

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

WEBERT JÚNIO ARAÚJO

AVALIAÇÃO DE ONTOLOGIA COM BASE NA
COMPARAÇÃO A UM CORPUS:
um estudo da OntoAgroHidro da EMBRAPA

Belo Horizonte

2016

WEBERT JÚNIO ARAÚJO

AVALIAÇÃO DE ONTOLOGIA COM BASE NA
COMPARAÇÃO A UM CORPUS:
um estudo da OntoAgroHidro da EMBRAPA

Documento apresentado à banca examinadora do Programa de Pós-Graduação da Escola de Ciência da Informação da Universidade Federal de Minas Gerais – UFMG, como requisito normativo para título de Mestre.

Linha de Pesquisa: Organização e Uso da Informação.

Orientadora: Gercina Ângela Borém de Oliveira Lima

Belo Horizonte

2016

A663a

Araújo, Webert Júnio.

Avaliação de ontologia com base na comparação a um corpus: um estudo da OntoAgroHidro da EMBRAPA [manuscrito] / Webert Júnio Araújo. – 2016.

136 f. : enc., il.

Orientador: Gercina Ângela Borém de Oliveira Lima.
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Referências: f. 89-100.

Apêndices: f. 101-136.

1. Ciência da informação – Teses. 2. Ontologias (Recuperação da Informação) – Teses. 3. Sistemas de organização – Teses. 4. Sistemas de recuperação da informação. I. Título. II. Lima, Gercina Ângela Borém de Oliveira. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 025.4.03

Ficha catalográfica: Biblioteca Profª Etelvina Lima, Escola de Ciência da Informação da UFMG.



UFMG

Universidade Federal de Minas Gerais
Escola de Ciência da Informação
Programa de Pós-Graduação em Ciência da Informação

FOLHA DE APROVAÇÃO

"AVALIAÇÃO DE ONTOLOGIA COM BASE NA COMPARAÇÃO A UM CORPUS: UM ESTUDO DA ONTOAGROHIDRO DA EMBRAPA"

Webert Junio Araujo

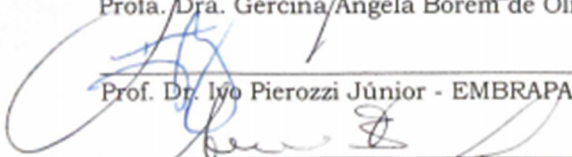
Dissertação submetida à Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais, como parte dos requisitos à obtenção do título de "**mestre em Ciência da Informação**", linha de pesquisa "**Organização e Uso da Informação**".

Dissertação aprovada em: 01 de junho de 2016.

Por:



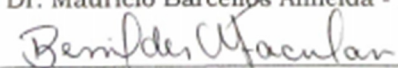
Prof. Dra. Gercina/Ângela Borém de Oliveira Lima - ECI/UFMG (Orientadora)



Prof. Dr. Ivo Pierozzi Júnior - EMBRAPA

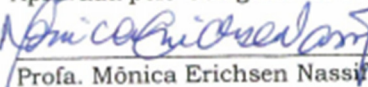


Prof. Dr. Mauricio Barcellos Almeida - ECI/UFMG



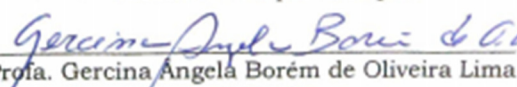
Profa. Dra. Benildes Coura Moreira dos Santos Maculan - ECI/UFMG

Aprovada pelo Colegiado do PPGCI



Profa. Mônica Erichsen Nassif
Decano do Colegiado do
Colegiado do PPGCI

Versão final Aprovada por



Prof. Gercina/Ângela Borém de Oliveira Lima
Orientadora



UFMG

Universidade Federal de Minas Gerais
Escola de Ciência da Informação
Programa de Pós-Graduação em Ciência da Informação

ATA DA DEFESA DE DISSERTAÇÃO DE **WEBERT JUNIO ARAUJO**, matrícula:
2014660438


Às 10:00 horas do dia 01 de junho de 2016, reuniu-se na Escola de Ciência da Informação da UFMG a Comissão Examinadora aprovada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação em 05/05/2016, para julgar, em exame final, o trabalho intitulado *Avaliação de ontologia com base na comparação a um corpus: um estudo da OntoAgroHidro da Embrapa*, requisito final para obtenção do Grau de MESTRE em CIÊNCIA DA INFORMAÇÃO, área de concentração: Produção, Organização e Utilização da Informação, Linha de Pesquisa: Organização e Uso da Informação. Abrindo a sessão, a Presidente da Comissão, Profa. Dra. Gercina Ângela Borém de Oliveira Lima, após dar conhecimento aos presentes do teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Logo após, a Comissão se reuniu sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:


Profa. Dra. Gercina Ângela Borém de Oliveira Lima - Orientadora	APROVADO
Prof. Dr. Ivo Pierozzi Júnior	APROVADO
Prof. Dr. Mauricio Barcellos Almeida	APROVADO
Profa. Dra. Benildes Coura Moreira dos Santos Maculan	APROVADO

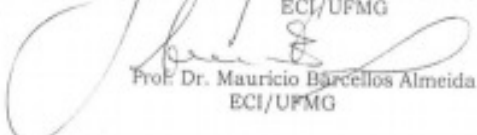
Pelas indicações, o candidato foi considerado APROVADO.


O resultado final foi comunicado publicamente ao candidato pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a sessão, da qual foi lavrada a presente ATA que será assinada por todos os membros participantes da Comissão Examinadora.

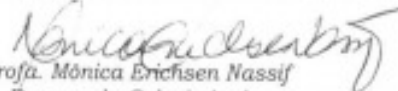
Belo Horizonte, 01 de junho de 2016


Profa. Dra. Gercina Ângela Borém de Oliveira Lima
ECI/UFMG


Prof. Dr. Ivo Pierozzi Júnior
EMBRAPA


Prof. Dr. Mauricio Barcellos Almeida
ECI/UFMG


Profa. Dra. Benildes Coura Moreira dos Santos Maculan
ECI/UFMG


Profa. Mônica Erichsen Nassif
Decano do Colegiado do
Programa de Pós-Graduação em
Ciência da Informação/UFMG

Obs: Este documento não terá validade sem a assinatura e carimbo da Coordenadora.

DEDICO

A Deus, que torna tudo isto realizável.

Aos meus pais, que sempre acreditaram na importância dos estudos e que não mediram esforços para que eu pudesse ter uma boa educação.

À minha esposa, pelo carinho, por acreditar nos meus sonhos e motivar-me sempre.

À minha Tia Glau (*in memoriam*), que sempre me encorajou a perseguir os meus sonhos, não importando com o tamanho dos obstáculos. Obrigado por me ensinar, ainda que em outras palavras, que “quem é dono de uma vontade firme modela o mundo ao seu feitio.”

AGRADECIMENTOS

A Deus, que me dá forças para vencer todos os obstáculos e seguir em frente. Obrigado por tudo que tens me concedido, Senhor!

Aos meus amados pais, Rosa e Antônio, que não tiveram acesso à educação, mas sempre acreditaram no conhecimento e na educação como ferramentas de transformação. Obrigado por todos os valores que me ensinaram. Obrigado por cada gota de suor que derramaram para que eu pudesse chegar até aqui.

À minha amada esposa, Andresa, pela paciência, carinho, críticas no momento certo e por estar sempre ao meu lado em todos os momentos. Obrigado por todo incentivo, *Young Lady*. Te amo!

Aos meus irmãos, Kel, Tim e Taka, que sempre se orgulharam de mim e que me motivam a continuar em busca dos meus sonhos.

À minha orientadora Gercina Lima, que acreditou no meu potencial e me guiou nos caminhos complexos da pesquisa. Obrigado, Professora, pela forma como você acredita nos seus orientandos e os motiva a fazer sempre o melhor!

Ao Ivo Pierozzi Júnior, pela paciência e disponibilidade para as várias conversas por videoconferência.

Ao Caio, que é uma fonte de inspiração para que eu acredite na realização de todos os meus sonhos.

A todos os meus familiares, por acreditarem sempre no meu potencial e por todas as mensagens de incentivo.

À equipe do LTI (Christiano, Lucas, Luiz, Wagner e Walisson), por ter-me acolhido desde a graduação. Obrigado pelas experiências, conversas e amizade.

Aos professores da ECI, nos quais me inspirei para buscar o conhecimento sempre.

Ao professor Maurício, que plantou o desafio da pesquisa em ontologias em mim na disciplina “Introdução às ontologias”.

À professora Benildes, exemplo de pesquisadora, e que muito me auxiliou durante o período do mestrado.

Aos companheiros do Grupo MHTX, pelas prévias, discussões e todas as experiências que vivemos juntos.

Ao Júnio Lopes por todo seu empenho em me ajudar na normalização do trabalho.

Aos funcionários da Escola de Ciência da Informação, que tratam os alunos como parte da família. Obrigado pelas conversas nos corredores, pelos desabafos, pelas conversas boas sobre futebol, que nos fazem esquecermo-nos um pouco da pesquisa e relaxar. Em especial a: Cláudia, Eliedir, Gizele, Lucimary e Willer.

À turma do 1.º período do curso de Biblioteconomia 2014/2, pela experiência fantástica no Estágio Docente. Obrigado por ter-me acolhido.

À banca de defesa, por ter aceitado meu convite de prontidão.

A todas as pessoas que torceram por mim, fica aqui minha gratidão!

*O coração do homem traça o seu caminho,
mas o Senhor lhe dirige os passos.*

Provérbios 16-9

*Se tiveres conhecimento, deixa que as outras
pessoas acendam as suas velas na tua.*

Thomas Fuller

RESUMO

Esta pesquisa teve como objetivo norteador avaliar uma ontologia de domínio por meio de uma proposta de avaliação baseada na comparação a um corpus de texto. 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 escolha desse tema se baseou nos problemas relacionados à existência de ontologias ineficientes, ou seja, que não representam de maneira adequada o domínio que se propõem a representar. Assim, acredita-se que esses problemas possam ser minimizados com a temática avaliação de ontologias, que propõe verificar a adequabilidade da ontologia na representação do conhecimento de determinado domínio. A pesquisa caracteriza-se como um estudo de caso e envolveu procedimentos de natureza exploratória, qualitativa e aplicada, iniciando-se com a exploração acerca dos temas relevantes para a pesquisa, com levantamento bibliográfico na literatura da área de Ciência da Informação e áreas adjacentes. O fundamento teórico-metodológico perpassou conceitos sobre temas tais como organização e representação do conhecimento, definições de conceito e termo, ontologias e questões de competência. A revisão de literatura abordou a temática avaliação de ontologias. 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 de texto 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 de textos selecionado para avaliação. Isso mostra 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. Concluiu-se que a avaliação de ontologia aplicada foi satisfatória, pois permitiu identificar vários pontos de melhoria na ontologia a-avaliada.

