caixas de busca dos tesouros e selecionaram-se somente os descritores que possuíam alguma relação com os conceitos das questões de competência.

(3) Organização de uma tabela com as expressões:

O conjunto de expressões extraídas da coleção de documentos e os descritores extraídos dos tesouros AGROVOC e NAL chegaram-se a 130 no total. Essa foi a amostra total de expressões que representam os 11 conceitos das questões de competência e conceitos relacionados do domínio de conhecimento.

4.4 Etapa 4 – Avaliação e inclusão das expressões na OntoAgroHidro

Esta etapa tratou-se da avaliação da ontologia e da inclusão de novas expressões em sua estrutura. Para isso, seguiram-se os procedimentos: (1) busca das expressões selecionadas na Etapa 3 na estrutura da ontologia; (2) identificação do número de conceitos representados pela lista de expressões; (3) levantamento do número de expressões representadas na ontologia e avaliação dos resultados; (4) inclusão de expressões na ontologia; (5) análise e validação da inclusão das expressões por especialistas do domínio; (6) correção dos erros apontados pelos especialistas do domínio.

(1) Busca das expressões selecionadas na Etapa 3 na estrutura da ontologia:

Utilizou-se o código da OntoAgroHidro em OWL (Web Ontology Language) e com o uso do recurso “Ctrl + F” buscou-se uma expressão de cada vez na estrutura da ontologia. Além

⁷Foi necessário fazer essa tradução porque a OntoAgroHidro é bilíngue.

⁸Disponível em: <http://aims.fao.org/standards/agrovoc/functionalities/search>

⁹Disponível em: <http://agclass.nal.usda.gov/dne/search.shtml>

¹⁰Foi necessário buscar as expressões na língua inglesa, pois o tesouro NAL não possui descritores na língua portuguesa e o AGROVOC, embora seja multilíngue, possui descritores no português europeu, que apresenta algumas variações com relação ao português brasileiro.

disso, navegou-se na estrutura da OntoAgroHidro, por meio do *Software Protégé*, com o propósito de identificar quais das 130 expressões estavam representadas pelos conceitos da ontologia.

(2) Identificação do número de conceitos representados pela lista de expressões:

Como várias expressões podem representar um único conceito, foi necessário identificar na lista de expressões quantos conceitos ela representa. Assim, foi possível comparar o número de conceitos representados pela lista de expressões com a quantidade desses conceitos que são representados pela OntoAgroHidro.

A identificação do número de conceitos representado pela lista de expressões foi feita de forma intelectual. Constatou-se que a lista de expressões representa 82 conceitos no total.

(3) Levantamento do número de expressões representadas na ontologia e avaliação dos resultados:

Nesta fase, realizou-se o levantamento da quantidade de conceitos da lista de expressões representados na estrutura da ontologia. Os resultados encontrados foram detalhados de acordo com cada conceito das questões de competência. Cabe ressaltar que o objetivo foi identificar se a ontologia representava o sentido (significado) de cada expressão, ou seja, considerou-se que a ontologia poderia não possuir a expressão exata em sua estrutura, mas que algum dos seus conceitos poderia estar representando aquele mesmo sentido (significado).

Depois do levantamento, fez-se a avaliação com base no número de conceitos identificados na lista de expressões (82 conceitos) em comparação com a quantidade desses conceitos representados pela ontologia. Esse parâmetro de avaliação se respalda em Brewster *et al.* (2004), quando afirmam que todo conceito na estrutura de uma ontologia é uma representação reduzida de várias possibilidades lexicais (no caso desta pesquisa, são as expressões extraídas dos documentos analisados e dos tesouros).

O resultado da comparação pode ser visto na Tabela 3, na seguinte forma: a primeira coluna apresenta a qual questão de competência se refere; a segunda coluna mostra o número de expressões identificadas na coleção de textos e nos tesouros; a terceira coluna apresenta o número de conceitos representados pelas expressões da segunda coluna; a quarta coluna mostra a quantidade de conceitos representados na OntoAgroHidro; e, por fim, a quinta e última coluna apresenta a representatividade da ontologia em porcentagem.