Palavras-chave: Avaliação de ontologias. Ontologia de domínio. Sistema de Organização do Conhecimento. Organização e Representação do Conhecimento.

ABSTRACT

The guiding aim of this research was to evaluate a domain ontology through an evaluation proposal based on comparison to a text corpus. The ontology proposed to evaluate in this study was carried out within the EMBRAPA Campinas, it is called OntoAgroHidro and represents the impact of climate changes on agriculture and water resources. The selection of this theme is based on problems related to the existence of inefficient ontologies, ie, that do not represent properly the domain that propose to represent. Therefore, it is believed that the exposed problems can be minimized with the thematic ontology evaluation, which proposes to verify the suitability of the ontology in the representation of particular domain of knowledge. The research is characterized as exploratory, qualitative and applied, starting with the exploration of the relevant topics to the research in the Information Science and Computer Science literatures. The theoretical and methodological foundation permeated concepts on topics such as organization and knowledge representation, definitions of term and concept, ontologies and competency questions. The literature review covered the thematic ontology evaluation. The research methodology was based on a proposal for a data-driven ontology evaluation developed by Brewster et al (2004). The proposal was adapted and has four steps: 1) definition of competency questions; 2) selection of the collection of documents that compose the corpus for the ontology evaluation; 3) selection of expressions related to the concepts of the competency questions; 4) evaluation and inclusion of expressions in the ontology. The results demonstrated that the evaluated ontology presented a rate of approximately 70% of the text corpus concepts selected for the evaluation. This shows the need to improve the representation of some concepts in the ontology. The results also indicate particularities in the representation of concepts by OntoAgroHidro, indicating that concepts related to water resources domain had a better representativeness. We concluded that the applied ontology evaluation was satisfactory, as identified several areas for improvement in the evaluated ontology.

Keywords: Ontology evaluation. Domain ontology. Knowledge Organization System. Knowledge Organization and Representation.

LISTA DE ILUSTRAÇÕES

FIGURA 1 – Árvore de Porfírio.....	26
FIGURA 2 – Tipos de Sistema de Organização do Conhecimento	30
FIGURA 3 – Triângulo do Conceito de Dahlberg	32
FIGURA 4 – Visão geral da OntoAgroHidro e suas classes e subclasses principais	46
FIGURA 5 – Visão geral das etapas de avaliação da ontologia.....	66
FIGURA 6 – Recorte da estrutura de classes em que o conceito cultivo orgânico deveria estar representado.....	79
FIGURA 7 – Recorte da estrutura de classes em que o conceito população deveria estar representado.	80
FIGURA 8 – Recorte da estrutura de classes em que o conceito desmatamento deveria estar representado.....	80
FIGURA 9 – Recorte da estrutura de classes em que o conceito nascente deveria estar representado.	81
QUADRO 1 As categorias de Aristóteles e seus respectivos exemplos	25
QUADRO 2 Sistemas de Organização do Conhecimento Tradicionais	29
QUADRO 3 Visão global das propostas de avaliação.....	58
QUADRO 4 Adaptações da proposta de avaliação de Brewster et al (2004) para a presente pesquisa	63

LISTA DE TABELAS

TABELA 1 Resultados da busca dos documentos na base BDPA.....	70
TABELA 2 Quantidade de documentos analisados.....	72
TABELA 3 Representatividade da OntoAgroHidro por questão de competência.....	77
TABELA 4 Representatividade da OntoAgroHidro discriminada por cada conceito das questões de competência.....	78

LISTA DE ABREVIATURAS E SIGLAS

ADEC	American Distance Education Consortium
AGNIC	Agricultural Network Information Center
AGROVOC	Agricultura e Vocabulário
AIMS	Agricultural Information Management Standards
ARIST	Annual Review of Information Science and Technology
ARS	Agricultural Research Service
BCI	Biblioteconomia e Ciência da Informação
BDPA	Base de Dados da Pesquisa Agropecuária
BRAPCI	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação
CC	Ciência da Computação
CI	Ciência da Informação
CORE	Collaborative Ontology Reuse and Evaluation
CUAHSI	Consortium of Universities for the Advancement of Hydrologic Science
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
FAO	Food and Agriculture Organization of United Nations
FSRIO	Food Safety Research Information
GO	Gene Ontology
IA	Inteligência Artificial
IICA	Inter-American Institute for Cooperation on Agriculture
KOS	Knowledge Organization Systems
LOD	Linked Open Data
MHTX	Mapa Hipertextual
NAL	National Agricultural Library
NASA	National Aeronautics and Space Administration

OANC	American National Corpus
OC	Organização do Conhecimento
OWL	Web Ontology Language
QCS	Questões de Competência
RC	Representação do Conhecimento
RI	Recuperação da Informação
SKOS	Simple Knowledge Organization System
SKOS-XL	Simple Knowledge Organization System Extension for Labels
SMO	Small Molecule Ontology
SOC	Sistemas de Organização do Conhecimento
SWEET	Semantic Web for Earth and Environmental Terminology
TOVE	Toronto Virtual Enterprise Ontology Project
UFMG	Universidade Federal de Minas Gerais
USDA	United States Department of Agriculture

Sumário

1. INTRODUÇÃO	15
1.1 Justificativa e problema	17
1.2 Objetivos	20
1.2.1 Objetivo geral	20
1.2.2 Objetivos específicos	20
1.3 Estrutura da dissertação	21
2. REFERENCIAL TEÓRICO E METODOLÓGICO	23
2.1 Conceitos básicos relacionados à Representação do Conhecimento	23
2.2 Representação do Conhecimento no contexto da Biblioteconomia e Ciência da Informação	24
2.3 Definições de conceito e termo utilizadas neste trabalho.....	31
2.3.1 O Conceito	31
2.3.2 O Termo (expressão verbal).....	33
2.4 Tesouros AGROVOC e NAL	34
2.5 Ontologias.....	36
2.5.1 Definições	37
2.5.2 Características, tipos e aplicações das ontologias	39
2.5.3 Questões de competência em ontologias	41
2.5.4 OntoAgroHidro	43
3 AVALIAÇÃO DE ONTOLOGIAS: UMA BREVE REVISÃO	47
3.1 Avaliações com base na comparação a um modelo padrão.....	49
3.2 Avaliações com base no uso da ontologia em uma aplicação/tarefa	51
3.3 Avaliações com base na comparação a uma fonte de dados.....	53
3.4 Avaliações feitas por humanos	55
3.5 Considerações do capítulo	57
4 METODOLOGIA	61
4.1 Caracterização da pesquisa	61
4.2 Método	62
4.3 Percurso metodológico	64
5 DESCRIÇÃO DOS PROCEDIMENTOS E RESULTADOS	67

5.1 Etapa 1 – Definição das questões de competência.....	67
5.2 Etapa 2 – Seleção da coleção de documentos que compõe o corpus de texto para avaliação da ontologia.....	68
5.3 Etapa 3 – Seleção das expressões relacionadas aos conceitos das questões de competência	73
5.4 Etapa 4 – Avaliação e inclusão das expressões na OntoAgroHidro.....	74
5.5 Considerações do capítulo	83
6 CONSIDERAÇÕES FINAIS	85
6.1 Limitações da pesquisa	87
6.2 Trabalhos futuros.....	88
REFERÊNCIAS	89
APÊNDICES.....	101
APÊNDICE A – Lista de documentos analisados para extração das expressões	101
APÊNDICE B – Expressões (em português e inglês) extraídas da coleção de documentos e dos tesouros	110
APÊNDICE C – Representação das expressões pela OntoAgroHidro.....	116
APÊNDICE D – Relatório das expressões incluídas na OntoAgroHidro	123

1. INTRODUÇÃO

Ao longo de sua existência, o homem constantemente se preocupou em conhecer e compreender as coisas ao seu redor. Com isso, obteve grande progresso durante este caminhar em busca do desconhecido, o que pode ser comprovado pelo legado deixado pelos filósofos e cientistas em diferentes épocas. Até hoje, há uma persistência na busca pelo inexplorado – o conhecimento sempre foi objeto de desejo da humanidade e sempre esteve presente na vida do homem. Dos mais simples, com base principalmente no senso comum, aos filosóficos e científicos, com uma noção mais crítica e baseada em métodos. A evidente relevância do conhecimento para a humanidade justifica seu estudo e investigação para que se possa aprofundar mais em questões de seu interesse e contribuir com a descoberta do novo.

Assim como é importante adquirir conhecimento, a sua organização e representação sempre foram úteis para os seres humanos. É por isso que, com o passar do tempo, o homem foi desenvolvendo técnicas de representação e organização de seus conhecimentos para facilitar o seu compartilhamento e a comunicação, que, doravante, auxiliam em sua compreensão e acesso. Assim, vêm sendo estudadas e desenvolvidas desde as classificações das ciências, elaboradas por filósofos, as representações arborescentes, até a criação dos sistemas de classificação bibliográficas, que buscam uma ordenação lógica para os saberes.

Muitas das ferramentas de representação estudadas no passado servem de base para as atuais e há a necessidade de evolução dessas ferramentas. Essa necessidade se deve a vários fatores, entre eles, o contexto atual balizado pelas tecnologias e que exige cada vez mais o acesso ao conhecimento de forma rápida, além do desenvolvimento da ciência, que exige novas estruturas de representação.

Vários domínios do saber se preocupam em organizar o conhecimento, 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.

As ontologias são estudadas em mais de uma área do conhecimento. Na literatura, encontram-se estudos de ontologia na Ciência da Computação (CC), Filosofia, Inteligência Artificial (IA), Biblioteconomia e Ciência da Informação (BCI), entre outros. No contexto da Biblioteconomia e Ciência da Informação, os estudos foram iniciados por volta dos anos 1990 e têm ganhado destaque nas pesquisas da área por possuir grande potencial na representação e por contribuir com a recuperação da informação, permitindo inferências por sistemas inteligentes.

Ontologias podem ser consideradas, entre outras abordagens, sistemas de organização do conhecimento que possibilitam a representação de coisas existentes no mundo real. Elas permitem que haja uma representação única de determinada realidade de um dado 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 dispõem a representar, seja por falta de conhecimento sobre modelagem da equipe de desenvolvimento, seja por falta de critérios, padrões e metodologia. Pesquisadores do tema, Jones, Bench-Capon, Visser (1998); Fernández (1999); Uschold e Gruninger (1996), confirmam que o grande problema é a falta de um padrão metodológico que forneça orientações para o desenvolvimento de ontologias. Assim,

elas acabam sendo mais artesanais do que científicas, como afirmam Jones, Bench-Capon e Visser (1998). 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.

Encontram-se na literatura várias propostas para avaliação de ontologias. Porém, grande parte delas possui um viés mais computacional, ou seja, são automáticas, utilizando um agente de *software* que analisa a ontologia com base em critérios configurados no *software*. Poucas propostas são realizadas a partir da análise humana. A proposta desta pesquisa envolve uma abordagem de avaliação que abrange a máquina, mas que é complementada pela análise de um humano, ou seja, é uma abordagem semiautomática.

1.1 Justificativa e problema

As ontologias criam um padrão, facilitando o intercâmbio de informações entre pessoas e agentes de *software*. Além disso, elas podem contribuir com a recuperação da informação, permitindo a atribuição de semântica nas consultas realizadas pelos usuários, o que pode melhorar o índice de relevância dos resultados em comparação com as clássicas buscas por palavras-chave.

Como já dito, há várias propostas de construção de ontologias em diversos campos do conhecimento, como na medicina, engenharia de *software*, bibliotecas digitais. Porém, ainda não há uma metodologia geral unificada para a construção delas. As

metodologias existentes encontram-se dispersas e individualizadas. Isso acaba abrindo margem para o desenvolvimento de ontologias ineficientes. Assim, para verificar a adequabilidade da ontologia na representação do conhecimento de determinado domínio, faz-se necessário avaliá-la. Por isso teve início a demanda de métodos de avaliação de ontologias, visto que muitas são disponibilizadas para uso e reuso do conhecimento sem passar por qualquer tipo de avaliação.

A avaliação de ontologia é uma temática relevante para os estudos da área, pois pode mensurar sua eficiência na representação do conhecimento de um dado domínio. Além disso, a avaliação contribui para a identificação de possíveis falhas na modelagem da ontologia, melhorando a representação do conhecimento do domínio modelado, necessidade que pode ser comprovada de acordo com Correa:

Todo processo de modelagem é apenas uma representação dependente do ponto de vista do desenvolvedor, portanto normalmente é um processo falho, com isso, uma avaliação de um processo de modelagem é necessária para o refinamento de um processo adequado. (CORREA, 2009, p. 48).

Dessa forma, a presente pesquisa se justifica por buscar contribuir com os estudos sobre ontologias, com foco em uma abordagem de avaliação com base na comparação da ontologia a um corpus¹ de textos. Além disso, levam-se em consideração os conhecimentos dos especialistas do domínio nessa avaliação.

Ademais, o presente estudo se insere no quadro de pesquisas da BCI, pois objetiva avaliar um sistema de organização do conhecimento, que é uma ferramenta importante no contexto da área, dado o potencial que possui na representação e recuperação da informação e na organização do conhecimento, temáticas essas que estão intrinsecamente ligadas ao Fazer da CI². Como afirma Vickery (1986), a CI está voltada para o entendimento das manifestações da informação e suas formas de representação e organização. Além disso, são também preocupações da CI os

¹ No contexto desta pesquisa, considera-se 'corpus' como um coleção de documentos concernentes a uma área específica de conhecimento.

² De acordo com o Dicionário de Biblioteconomia e Arquivologia, o Fazer da CI está ligado à investigação das propriedades e o comportamento da informação, as forças que governam o fluxo da informação e os meios de seu processamento para uma ótima acessibilidade e usabilidade.

problemas relacionados ao processo de comunicação do conhecimento entre pessoas nos contextos social, institucional e individual, como diz Saracevic (1996). Uma das características das ontologias é justamente o compartilhamento do vocabulário, para facilitar a comunicação entre a comunidade de dado domínio e eliminar a possibilidade de ruídos na comunicação.

No presente trabalho, propõe-se avaliar uma ontologia que representa os recursos hídricos e a agricultura, domínios socialmente importantes. Portanto, é fundamental facilitar a comunicação e o compartilhamento e auxiliar no processo de recuperação da informação nesses domínios.

A abordagem utilizada neste trabalho se justifica porque é fundamentada na literatura sobre avaliação de ontologias. Além disso, pressupõe uma análise dos especialistas do domínio, que são as pessoas com o conhecimento técnico e experiência em determinado campo. Portanto, são fontes confiáveis quando se trata de avaliação de ontologias, uma vez que estarão lidando com conhecimentos específicos de sua área.

O ambiente da presente pesquisa é a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA Campinas), que, por meio de um Convênio firmado com a Universidade Federal de Minas Gerais (UFMG) e intermediado pelo Grupo de Pesquisa MHTX (Mapa Hipertextual – MHTX), visa à aplicação prática das pesquisas desenvolvidas no âmbito da Universidade ao contexto da Empresa.

A EMBRAPA criou a RedeAgroHidro, que tem o objetivo de entender como as mudanças climáticas e a agricultura afetam ou têm potencial para impactar os recursos hídricos nos diversos biomas brasileiros e os seus efeitos na sustentabilidade ambiental e econômica das comunidades rurais. Nessa Rede, são desenvolvidos vários projetos, sendo um deles o projeto intitulado “Impactos da agricultura e das mudanças climáticas nos recursos hídricos: diagnose e propostas de adaptação e mitigação em bacias hidrográficas nos biomas brasileiros”. Esse projeto tem como objetivo a análise dos impactos das mudanças climáticas e da agricultura nos recursos hídricos, produzindo conhecimentos e estratégias para um uso sustentável do solo e

da água, possibilitando a conservação dos recursos hídricos e a competitividade da agricultura em diferentes biomas brasileiros. Como forma de divulgar e compartilhar as informações dentro do projeto, foi criada a ontologia “OntoAgroHidro” com vistas a representar todo o universo de conceitos relacionados ao contexto do projeto e, assim, contribuir com o compartilhamento e consenso de todos os termos representados no domínio em questão (REDE AGROHIDRO, 2015).

Assim, as questões que norteiam o presente estudo são: (1) Como avaliar uma ontologia com base na comparação a um corpus de texto? (2) A OntoAgroHidro representa os conceitos principais presentes nas questões de competência?

Diante desses questionamentos, apresentam-se os objetivos descritos a seguir.

1.2 Objetivos

Esta dissertação visa contribuir com a melhoria da ontologia em questão. A seguir são apresentados os objetivos que nortearam a pesquisa.

1.2.1 Objetivo geral

Avaliar a OntoAgroHidro com base em uma proposta de avaliação baseada na comparação a um corpus de texto.

1.2.2 Objetivos específicos

1. Avaliar o desempenho da OntoAgroHidro na representação de conceitos

do domínio modelado com base em questões de competência.

2. Sugerir mudanças na OntoAgroHidro para melhorar a modelagem do domínio, a partir dos resultados da avaliação.

3. Contribuir para o enriquecimento conceitual e terminológico da ontologia.

1.3 Estrutura da dissertação

Esta dissertação está estruturada em seis capítulos, divididos da seguinte forma:

O capítulo 1, introdução, apresenta as motivações para o estudo, o problema de pesquisa e as justificativas para a sua realização, e por fim, descrevem-se os objetivos geral e específicos, que sintetizam o propósito desta pesquisa.

O capítulo 2, referencial teórico e metodológico, discorre sobre as bases teóricas que fundamentam a pesquisa e os insumos metodológicos, perpassando por temas como representação e organização do conhecimento, conceito e termo, tesouros, ontologias e questão de competência no âmbito dos estudos sobre ontologias.

O capítulo 3 traz a revisão de literatura, que aborda os trabalhos já publicados sobre avaliação de ontologias, mostrando diversas propostas para avaliar ontologias, desde propostas mais simples a outras sofisticadas e que exigem mais recursos para sua execução.

O capítulo 4 caracteriza a metodologia da pesquisa, e indica os métodos e técnicas utilizados, as estratégias de busca para a pesquisa bibliográfica e apresenta de forma geral as etapas percorridas para o desenvolvimento do estudo.

O capítulo 5 trata do detalhamento dos procedimentos metodológicos adotados para avaliação da ontologia OntoAgroHidro. Descrevem-se cada uma das etapas para

avaliação da ontologia, assim como os procedimentos de cada etapa. Apresentam-se, também, os resultados empíricos da aplicação de cada procedimento e analisam-se os resultados obtidos.

Por fim, o capítulo 6 apresenta as considerações finais, onde pondera-se sobre os objetivos deste estudo, os resultados alcançados, as contribuições e as principais dificuldades encontradas no percurso desta pesquisa. Além disso, sugerem-se algumas possibilidades de estudos futuros.

2. REFERENCIAL TEÓRICO E METODOLÓGICO

Nesta seção, apresentam-se as bases teóricas que fundamentam a pesquisa com o propósito de contextualizar a temática na área de estudo. Está dividida da seguinte maneira: a subseção 2.1 aborda os conceitos básicos relacionados à Representação do Conhecimento (RC), ou seja, “representação” e “conhecimento”; a subseção 2.2 trata da RC no contexto da Biblioteconomia e Ciência da Informação (BCI), mostrando a sua relação com a Organização do Conhecimento (OC) e algumas definições de autores conceituados da área, assim como os Sistemas de Organização do Conhecimento (SOC) abordados pela Ciência da Informação (CI); na subseção 2.3, apresentam-se as definições de conceito e termo; a subseção 2.4 aborda os tesouros AGROVOC e NAL; por fim, a subseção 2.5 apresenta algumas informações referentes às ontologias, como definições, características, tipologias, questões de competência e a ontologia OntoAgroHidro.

2.1 Conceitos básicos relacionados à Representação do Conhecimento

Antes de partir para a definição de RC, é necessário desmembrar os termos e defini-los separadamente. Iniciando por “conhecimento” (do latim *cognoscere*, “ato de conhecer”), de acordo com o dicionário Aurélio (FERREIRA, 2010, p. 189), é o “ato ou efeito de conhecer; informação ou noção adquirida pelo estudo ou pela experiência”. Aranha e Martins (1993, p. 40) o definem como “o pensamento resultante do vínculo entre o sujeito que conhece e o objeto a ser conhecido.”

Peter Burke (2003), em uma proposta de definição de informação e conhecimento, diz que o conhecimento é algo que vai além da informação. Ele o define como aquilo que representa o que foi “cozido”, ou seja, sistematizado pelo pensamento. Em suas definições de conhecimento, Brascher e Café (2008) dizem que ele é o resultado da cognição e que é o conteúdo ideal da consciência humana.

Com relação ao conceito de “representação”, segundo o dicionário Aurélio (FERREIRA, 2010, p. 658), é “aquilo que a mente produz, o conteúdo concreto do que é apreendido pelos sentidos, a imaginação, a memória ou o pensamento”. A palavra “representar” tem sua origem na palavra latina *repraesentare*, cujos significados são “fazer presente”, “apresentar novamente”, e “trazer à memória” (SALES; CAFÉ, 2009, p. 100). Alvarenga (2003, p. 20) diz que “representar significa colocar algo no lugar de”.

É possível encontrar conceituações e citações de autores que colocam a RC no contexto da Inteligência Artificial (IA) e Ciência da Computação (CC), como Mypoulos (1981), que afirma que a RC é um problema central na IA, pois o desenvolvimento de sistemas inteligentes se baseia no conhecimento especializado e na facilidade de manipulação desse conhecimento. Assim, como é possível encontrar autores como Brascher e Café (2008), que afirmam que a RC está relacionada à BCI, mais especificamente à Organização do Conhecimento (OC).