Tabela 3: Representatividade da OntoAgroHidro por questão de competência

Questões	Nº de expressões	Nº de conceitos repres. pelas expressões ¹¹	Nº de conceitos repres. na OntoAgroHidro	Representatividade da OntoAgroHidro
Questão 1	66	50	38	76%
Questão 2	40	25	14	56%
Questão 3	24	7	3	42,85%
Total	130	82	55	67,07%

Fonte: Elaborada pelos autores.

Percebeu-se, com base nos dados apresentados na Tabela 3, que a representatividade da questão 1 (Os cultivos orgânicos ocasionam algum impacto na qualidade de água de nossas bacias?) é a maior (76%) em comparação com as questões de competência 2 e 3. Isso está associado ao conceito de “qualidade de água”, uma vez que a OntoAgroHidro representa grande parte das expressões ligadas a esse conceito, que são, em sua maioria, parâmetros físicos, químicos e biológicos para medir a qualidade de água.

A Tabela 4 aborda com mais detalhes a representatividade da OntoAgroHidro por cada conceito oriundo das questões de competência. Isso possibilita um conhecimento mais específico de quais são os conceitos que não estão bem representados na OntoAgroHidro e aponta para a necessidade de melhorias na representação para que se possa ter um melhor modelo do domínio.

Tabela 4: Representatividade da OntoAgroHidro discriminada por cada conceito das questões de competência

Questão	Conceito buscado	Nº de expr.	N.º de conceitos repres. pelas expr.	Conceitos repres. na OntoAgroHidro	Representatividade da OntoAgroHidro
Questão 1	Qualidade de água	47	47	37	78,72%
	Cultivo orgânico	13	1	0	0%
	Bacia	6	2	1	50%
Questão 2	Água	18	15	11	73,33%
	Reservatório de abast. de água	14	6	2	33,33%
	População	5	3	0	0%
	Tempo	3	1	1	100%
	Desmatamento	10	1	0	0%

¹¹O número de conceitos se baseia em uma análise de quantos conceitos estão representados na lista de expressões.

Questão 3	Nascente	5	1	0	0%
	Quantidade de água	4	1	1	100%
	Rio	5	4	2	50%

Fonte: Elaborada pelos autores.

Em uma abordagem analítica da Tabela 4, realizou-se a separação dos conceitos relacionados a “recursos hídricos”, tais como: qualidade de água, bacia, água, reservatório de abastecimento de água, nascente, quantidade de água e rio. Avaliando a representatividade da OntoAgroHidro em relação aos conceitos mencionados, constatou-se que a ontologia representa 71,05% dos conceitos (ou seja, 54 conceitos de 76). A ideia de apresentar esse dado remete ao objetivo principal para a criação da OntoAgroHidro. A saber, o objetivo de criação da ontologia é a representação do conhecimento sobre os impactos das **mudanças climáticas** e da **agricultura** nos **recursos hídricos**. Dessa forma, pode-se afirmar, que para ser eficiente em seu propósito, a ontologia precisa representar principalmente esses três subdomínios (mudanças climáticas, agricultura e recursos hídricos) de forma bem específica. Como todas as três questões de competência abordavam algum conceito relacionado à ‘água’, no que concerne ao subdomínio ‘recursos hídricos’, essa pesquisa pôde avaliar o desempenho da ontologia na representação.

Também, pode-se notar que a Tabela 4 demonstra que os conceitos: bacia, reservatório de abastecimento de água e rio obtiveram uma baixa representatividade na ontologia, se comparados com o índice de representatividade de conceitos tais como: água, qualidade de água e outros.

(4) Inclusão de expressões na ontologia:

Para inclusão de expressões na OntoAgroHidro, analisaram-se as expressões que não estavam representadas na estrutura da ontologia, e com base no uso dos tesouros AGROVOC e NAL, e outras fontes de informação do domínio da agricultura¹² (como glossários, artigos, etc.), incluíram-se na estrutura da ontologia, as expressões consideradas relevantes para o domínio modelado pela ontologia. Para isso, utilizou-se o editor de ontologias Protégé, em que foi necessário analisar toda a estrutura de classes da ontologia e decidir em qual classe, subclasse ou propriedade determinada expressão deveria ser incluída, tornando-se essa expressão, um conceito na ontologia.