No contexto deste trabalho, o foco está nos estudos de RC sob a ótica da BCI, em que são apresentados os modelos fundamentais que contribuíram para o desenvolvimento dos sistemas de organização atuais; entre eles, as ontologias. O objetivo de abordar os temas Organização e Representação do Conhecimento é contextualizar a relação das ontologias com essas temáticas que já vêm sendo estudadas no contexto da BCI há mais tempo.

2.2 Representação do Conhecimento no contexto da Biblioteconomia e Ciência da Informação

Antes de contextualizar a RC na BCI, aborda-se a evolução dos primeiros sistemas de organização do conhecimento até chegar aos sistemas atuais. De acordo com Aranalde (2009, p. 90), Aristóteles é o responsável pelo “primeiro trabalho sistemático de elaboração e de abordagem de categorias para a organização do conhecimento”.

O filósofo elencou dez categorias (as categorias podem ser vistas no QUADRO 1) que tornavam possíveis a classificação das coisas existentes no mundo.

QUADRO 1
As categorias de Aristóteles e seus respectivos exemplos

Categoria	Exemplos
Substância	homem, cão, pedra
Qualidade	azul, virtual, bonito
Quantidade	grande, comprido, pouco
Relação	duplo, empregado, mais barulhento
Duração	ontem, 2001, de noite
Lugar	Brasil, ali, no quintal
Ação	escrevendo, falando, correndo
Sufrimento	cortado, vitorioso, apreendido
Maneira de ser	feliz, saudável, gelado
Posição	horizontal, abaixo, supra

Fonte: PIEDADE, 1983, p. 20.

Com a evolução do conhecimento, foram propostas diversas outras formas de categorizá-lo. Além das categorias elencadas por Aristóteles, surgiram trabalhos clássicos, como a “Árvore de Porfírio” (FIG. 1) e a “Árvore Baniana”. A primeira delas, baseada nas categorias de Aristóteles, apresenta a forma dicotômica de representação com relacionamentos hierárquicos. De acordo com Japiassu e Marcondes (2008), a Árvore de Porfírio é

uma representação sob a forma de uma árvore, destinada a ilustrar a subordinação dos conceitos, a partir do conceito mais geral, que é o de substância, até chegar ao conceito homem, o de menor extensão, mas o de maior compreensão (JAPIASSU; MARCONDES, 2008, p. 18)

Já a “Árvore Baniana” de Ranganathan é policotômica e permite relacionamentos multidimensionais. Nesse tipo de árvore, os ramos de um tronco se enxertam em

outros troncos, fazendo assim uma analogia aos assuntos em uma classificação que se conecta com outros. Essas formas de representação são apenas algumas das várias que foram desenvolvidas ao longo dos anos, e que contribuíram para a existência dos novos instrumentos de representação que temos hoje no contexto digital.

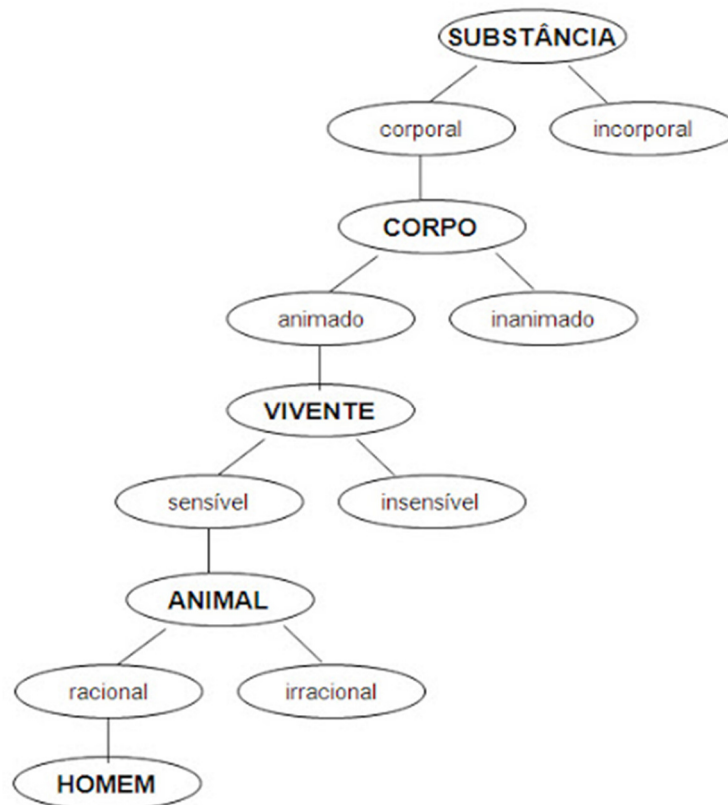


FIGURA 1 – Árvore de Porfírio
Fonte: Japiassu e Marcondes, 2008, p. 18.

No âmbito da BCI, percebe-se que a RC está associada à Organização do Conhecimento. Há pesquisadores, como Tristão et al (2004), que consideram as áreas similares, e há aqueles, como Dahlberg (2006) e Brascher e Café (2008), que defendem que são distintas.

Este trabalho baseia-se nas ideias dos autores que tratam a Representação do Conhecimento e a Organização do Conhecimento como distintas, mas que possuem

uma conexão estreita. Cabe ressaltar que há ainda na literatura definições de Organização da Informação e Organização do Conhecimento com o mesmo sentido. Porém, essa divergência conceitual não será abordada neste trabalho, por não ser relevante para o presente estudo.

Com base na linha conceitual aqui abordada, a Representação do Conhecimento tem sua origem dentro da Organização do Conhecimento. A RC possibilita a OC por meio de instrumentos de representação ou Sistemas de Organização do Conhecimento (SOC) — termo mais comum na área —, colocando as atividades de organização em prática. Assim, será abordado primeiramente o conceito de OC e depois o de RC.

De acordo com a literatura da área, a preocupação com a OC não é recente. Ao longo dos anos, pesquisadores como Charles A. Cutter, W. C. Berwick Sayers, Henry Bliss, Ernest Cushingon Richardson vêm propondo teorias e técnicas para a OC (HJORLAND, 2007). Com base em estudos similares a esses, os sistemas de organização do conhecimento bibliográficos e os alfabéticos, usados por bibliotecários como instrumentos de representação e organização do conhecimento, contribuíram para o desenvolvimento da OC.

Uma das definições de OC que possui mais destaque na literatura é a de Dahlberg (2006), que diz que a organização do conhecimento é a ciência que determina a estruturação e a sistematização dos conceitos, de acordo com suas características, que podem ser descritas como elementos de herança do objeto, e a aplicação dos conceitos e classes dos conceitos ordenados pela indicação de valores dos referentes conteúdos dos objetos ou assuntos.

As definições de Dahlberg (2006), Brascher e Café (2008) pontuam que a OC está relacionada à organização de conceitos, visto que a hierarquia entre os conceitos e as relações entre eles possibilitam a representação semântica de um domínio. Brascher e Café definem a organização do conhecimento como

o processo de modelagem do conhecimento que visa à construção de representações do conhecimento. Esse processo tem por base a análise do

conceito e de suas características para o estabelecimento da posição que cada conceito ocupa num determinado domínio, bem como das suas relações com os demais conceitos que compõem esse sistema nocional (BRASCHER E CAFÉ, 2008, p. 8).

Para Vickery (2008), organizar o conhecimento é agrupar aquilo que conhecemos em uma estrutura sistematicamente organizada. Com base no raciocínio abordado neste trabalho, considera-se então a OC como o processo de modelagem do conhecimento que visa à construção de suas representações.

Com relação ao conceito de “representação”, Alvarenga (2003, p. 20) aponta que representar significa o “ato de colocar algo no lugar de”. Dessa forma, toda representação é algo incompleto, visto que tem o objetivo de tirar a complexidade daquilo que se está representando; assim, pode-se dizer que não existe representação perfeita. A única “representação” perfeita seria o próprio artefato que se deseja representar.

Segundo Brascher e Café (2008), a representação do conhecimento é resultado de

um processo de análise de domínio e procura refletir uma visão consensual sobre a realidade que se pretende representar [...] é feita por meio de diferentes tipos de Sistemas de Organização do Conhecimento (SOC) (BRASCHER;CAFÉ, 2008, p. 6-8).

A RC é feita por meio dos instrumentos de representação, ou Sistemas de Organização do Conhecimento, que expressam a interpretação organizada e sistematizada do objeto a ser representado.

Os Sistemas de Organização do Conhecimento (SOC) foram evoluindo ao longo dos anos. Começando com os modelos criados pelos filósofos, que possuíam o objetivo de explicar o funcionamento das coisas no mundo e de classificar o conhecimento humano, chegando aos sistemas bibliográficos de organização do conhecimento, que surgiram para serem aplicados no arranjo de livros nas estantes. No início, esses sistemas bibliográficos eram criados sem notação, mas, com o crescente volume de livros, logo se viu a necessidade do desenvolvimento de sistemas bibliográficos de organização do conhecimento com notação. No quadro a seguir, apresentam-se

alguns dos principais Sistemas de Organização do Conhecimento que foram desenvolvidos. Faz-se uma separação entre os sistemas filosóficos, bibliográficos sem notação e bibliográficos com notação.

QUADRO 2
Sistemas de Organização do Conhecimento Tradicionais

TIPO	SOC	ANO	CLASSIFICACIONISTA
Filosófico	Classificação de Locke	1688	John Locke
	Classificação de Comte	1822-51	Auguste Comte
	Classificação de Spencer	1864	Herbert Spencer
Bibliográficos sem notação	Classificação de Aldus	1505	Aldo Manuzzi
	Classificação de Leibnitz	1718	Leibnitz
	Classificação de Smith	1882	L. P. Smith
Bibliográficos com notação	Classificação Decimal de Dewey	1876	Melvil Dewey
	Classificação da Biblioteca do Congresso	1902	Biblioteca do Congresso – EUA
	Classificação Decimal Universal	1905	FID

Fonte: Adaptado de Miranda, 2005, p. 123, 125 e 126.

No atual contexto, balizado pelas tecnologias de informação e comunicação, o termo Knowledge Organization Systems (KOS), segundo Hodge (2000), foi proposto em 1998 pelo Networked Knowledge Organization Systems Working Group, na primeira conferência da ACM Digital Libraries, em Pittsburgh, Pensilvânia, e abrange sistemas de classificação, cabeçalhos de assunto, arquivos de autoridade, taxonomias, tesouros, redes semânticas, ontologias, dentre outros.

De acordo com Carlan (2010, p. 97), os SOCs “são sistemas conceituais que

representam determinado domínio por meio da sistematização dos conceitos e das relações semânticas que se estabelecem entre eles.” Já Hjørland (2007) descreveu os SOC's como ferramentas que abordam a interpretação organizada de estruturas do conhecimento, também conhecidos como ferramentas semânticas. Além disso, são estruturas sistemáticas que têm o propósito de construir modelos abstratos da realidade por meio da representação dos conceitos de um domínio.

Zeng (2008, p. 161) apresenta a visão geral de alguns tipos de SOC agrupados de acordo com a complexidade de suas estruturas e principais funções, como é mostrado na FIG. 2.

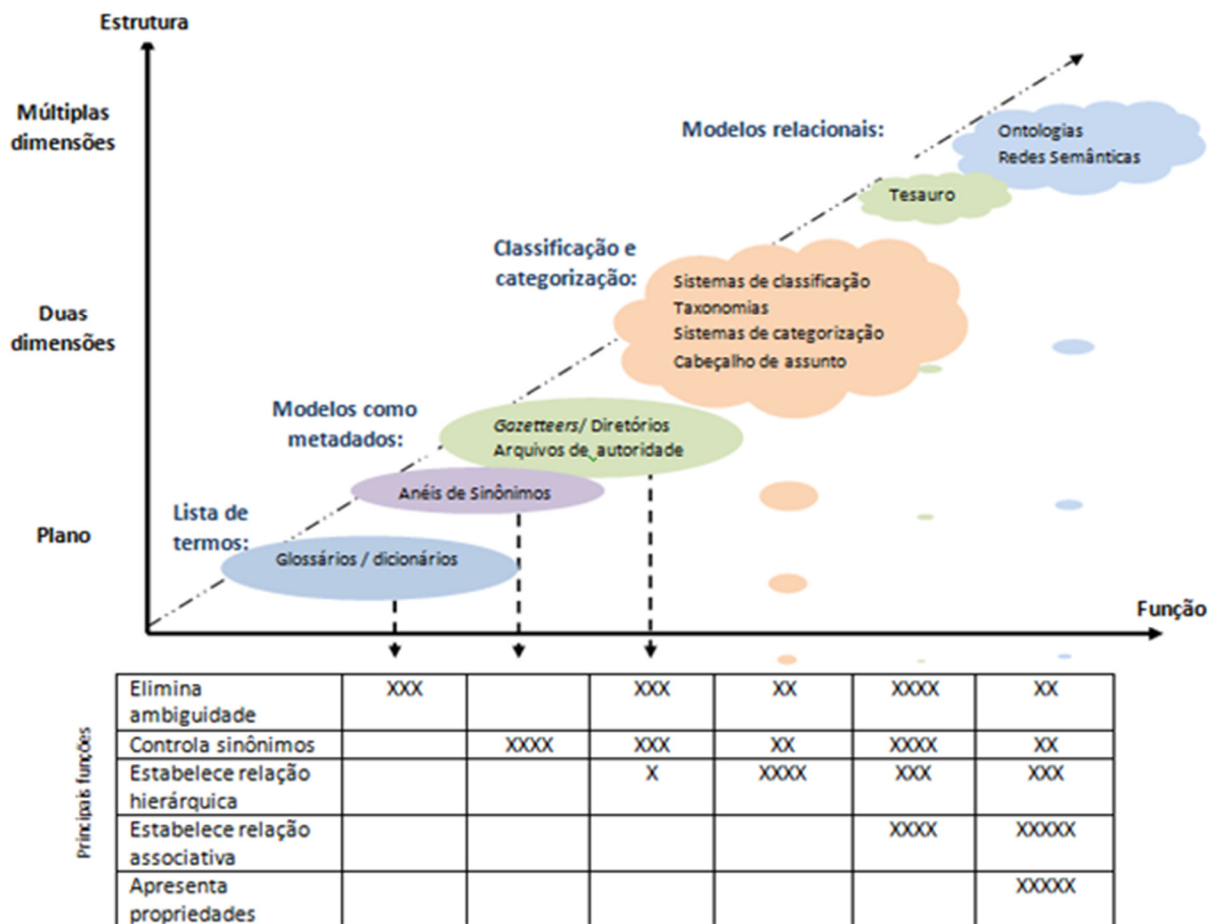


FIGURA 2 – Tipos de Sistema de Organização do Conhecimento
 Fonte: Adaptado de Zeng, 2008, p. 161.

Na literatura das áreas de BCI, encontram-se vários instrumentos nomeados Sistemas

de Organização do Conhecimento; entre eles: tesouros, taxonomias sistemas de classificação, redes semânticas, ontologias. Neste trabalho, não serão abordados todos esses instrumentos, somente as ontologias, que serão mais bem detalhadas na subseção 2.5.

2.3 Definições de conceito e termo utilizadas neste trabalho

Com o propósito de deixar claras as linhas teóricas utilizadas neste trabalho, apresenta-se uma breve definição de conceito e termo.

2.3.1 O Conceito

A origem etimológica da palavra “conceito” vem do latim *conceptus* (do verbo *concipere*) e expressa “coisa concebida” ou “formada na mente”. Na língua geral, há ambiguidade e esse termo é usado em diferentes sentidos, podendo significar “noção”, “juízo”, “opinião”, “ideia” ou “pensamento”. (FERRATER-MORA, 2004³ citado por MACULAN, 2015).

De acordo com Maculan (2015), os estudos sobre “conceito” têm origem na Filosofia Clássica e existem diferentes entendimentos para o seu significado. Neste trabalho, não iremos tratar das discussões sobre conceito na Filosofia, apenas mencionar as definições que possuem relação com o significado de conceito aqui tratado.

Entre as definições de conceito encontradas na literatura das áreas de Filosofia, Terminologia, Biblioteconomia e Ciência da Informação, estão as seguintes:

- Soma total dos enunciados verdadeiros sobre um objeto (DAHLBERG, 1978).

³ FERRATER-MORA, J. *Dicionário de filosofia*. 2. ed. São Paulo: Loyola, 2004. t.1-4.

- Unidade do conhecimento criada pela combinação única de características (ISO 1087-1, 2000).
- Construto mental que representa um objeto individual material ou imaterial (GOMES, 1990).
- Representação mental de um objeto, que somado a uma unidade lexical que o denomina, forma o termo (SALES, 2008).
- Produto da cognição humana (SMITH, 2004).

Percebe-se, com base nas definições apresentadas, que o conceito é algo abstrato que precisa ser representado por um signo, para que possa ser utilizado, principalmente no âmbito de sistemas de organização do conhecimento.

Tratando da definição de Dahlberg sobre conceito, recorre-se à sua Teoria do Conceito, em que ela o apresenta como produto de uma tríade: referente, que representa aquilo que será conceitualizado (pode ser um objeto, atividade, característica, processo, fato, etc.); características, que se tratam dos enunciados verdadeiros ou atributos de um referente; e forma verbal (termo), que se refere ao vocábulo usado para significar o referente. Essa tríade é apresentada no Triângulo do Conceito (FIG. 3), que é uma síntese da Teoria do Conceito.

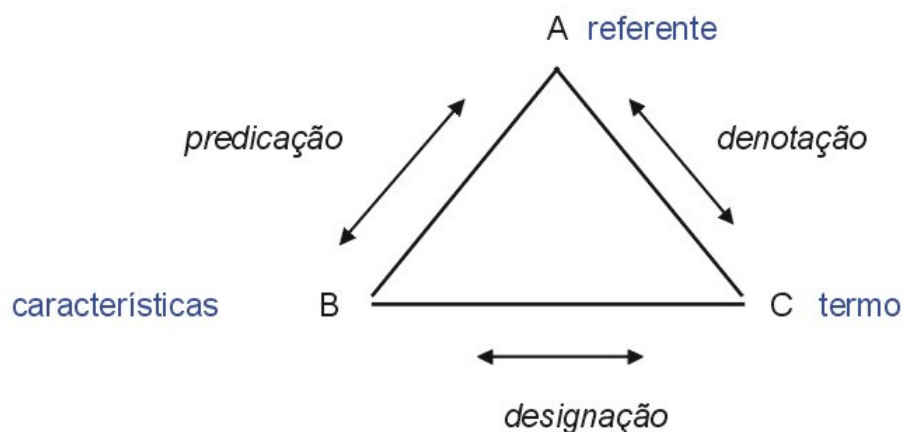


FIGURA 3 – Triângulo do Conceito de Dahlberg
Fonte: Adaptado de DAHLBERG, 1978.

Uma exemplificação de um objeto qualquer da realidade sob o prisma do Triângulo do Conceito pode ser útil para uma melhor compreensão. O objeto árvore, por exemplo, em que o referente é aquilo que se pretende conceitualizar; portanto, o conceito de árvore. As características podem ser: é um vegetal, possui tronco lenhoso, é uma planta de grande porte, possui raízes pivotantes, forma ramos bem acima do nível do solo; e a forma verbal é o termo “árvore”.

No âmbito desta pesquisa, estaremos nos referindo ao sentido de conceito como o elemento composto pela tríade: referente, significado, expressão verbal (termo); ou seja, aquilo que possui um significado (características verdadeiras) e que é representado por uma ou mais expressões verbais.

2.3.2 O Termo (expressão verbal)

O termo é objeto de estudo da Terminologia e é definido por Cabré (1999,) como unidades lexicais [palavras ou expressões] acionadas pelas suas condições pragmáticas de conformidade com um tipo de comunicação. A norma ISO 1087-1 (2000) define termo como unidade linguística que designa, ou seja, que representa um conceito.

Os estudos sobre o “termo” vão além da área da Terminologia, podem ser estudados também no âmbito da Biblioteconomia e Ciência da Informação. Cunha e Cavalcanti (2008, p. 360), definem “termo” como “uma palavra ou expressão empregada para a inclusão temática de um item em sistemas de informação e para recuperação posterior”. Os autores apresentam uma série de tipos de termos, como: termo abrangente, termo associado, termo autorizado, termo candidato a descritor, termo controlado, termo de acesso, etc. Percebe-se na definição dos autores que a ideia de termo está bem associada com uma das atividades da biblioteconomia, o tratamento da informação, em que um recurso de informação (um documento, por exemplo) é representado tematicamente por um termo.

De acordo com Maculan (2015), termo é

uma unidade-padrão, podendo ser constituído de uma ou mais palavras, formando uma expressão, e é considerado como uma unidade lexical em um dado domínio, utilizada para designar um conceito (MACULAN, 2015, p. 79).

A definição apresentada por Maculan (2015) está relacionada com a definição de termo utilizada neste trabalho, ou seja, entende-se termo como um signo verbal utilizado para representar um significado (ou conceito) em um domínio específico.

A palavra “termo” geralmente vem associada às palavras “conceito” ou “significado”, isso porque o termo é o signo utilizado para representar o conceito. Dessa forma, o termo é dependente de um conceito, de algo para representar.

2.4 Tesouros AGROVOC e NAL

Tesouros são vocabulários controlados organizados por meio de termos descritores relacionados de forma semântica, por meio das relações de hierarquia, equivalência e de associação; e cobrem de forma específica um domínio do conhecimento (AMERICAN, 2003; CUNHA, CAVALCANTI, 2008). No contexto desta pesquisa, utilizaram-se os tesouros do domínio da agricultura AGROVOC e NAL como insumos metodológicos.