Durante o processo de inclusão de expressões na OntoAgroHidro, foram criadas algumas propriedades que não existiam na ontologia, com o propósito de expandir o vocabulário relacionado à expressão adicionada na ontologia. Para tanto, com base em um

¹²Essas fontes de informação serviram de guia para a inclusão das expressões da OntoAgroHidro.

trabalho desenvolvido por Ramalho (2015), criaram-se as subpropriedades: *altLabel* (Rótulo Alternativo); *prefLabel* (Rótulo preferido); *hiddenLabel* (Rótulo oculto).

Em seguida, inseriram-se informações adicionais a respeito das expressões, tais como sinônimos, expressões correspondentes na língua inglesa e definições¹³. Algumas definições foram retiradas da base lexical Onto.PT¹⁴. Isso contribuiu para que a ontologia ficasse com uma representatividade mais rica, do ponto de vista terminológico.

Ao final, foi elaborado um relatório compilando as expressões incluídas na ontologia com os detalhes sobre a localização da expressão na estrutura da ontologia e quais anotações foram atribuídas a essa expressão. Posteriormente, esse relatório foi enviado a dois especialistas do domínio para análise e validação. Além disso, foi enviado também o arquivo da OntoAgroHidro em OWL com as alterações realizadas em sua estrutura. Foram incluídas 21 novas expressões na ontologia, assim como seus respectivos rótulos (rótulo preferido, rótulo alternativo, definição).

(5) Análise e validação da inclusão das expressões por especialistas do domínio:

Após a inclusão das expressões na estrutura da OntoAgroHidro, dois especialistas do domínio analisaram e validaram a inclusão de cada expressão na estrutura da ontologia, com base no relatório enviado e na própria ontologia, por meio do software Protégé.

A análise dos especialistas envolveu: a classe, subclasse ou propriedade em que a expressão foi incluída, os rótulos (*labels*) adicionados, ou seja, se aqueles rótulos são os melhores, as instâncias (*individuals*) e as definições (*comments*).

(6) Correção dos erros apontados pelos especialistas do domínio:

Após os especialistas do domínio apresentarem um parecer com relação às expressões incluídas na ontologia, todas as correções indicadas foram realizadas. As principais correções apontadas pelos especialistas do domínio referiam-se às expressões que demandavam um conhecimento mais aprofundado do domínio.

4.5 Discussão

A proposta de avaliação da OntoAgroHidro apresentada neste trabalho perpassou por quatro etapas, subdivididas em 14 procedimentos, como pode ser observado na Figura. 1 – Visão geral das etapas de avaliação da ontologia – da seção 3, que apresenta a proposta de modo esquematizado.

¹³ As definições foram retiradas da base lexical Onto.PT, dos tesouros AGROVOC e NAL e da Web (quando não eram encontrados nas ferramentas: Onto.PT, AGROVOC e NAL). Cabe ressaltar que todas as definições foram validadas pelo especialista do domínio.

¹⁴ Onto.PT é uma base lexical para a língua portuguesa estruturada de forma semelhante à base lexical em inglês, WordNet. A Onto.PT está disponível em: <http://ontopt.dei.uc.pt>

Os resultados demonstraram que a ontologia avaliada apresenta um percentual de aproximadamente 70% dos conceitos do corpus de textos selecionado para a avaliação. Isso mostrou a necessidade de melhoria da representatividade de alguns conceitos pela ontologia. Os resultados também indicaram particularidades na representação de conceitos pela OntoAgroHidro, indicando que conceitos relacionados ao domínio de recursos hídricos possuíam uma melhor representatividade.

Fundamentado nos resultados desta pesquisa, acredita-se que o problema que impediu que a ontologia tivesse uma melhor representatividade dos conceitos da amostra selecionada para essa pesquisa foi a ausência de uma modelagem mais específica do domínio. Alguns conceitos do domínio não estavam representados pela OntoAgroHidro, pois faltou especificação, ou seja, existia o conceito mais abrangente, mas o conceito mais específico não estava representado na estrutura da ontologia. Essas falhas na representação identificadas na OntoAgroHidro podem ser solucionadas com a inclusão de novos conceitos na ontologia, que foi exatamente o que se começou a fazer nesta pesquisa. Ademais, o reuso de outros sistemas de organização do conhecimento, como tesouros, também pode contribuir no enriquecimento semântico da OntoAgroHidro e eliminar essa falta de conceitos específicos do domínio.