O tesouro AGROVOC, combinação das palavras agricultura e vocabulário, é um vocabulário controlado, multilíngue, que cobre todas as áreas de interesse da Organização das Nações Unidas para Alimentação e Agricultura (em inglês, Food and Agriculture Organization of United Nations, FAO). Essas áreas incluem alimentação, nutrição, agricultura, pescaria, silvicultura, meio ambiente, etc. O vocabulário possui mais de 32 mil conceitos, disponíveis em 23 idiomas, são eles: árabe, chinês, tcheco, inglês, francês, alemão, hindu, húngaro, italiano, japonês, coreano, laosiano, malaio, persa, polonês, português, russo, eslovaco, espanhol, telugu, tailandês, turco e ucraniano.

O AGROVOC foi primeiramente publicado no início da década de 1980 pela FAO em inglês, espanhol e francês. Seu objetivo era servir como um vocabulário controlado para indexação de publicações na agricultura e tecnologia. Em 2000, o AGROVOC abandonou a publicação impressa e migrou para o meio digital, com os dados armazenados por uma base de dados relacional. Em 2004, testou-se uma conversão para a linguagem OWL (Web Ontology Language)⁴ e em 2009 o AGROVOC se tornou um recurso SKOS (Simple Knowledge Organization System)⁵. Atualmente, está disponível em 21 idiomas como esquema conceitual em SKOS-XL (Simple Knowledge Organization System eXtension for Labels)⁶; também publicado como um conjunto Linked Open Data (LOD)⁷ alinhado a 13 conjuntos de dados relacionados à agricultura e pode ser acessado por meio do portal da Agricultural Information Management Standards (AIMS).

O AGROVOC pode ser utilizado para a indexação de documentos, para a familiarização com algum termo, pois é possível acessar a sua definição e as relações com outros termos. O tesouro é utilizado por pesquisadores, bibliotecários e gerentes de informação para recuperar e organizar dados em sistemas de informação agrícola.

O NAL Thesaurus é uma ferramenta Web, originalmente desenvolvida pela National Agricultural Library com o objetivo de conhecer as necessidades do Serviço de Pesquisa Agrícola (em inglês, Agricultural Research Service – ARS) do Departamento de Agricultura Americano (em inglês, United States Department of Agriculture – USDA). A primeira edição foi publicada em 2002 apenas no meio digital.

⁴OWL: Web Ontology Language (OWL) é uma linguagem da Web Semântica desenvolvida para representar conhecimento rico e completo sobre coisas, grupos de coisas e as relações entre essas coisas. (W3C, 2012).

⁵SKOS: Simple Knowledge Organization System é uma recomendação da W3C desenvolvida para representar tesouros, esquemas de classificação, taxonomias, cabeçalhos de assunto, ou qualquer outra estrutura de vocabulário controlado, numa linguagem que possa ser interpretada pela máquina. (W3C, 2012).

⁶SKOS-XL: Simple Knowledge Organization System eXtension for Labels é uma extensão para o SKOS que fornece um suporte adicional para descrever e conectar entidades léxicas. (W3C, 2012).

⁷LOD: Linked Open Data é um conjunto de dados conectados (Linked Data) de conteúdo aberto.

Em 2007, as definições dos termos do tesouro foram publicadas como o Glossário de Termos Agrícolas. Desenvolvem-se a interface de busca e a navegação alfabética com o propósito de facilitar a localização das definições em mais de 2400 termos. No mesmo ano, a National Agricultural Library publica a versão do tesouro em espanhol com a colaboração da American Distance Education Consortium (ADEC). Desde 2007, o Inter-American Institute for Cooperation on Agriculture (IICA) e a National Agricultural Library em parceria mantêm o tesouro para dar suporte ao avanço da informação agrícola no continente americano.

Em 2010, o tesouro fica disponível como tecnologia *Linked Open Data*, que permite que as informações contidas no tesouro sejam “interpretadas” pelos computadores. Isso possibilita uma maior integração entre o NAL Thesaurus e os recursos desenvolvidos no âmbito da Web Semântica.

O NAL é organizado em 17 categorias de assunto, que são utilizadas para navegar no tesouro em uma disciplina específica. O escopo de assunto da agricultura é amplamente definido e inclui a terminologia de apoio das ciências biológicas, física e ciências sociais. O NAL é usado principalmente para indexação com o propósito de melhorar a recuperação da informação agrícola. Além disso, a Food Safety Research Information (FSRIO) e a Agricultural Network Information Center (AgNIC) também utilizam o tesouro como um vocabulário para indexação para seus sistemas de informação e para localização da informação nos *websites* e base de dados.

2.5 Ontologias

Nesta subseção, apresenta-se uma noção sobre ontologias, começando pelas definições, depois suas características, tipologia e utilidade.

2.5.1 Definições

O termo “ontologia” vem do grego: *onto* (ser) e *logos* (palavra) e, de acordo com Moreira et al. (2004), foi cunhado por Rudolf Goclenius em seu *Lexicon philosophicum* e Jacob Lorhard em seu *Theatrum philosophicum*, em 1613. Na literatura, há várias definições para o termo; algumas são consensuais e apenas se completam e outras seguem linhas diferentes. Sob o ponto de vista da Filosofia, a ontologia é uma teoria sobre a natureza da existência.

Constatam-se, por meio da literatura, estudos de ontologia envolvendo a Biblioteconomia e Ciência da Informação (BCI), Ciência da Computação (CC), Filosofia, entre outros campos. É útil distinguir “Ontologia” – aqui a palavra “Ontologia” aparece no singular e com letra maiúscula em conformidade com o que é discutido por Guarino (1998) – como disciplina teórica da Filosofia de “ontologias” (no plural e com letra minúscula) como um artefato para aplicações, geralmente empregado no âmbito da Biblioteconomia e Ciência da Informação e Ciência da Computação.

No contexto da CI, Vickery (1997) foi um dos pioneiros neste campo a dar atenção ao assunto ontologia na organização e representação do conhecimento. Logo depois, outros pesquisadores da área começaram a estudar o tema, como Soergel (1999), Gilchrist (2003). Os estudos da temática ainda são incipientes na área, mas têm evoluído nos últimos anos, dada a potencialidade das ontologias para contribuir com a CI.

Soergel (1999) diz que “ontologia” é uma classificação de categorias básicas. O autor faz ainda uma crítica à falta de comunicação entre as áreas de conhecimento, pois, de acordo com ele, o termo “ontologia” é apenas um novo nome para “classificação”, termo que já é usado pela BCI há muito tempo. Assim, o autor sugere que os estudos sobre ontologias possam basear-se em técnicas já consolidadas no contexto da BCI.

Segundo Almeida e Bax (2003), uma ontologia pode ser concebida como uma linguagem, pois estabelece um conjunto de termos que, posteriormente, poderão ser usados para formular consultas sobre uma base de conhecimento. Nessa mesma linha, Furgeri (2006) define ontologia como uma técnica que permite organizar a informação e auxiliar sua recuperação, por meio de um vocabulário de termos categorizados que descrevem determinada realidade.

O objetivo da ontologia, apontado por Souza e Alvarenga (2004), tem grande relação com essa pesquisa. Eles dizem que o objetivo da ontologia é “criar um vocabulário controlado para se trocar informações entre os membros de uma comunidade, sejam eles humanos ou agentes inteligentes” (SOUZA; ALVARENGA, 2004, p. 137). A ideia de a ontologia funcionar como uma linguagem comum para um determinado domínio e assim possibilitar o compartilhamento do conhecimento é bastante citada na literatura. Gómez-Pérez e Pazos (1995), por exemplo, dizem que as ontologias são plataformas que possibilitam o reuso do conhecimento por meio do estabelecimento de vocabulários comuns e interpretações semânticas dos termos. Nessa mesma linha, Uschold e Gruninger (1996) dizem que uma ontologia padroniza a linguagem de um domínio e facilita a comunicação entre os usuários, por meio de um vocabulário compartilhado do domínio.

Sob o ponto de vista da Ciência da Computação, o termo ‘ontologia’ surgiu em 1967 em um trabalho sobre modelagem de dados de G. H. Mealy (SMITH, 2004). A ontologia é definida no contexto da CC como um artefato de engenharia, composto de um vocabulário de termos organizados em uma taxonomia, as definições dos termos e um conjunto de axiomas formais usados para criar relações e limitar as interpretações (NOY; HAFNER, 1997). Sob esse enfoque, Guarino (1998) define ontologia como

[...] um artefato constituído de um vocabulário usado para descrever uma certa realidade, mais um conjunto de fatos explícitos e aceitos que dizem respeito ao sentido pretendido para as palavras do vocabulário. Esse conjunto de fatos tem a forma da teoria da lógica de primeira ordem, em que as palavras do vocabulário aparecem como predicados unários ou binários (GUARINO, 1998, p.4, tradução própria⁸).

⁸ [...] an ontology refers to an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the

Gruber (1993, p. 1) traz uma das principais definições do termo; ele diz que “ontologia é uma especificação formal, explícita e compartilhada de uma conceitualização.” A palavra “formal”, apontada por Gruber, quer dizer que a ontologia deve possuir um nível de formalidade que possa ser compreendida pelas máquinas; já a palavra “explícita” significa que todos os recursos empregados na ontologia devem ser claros; “compartilhada” quer dizer que aquele conhecimento trazido pela ontologia deve ser consensual, deve ser aceito por um grupo de pessoas e não apenas por alguns indivíduos; quanto à palavra “conceitualização”, trata-se de um modelo abstrato de algum fenômeno do mundo por meio da identificação dos conceitos relevantes daquele fenômeno (STUDER et al., 1998).

Para Chandrasekaram et al. (1999), uma ontologia é a representação de um vocabulário, geralmente especializado de algum domínio. De modo mais preciso, não é o próprio vocabulário que se qualifica como uma ontologia e sim a conceitualização que os termos presentes no vocabulário pretendem capturar. Percebe-se nessa definição que a relevância de uma ontologia é esclarecer a estrutura de conhecimento e não apenas listar termos de certo domínio.

2.5.2 Características, tipos e aplicações das ontologias

As ontologias são modeladas para representar os conceitos de diferentes domínios, o que é facilitado pelo fato de elas serem elaboradas de maneira a representar explicitamente os conceitos e seus relacionamentos por meio da formalização desses relacionamentos através de linguagens legíveis pelas máquinas. Os componentes gerais das ontologias que possibilitam que elas representem o conhecimento são:

- Classes e subclasses: são coleções de elementos compostos de um conjunto

vocabulary words. This set of assumptions has usually the form of a first-order logical theory, where vocabulary words appear as unary or binary predicate names (GUARINO, 1998, p. 4).

de atributos semelhantes. Compõem as unidades básicas de uma ontologia e formam conceitos que determinam certo objeto.