5 Considerações finais

Esta pesquisa partiu do problema da existência de ontologias ineficientes em vários domínios do conhecimento, que demandam a realização de avaliações para verificar a adequabilidade da ontologia na representação do conhecimento. Assim, realizou-se a aplicação de uma proposta de avaliação em uma ontologia de domínio com o intuito de verificar o desempenho da ontologia na representação do conhecimento do domínio.

Os procedimentos metodológicos adotados foram adequados para os objetivos desta pesquisa. A revisão de literatura permitiu identificar várias propostas de avaliação existentes. Constatou-se que na literatura ainda falta um método consolidado para avaliação de ontologias e a maioria das propostas segue objetivos bem particulares, em que são realizados testes próprios para uma ontologia específica de determinado domínio. Porém, vale ressaltar que a temática de avaliação de ontologias ainda é um campo emergente; portanto, é preciso reconhecer que as propostas existentes vêm contribuindo para o avanço do campo nos últimos anos.

Cabe informar que durante o desenvolvimento da pesquisa, algumas dificuldades foram encontradas. A principal delas foi a falta de uma metodologia consolidada para avaliação de ontologias. Outra dificuldade foi com relação à falta de estudos sobre avaliação de ontologias desenvolvidas para ontologias no idioma português; a maioria dos trabalhos

trata de ontologias no idioma inglês; sendo assim, toda a avaliação é direcionada para esse idioma, o que dificulta a existência de ferramentas que abordem as particularidades do idioma português.

Entre as limitações de estudo, cita-se o fato das etapas dois (seleção da coleção de documentos que compõe o corpus de texto para avaliação da ontologia) e três (seleção das expressões relacionadas aos conceitos das questões de competência) da metodologia ter sido realizada de forma manual, que é um processo demorado, subjetivo e inconsistente, visto que cada analisador humano pode interpretar de forma diferente. Dessa forma, acredita-se que a realização dessas etapas de forma automática com o uso de alguma técnica de Processamento de Linguagem Natural possa ser uma boa alternativa.

Em conclusão, espera-se que este estudo tenha contribuído para uma melhoria na representação do conhecimento dos domínios mencionados e para o avanço das pesquisas em Ciência da Informação e que possa ter contribuído especificamente com o desenvolvimento do campo de avaliação de ontologias.

Referências

ALMEIDA, M.B. A proposal to evaluate ontology content. **Applied Ontology**, Birmingham, v. 4, n. 3-4, p. 245-265, 2009.

BRANK, J.; GROBELNIK, M.; MLADENIC, D. A survey of ontology evaluation techniques. In: CONFERENCE ON DATA MINING AND DATA WAREHOUSES, 6., 2006, Ljubljana, Slovenia, 2006, [Proceedings....]. Ljubljana, Slovenia: [s.n], 2006. Disponível em: <http://ailab.ijs.si/dunja/sikdd2005/papers/BrankEvaluationSiKDD2005.pdf> Acesso em: 8 ago. 2017.

BREWSTER, C. *et al.* Data driven ontology evaluation. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 2004, Lisboa. **Proceedings...** Lisboa: [s.n], 2004. Disponível em: <http://oro.open.ac.uk/20045/> Acesso em: 8 ago. 2017.

CLARKE, E. L. *et al.* A task-based approach for Gene Ontology evaluation. **Journal of Biomedical Semantics**, v. 4, 2013. Disponível em: <http://jbiomedsem.biomedcentral.com/articles/10.1186/2041-1480-4-S1-S4> Acesso em: 10 jan. 2017.

FERNÁNDEZ LÓPEZ, M. Overview of methodologies for building ontologies. In: WORKSHOP ONTOLOGIES AND PROBLEM-SOLVING METHODS: LESSONS LEARNED AND FUTURE TRENDS DE LA CONFERENCIA INTERNACIONAL JOINT CONFERENCE FOR ARTIFICIAL INTELLIGENCE (IJCAI'99). 1999, Stockholm, Suécia. **Proceedings...** Stockholm, Suécia: [s.n], 1999. Disponível em: http://oa.upm.es/5480/1/Overview_Of_Methodologies.pdf Acesso em: 8 ago. 2017.