- **Propriedades/atributos:** descrevem as características e/ou qualidades das classes e subclasses.
- **Relações:** são responsáveis pelos relacionamentos semânticos entre os conceitos de um dado domínio.
- **Axiomas ou restrições:** são as regras declaradas sobre as relações. Permitem fazer inferências de conhecimento que não estão indicados nas taxonomias da ontologia.
- **Instâncias:** representam um determinado objeto de um conceito, ou seja, são os próprios dados da ontologia.

Ontologias podem ainda ser classificadas em tipos. Na literatura, encontra-se mais de uma tipologia para classificá-las. Podem ser quanto à função, grau de formalismo, aplicação, estrutura e conteúdo (ALMEIDA; BAX, 2003). Este trabalho baseia-se na tipologia abordada por Maedche (2002), cuja classificação de ontologias se dá de acordo com sua estrutura, como segue:

1. Ontologias de alto-nível, também conhecidas como “Ontologias de Fundamentação”: descrevem conceitos muito gerais, como espaço, tempo, evento, entre outros, que são independentes de um problema particular ou domínio.
2. Ontologias de domínio: descrevem o vocabulário relacionado a um domínio genérico, por meio da especialização de conceitos introduzidos nas ontologias de alto-nível.
3. Ontologias de tarefa: descrevem um vocabulário relacionado a uma tarefa ou a uma atividade genérica, por meio da especialização de conceitos introduzidos nas ontologias de alto-nível.
4. Ontologias de aplicação: são as ontologias mais específicas, utilizadas dentro das aplicações, que especializam as ontologias de domínio e as de tarefas.

Neste estudo, o foco é a ontologia de domínio, que será mais bem detalhada em seguida. As ontologias de domínio têm como característica um vocabulário específico e o objetivo principal é criar uma linguagem comum para o domínio representado, pois isso ajuda a reduzir a desordem terminológica e contribui para o compartilhamento do conhecimento entre os membros de uma mesma comunidade.

A literatura mostra inúmeras possibilidades de utilização das ontologias; entre elas, podem-se citar: recuperação da informação em ambientes digitais, processamento da linguagem natural, gestão do conhecimento, representação do conhecimento, comunicação, interoperabilidade, reuso do conhecimento. Isso porque as ontologias possuem várias vantagens, entre elas: permitem compartilhamento de conhecimento; fornecem uma descrição exata do conhecimento; por serem escritas em linguagem formal, ajudam a eliminar as ambiguidades da linguagem natural; possibilitam a comunicação entre pessoas; permitem a realização de inferências (GUARINO, 1998).

Para fins de estudo nesta pesquisa, foi escolhida uma ontologia na área específica da Agricultura e Recursos Hídricos, a OntoAgroHidro, que será descrita na subseção 2.5.4.

2.5.3 Questões de competência em ontologias

Questões de Competência (QCs), no contexto das ontologias, são expressões de perguntas que uma ontologia deve responder. Segundo Ren et al. (2014), são sentenças em linguagem natural que expressam modelos para tipos de perguntas às quais se quer que a ontologia seja capaz de responder corretamente.

A capacidade da ontologia de responder às questões de competência formuladas pode ser considerada um requisito funcional. Como afirmam Noy e McGuinness (2001), após a formulação das questões de competência, algumas perguntas servem como um teste decisivo; são elas: a ontologia possui informação suficiente para responder

a esses tipos de questões?; as respostas requerem um nível particular de detalhe ou a representação em uma área específica? Ainda de acordo com as autoras, essas questões são apenas um esboço e não precisam ser exaustivas.

Gruninger e Fox (1995) separam as QCs em dois tipos: as questões de competência informais e as formais. As primeiras são expressas em linguagem natural e não exigem conhecimento sobre Lógica Descritiva (Description Logics), facilitando o uso por pessoas não familiarizadas com essa Lógica; já as segundas, de competências formais, são expressas em linguagem formal e são criadas a partir das questões informais, usando axiomas. É preciso destacar que todos os termos usados na declaração das questões de competência formais precisam conter a terminologia da ontologia.

QCs desempenham um papel fundamental no contexto das ontologias. Elas podem ser utilizadas tanto na fase de desenvolvimento da ontologia quanto na fase de avaliação. Na fase de desenvolvimento, podem-se usar as QCs para ajudar a modelar o domínio, ou seja, por meio de questões às quais a ontologia deveria ter competência para responder, pode-se ter uma noção de quais são os conceitos relevantes do domínio e das relações entre eles. Na fase de avaliação da ontologia (que é abordada no presente trabalho), elas podem ser utilizadas para identificar as falhas da ontologia na modelagem do domínio e assim contribuir na avaliação de qualidade da ontologia.

Além dos autores citados neste trabalho, os estudos sobre QCs no âmbito das ontologias podem ser vistos nos trabalhos de Gruninger e Fox (1995) e Uschold e Gruninger (1996), que trataram das QCs no desenvolvimento de ontologias; Choi et al. (2010), que transformaram QCs informais em formais com o propósito de avaliar uma ontologia denominada Small Molecule Ontology (SMO); Presutti et al. (2012) e Suárez-Figueroa, Pradel e Hernández (2012), que utilizaram questões de competência como uma forma de capturar requisitos para ontologias.

2.5.4 OntoAgroHidro

A OntoAgroHidro é uma ontologia de domínio criada dentro de um projeto da Rede Agro-Hidro da EMBRAPA (como está descrito na seção 1.1 deste trabalho). A ontologia foi construída por especialistas da EMBRAPA utilizando a linguagem OWL 2.0⁹ e o software de edição de ontologias Protégé¹⁰ (versão 4). O objetivo principal da ontologia é a representação do conhecimento sobre os impactos das mudanças climáticas e da agricultura nos recursos hídricos. Como afirmam Bonacin, Nabuco e Pierozzi Júnior (2015), é um cenário bem ambicioso para se modelar, pois engloba particularidades do domínio das mudanças climáticas associadas com a agricultura e a hidrologia.

O processo de desenvolvimento da OntoAgroHidro envolveu o trabalho de engenheiros do conhecimento e especialistas do domínio. Para iniciar o processo de modelagem, criaram-se duas perguntas que representavam o macrocenário; são elas: 1) Quais são os impactos da agricultura e das mudanças climáticas nos recursos hídricos?; 2) Quais são os impactos da quantidade e qualidade de água na agricultura? (e vice-versa).

Especificamente, as fases de construção da OntoAgroHidro, de acordo com Bonacin, Nabuco e Pierozzi Júnior (2014), foram as seguintes:

1. Análise do problema e *design* cooperativo com os especialistas do domínio; para isso, foram usadas as ferramentas CMapTools¹¹ e yEd¹². Com base nas questões que motivaram a criação da ontologia, modelaram-se os primeiros rascunhos da ontologia usando mapas conceituais.

⁹ OWL 2.0: A Web Ontology Language 2 é linguagem de ontologia para a Web Semântica com o significado definido formalmente. OWL 2.0 fornece classes, propriedades, instâncias e valores de dados que são armazenados como Documentos da Web Semântica. (W3C, 2015).

¹⁰ Protégé é um software para desenvolvimento e edição de ontologias, criado na Universidade de Stanford dos E.U.A. Disponível em: <http://protege.stanford.edu/>.

¹¹ CMapTools: *Software* para desenvolvimento de mapas conceituais.

¹² yEd: *Software* para desenvolvimento de diagramas e mapas conceituais. <http://www.yworks.com/en/products/yfiles/yed>

2. Modelagem dos conceitos principais do domínio usando mapas conceituais como ponto de partida.
3. Escolha e especialização dos conceitos principais por meio da análise em artigos, relatórios, entre outros documentos que poderiam contribuir para descrever os conceitos principais da ontologia.
4. Análise de modelos de ontologias existentes para reuso. Durante essa fase, investigou-se a existência de modelos de representação do conhecimento que fossem capazes de modelar os conceitos principais da ontologia.
5. Os dois modelos existentes para reuso escolhidos foram as ontologias SWEET¹³ da NASA (*National Aeronautics and Space Administration*) e a CUASHI¹⁴, pois apresentavam uma valiosa contribuição com relação às definições e conceitos que seriam usados na OntoAgroHidro.
6. Por fim, a ontologia foi interativamente avaliada e submetida para checagem da consistência. Isso incluiu uma validação lógica (usando o *reasoner* Pellet¹²) e a validação com especialistas do domínio em dois Workshops.

A atual versão da OntoAgroHidro possui aproximadamente 8500 elementos, divididos entre classes, subclasses, propriedades, relações e instâncias. A FIG. 4 apresenta as sete classes principais da ontologia e suas respectivas subclasses. As classes principais são:

- **Definition:** esta classe refere-se às definições de conceitos necessários para caracterizar outros conceitos no domínio.
- **EnvironmentalTransformation:** esta classe refere-se à transformação ambiental como um fenômeno que pode ter causas antrópica e/ou natural. A transformação ambiental é causada por um agente; é resultado de um evento e altera um objeto.
- **EnvironmentalTransformationObject:** esta classe refere-se ao objeto em que a transformação ambiental ocorre. Na OntoAgroHidro, o foco é nos recursos

¹³ SWEET (Semantic Web for Earth and Environmental Terminology): <https://sweet.jpl.nasa.gov/>

¹⁴CUAHSI (Consortium of Universities for the Advancement of Hydrologic Science, Inc.): ontologia sobre hidrologia. <http://his.cuahsi.org/ontologyfiles.html>

hídricos, ou seja, o objeto alterado pela transformação ambiental.

- **EnvironmentalTransformationAgent:** esta classe refere-se aos agentes que causam a transformação ambiental.
- **Dimension:** esta classe refere-se às dimensões (no sentido de múltiplas variáveis/aspectos) de um problema especificado. As dimensões são especificadas e relacionadas à classe definição.
- **Event:** esta classe refere-se aos eventos que geram transformações ambientais. Um evento tem múltiplas dimensões; por exemplo, ocorre em um período e local, está associado com uma cultura de colheita. Um evento é descrito por um agente de informação.
- **InformationAgent:** esta classe refere-se aos agentes e recursos que contêm/proveem a informação na rede.

Os desenvolvedores da ontologia desejam que ela possa auxiliar no compartilhamento e integração da informação entre as diversas instituições e pesquisadores dentro do domínio modelado. Além disso, é de interesse deles o uso da ontologia como um recurso para melhorar a interoperabilidade entre bases de dados e assim permitir seu uso na recuperação da informação.

Como apontam Bonacin, Nabuco e Pierozzi Júnior (2015), a OntoAgroHidro já integra “uma conceitualização interessante do domínio ao qual pretendeu representar”, porém, ainda é necessária uma ampliação conceitual e terminológica, fazendo uso de outros sistemas de organização do conhecimento, como os tesauros, por exemplo.