FERNÁNDEZ, M.; CANTADOR, Iván; CASTELLS, Pablo. CORE: A tool for collaborative ontology reuse and evaluation. In: INTERNATIONAL WORKSHOP ON EVALUATION OF ONTOLOGIES FOR THE WEB, 4., 2006, Edinburgh, UK. **Proceedings...** Edinburgh, UK: [s.n], 2006. Disponível em: <http://oro.open.ac.uk/28592/1/eon06.pdf> Acesso em: 8 ago. 2017.

GÓMEZ-PÉREZ, A. Evaluation of ontologies. **International Journal of Intelligent Systems**, Alberta, Canadá, v. 16, n. 3, 2001. p. 391-409. Disponível em: [http://onlinelibrary.wiley.com/doi/10.1002/1098-111X\(200103\)16:3%3C391::AID-INT1014%3E3.0.CO;2-2/abstract](http://onlinelibrary.wiley.com/doi/10.1002/1098-111X(200103)16:3%3C391::AID-INT1014%3E3.0.CO;2-2/abstract) Acesso em: 8 ago. 2017.

HLOMANI, Hlomani; STACEY, D. A. An Extension to the Data-driven Ontology Evaluation. In: INTERNATIONAL CONFERENCE ON INFORMATION REUSE AND INTEGRATION, August 13-15, San Francisco. **Proceedings...** San Francisco: [s.n.], 2014.

JONES, D.; BENCH-CAPON, T.; VISSER, P. Methodologies for ontology development. [s.l.]: [s.n.], 1998. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.52.2437&rep=rep1&type=pdf> Acesso em: 8 ago. 2016.

LOZANO-TELLO, A.; GÓMEZ-PÉREZ, A. Ontometric: A method to choose the appropriate ontology. **Journal of Database Management**, 2004. Disponível em: http://oa.upm.es/6467/1/ONTOMETRIC_A_Method.pdf Acesso em: 10 set. 2016.

MAEDCHE, A.; STAAB, S. Measuring similarity between ontologies. In: EUROPEAN CONFERENCE ON KNOWLEDGE ACQUISITION AND MANAGEMENT, Madrid, Espanha. **Proceedings...** Madrid, Espanha: [s.n.], 2002. Disponível em: <http://disi.unitn.it/~accord/RelatedWork/Matching/Staab2002.pdf> Acesso em: 20 set. 2016.

MENDONÇA, Fabrício Martins. **Ontoforinfoscience**: metodologia para construção de ontologias pelos cientistas da informação: uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano. 2015. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, 2015.

PORZEL, R.; MALAKA, R. A task-based approach for ontology evaluation. In: EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE AND WORKSHOP ON ONTOLOGY LEARNING AND POPULATION IN. 16., Valencia, Espanha. **Proceedings...** Valencia, Espanha: [s.n.], 2004. Disponível em: <http://olp.dfki.de/ecai04/final-porzels.pdf> Acesso em: 8 ago. 2017.

RAMALHO, Rogério Aparecido Sá. Análise do modelo de dados SKOS: sistema de organização do conhecimento simples para a web, **Informação e Tecnologia**, Marília; João Pessoa, v. 2, n.1, p. 66-79, jan./jul., 2015. Disponível em: <http://periodicos.ufpb.br/ojs/index.php/itec/article/view/25995> Acesso em: 8 ago. 2017.

SABOU, M. *et al.* Ontology selection: Ontology evaluation on the real semantic web. In: INTERNATIONAL WORKSHOP ON EVALUATION OF ONTOLOGIES FOR THE WEB AT THE 15TH INTERNATIONAL WORLD WIDE WEB, 4., 2006, Edinburgh, UK. **Proceedings...** Edinburgh, UK: Knowledge Media Institute, 2006. Disponível em: http://ai.ia.agh.edu.pl/wiki/_media/pl:miw:2009:eon2006.pdf Acesso em: 8 ago. 2016.

USCHOLD, Mike; GRUNINGER, Michael. Ontologies: Principles, methods and applications. **Knowledge Engineering Review**, [S.l.], v. 11, n. 2, p. 93-136, 1996. Disponível em: <http://dx.doi.org/10.1017/S0269888900007797> Acesso em: 08 ago. 2016.

VRANDECIC, Denny. **Ontology evaluation**. 2010. Thesis. – Instituto de Tecnologia de Karlsruhe (KIT), Karlsruhe, 2010.