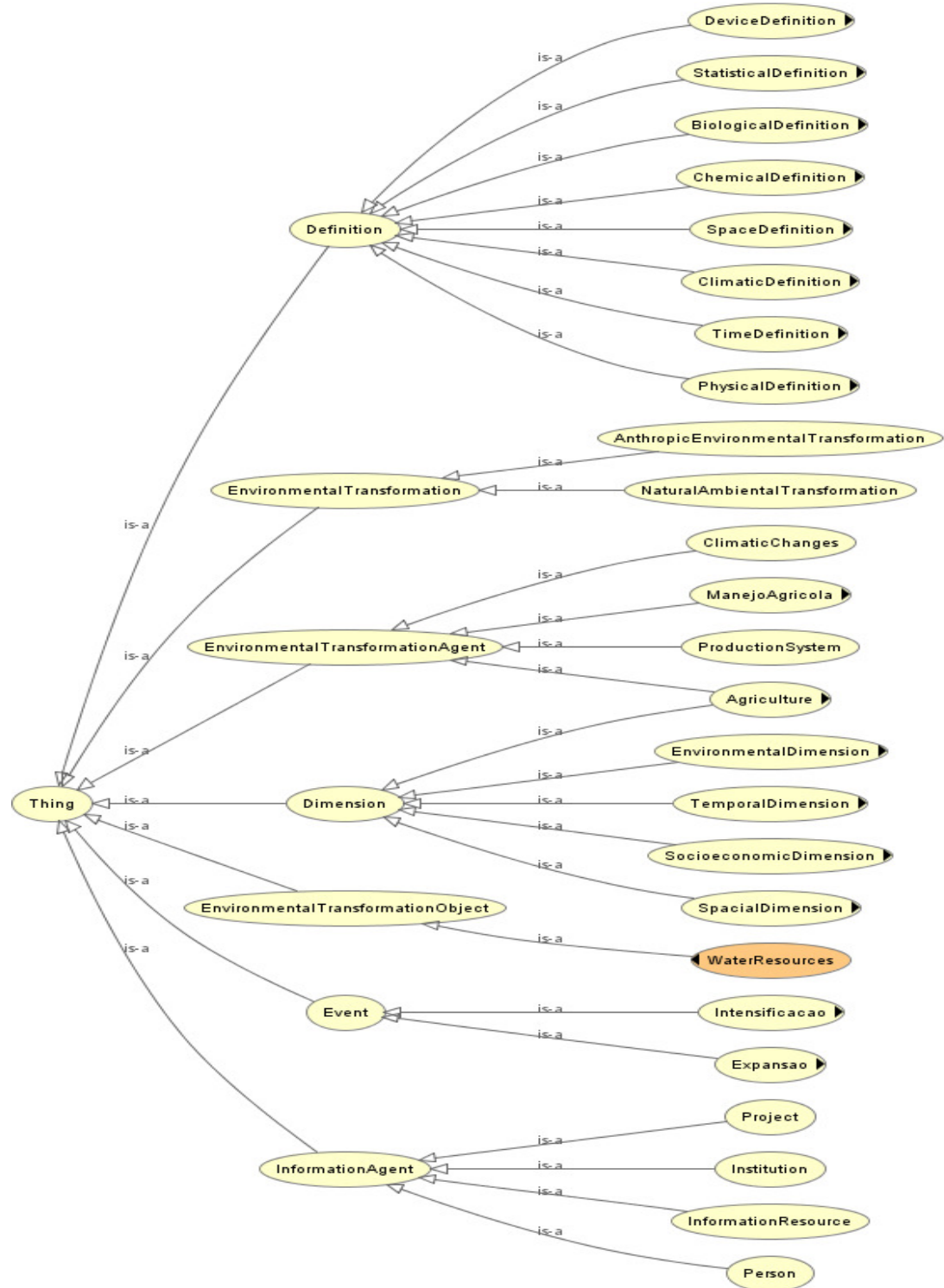


FIGURA 4 – Visão geral da OntoAgroHidro e suas classes e subclasses principais

Fonte: Elaborada pelo autor.

3 AVALIAÇÃO DE ONTOLOGIAS: UMA BREVE REVISÃO

O estudo da avaliação de ontologias aqui 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 sem uma metodologia padrão que orientasse a construção delas, 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 TOVE¹⁵, Methontology¹⁶, NeON¹⁷, Método 101¹⁸, Método Sensus¹⁹, não apresentam um consenso com relação às etapas para construção de ontologias, sendo direcionadas a diferentes propósitos e aplicações. Portanto, após leituras realizadas sobre esse tópico, constatou-se que a literatura ainda é incipiente com relação a modelos unificados para o desenvolvimento de ontologias. Então, com tal deficiência, a temática avaliação de ontologias vem como uma solução, pois tem o propósito de analisar a ontologia para ver se ela está representando de maneira adequada o modelo a que se propõe representar.

Goméz-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

¹⁵ TOVE (Toronto Virtual Enterprise Ontology Project): Desenvolvida por Gruninger e Fox (1995).

¹⁶ Methontology: Desenvolvida no laboratório de Inteligência Artificial da Universidade Politécnica de Madri pelos pesquisadores Goméz-Pérez, Fernández e Vicente (1996); Fernández, Goméz-Pérez e Juristo (1997).

¹⁷ NeON: Metodologia desenvolvida por Suárez-Figueroa et al. (2008).

¹⁸ Método 101: Desenvolvido por Noy e McGuinness (2001).

¹⁹ Método Sensus: Desenvolvido pelo grupo de linguagem natural – Information Sciences Institute da Universidade da Carolina do Sul.

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; afirma 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 própria²⁰).

Antes de abordar as propostas de avaliação de ontologias, cabe classificá-las para clarear o entendimento. A literatura mostra que não há um consenso com relação à classificação. Isso se deve à diversidade de propostas existentes. 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. Brank, Grobelnik e Mladenić (2006) apontaram outro tipo de classificação, que vai além das classificações dicotômicas (ou seja, que possuem mais de duas categorias para classificação), separando as propostas de avaliação de ontologias em quatro classes. São elas:

- com base na comparação da ontologia a um “modelo padrão” ou *Gold Standard*;
- com base no uso da ontologia em uma aplicação/tarefa e avaliação dos resultados;
- que envolvem comparações com fontes de dados não estruturados sobre um domínio que será coberto pela ontologia;

²⁰The 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)

- aquelas em que a avaliação é feita por humanos, que tentam estimar o quão bem a ontologia responde a um conjunto predefinido de critérios, padrões, requerimentos.

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.

A partir de 1995, várias propostas de avaliação de ontologias foram desenvolvidas, principalmente por pesquisadores ligados à Ciência da Computação. Como poderá ser visto nas próximas subseções, 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. Portanto, algumas propostas de avaliação de ontologias não apresentam resultados.

3.1 Avaliações com base na comparação a um modelo padrão

Nesse tipo de avaliação, a ontologia é comparada a outro modelo construído de forma ideal, conhecido como *Gold Standard*, “padrão ouro” ou ainda “modelo padrão.” Esse

modelo pode ser outra ontologia, que servirá de base para a avaliação. Na literatura, encontram-se alguns trabalhos de pesquisadores que utilizam essa abordagem; entre eles, Maedche e Staab (2002), Gangemi et al. (2006); Brank, Grobelnik e Mladenić (2006), Dellschaft e Staab (2006), Zavitanos, Paliouras e Vouros (2011), que são apresentados a seguir.

Maedche e Staab (2002) apresentam uma abordagem com base na medida de similaridade para comparar ontologias. A ontologia avaliada é comparada com um “padrão ouro” (*Gold Standard*), que pode ser outra ontologia (como é usada no trabalho dos autores) ou pode ser uma coleção de documentos de onde são retirados os termos representativos do domínio. No trabalho dos autores, a comparação de similaridades entre as ontologias é feita no nível lexical e conceitual. No primeiro, objetiva-se averiguar a semelhança entre os vocabulários das ontologias e, no segundo, objetiva-se comparar os conceitos e as relações entre elas.

Na proposta de Gangemi et al. (2006) para avaliação de ontologias, são apresentados dois modelos: uma metaontologia (ou seja, uma ontologia que possui ontologias e elementos de ontologias em seu domínio), denominada O^2 ; e uma ontologia para a seleção de outras ontologias, denominada oQual. Com base na metaontologia, é possível identificar três tipos principais de métricas para avaliação de ontologias: 1) medidas estruturais, que enfatizam sintaxe e semântica formal nas ontologias representadas como grafos; 2) medidas funcionais, que se referem ao uso pretendido da ontologia; 3) medidas de usabilidade, relacionadas ao nível de anotação da ontologia. A metaontologia O^2 é complementada pela ontologia oQual, que possibilita a realização de análises sobre elementos, processos e atributos da ontologia sob avaliação.

Na proposta de Brank, Grobelnik e Mladenić (2006), a abordagem para avaliação de ontologia permite a avaliação automática com base nas instâncias. É baseada no paradigma do *Gold Standard* e o foco principal é comparar quão bem determinada ontologia se assemelha ao *Gold Standard* no arranjo de instâncias dentro dos conceitos e o arranjo hierárquico entre os conceitos. Os autores dizem que a avaliação

é similar às outras que são baseadas na comparação a um *Gold Standard*; porém, na abordagem deles, a avaliação destaca os atributos dos conceitos da ontologia e não a descrição em linguagem natural dos conceitos e instâncias. Ressalta-se que o pré-requisito para a execução desse tipo de avaliação é que a ontologia avaliada precisa ser baseada no mesmo conjunto de instâncias que o *Gold Standard*.

Na proposta de Dellschaft e Staab (2006) apresenta-se um esquema para avaliação de ontologias com base em um *Gold Standard*. Os autores focam no problema das métricas para avaliação de ontologias, ou seja, qual das métricas citadas na literatura permite realizar uma avaliação multidimensional da ontologia, que abrange vários níveis e não apenas um, como acontece na maioria das propostas de avaliação. Assim, faz-se uma análise de várias métricas de avaliação de ontologia e em seguida são realizadas avaliações empíricas dessas métricas para no fim propor as métricas ideais para uma avaliação multinível de uma ontologia.

Zavitanos, Paliouras e Vouros (2011) apresentam uma proposta que transforma os conceitos e propriedades da ontologia em uma representação de espaço vetorial com o propósito de avaliá-la em comparação a um *Gold Standard*. As medidas de avaliação da proposta exploram a representação de espaço vetorial e calculam a semelhança entre as duas ontologias nos níveis léxico e relacional. Isso é feito com base em três etapas: 1) Transformação – os elementos da ontologia são transformados em distribuições probabilísticas; 2) Correspondência – comparam-se os elementos da ontologia com os elementos do *Gold Standard*; 3) Avaliação – a ontologia avaliada é penalizada quando seus elementos se diferem dos elementos do *Gold Standard*.

3.2 Avaliações com base no uso da ontologia em uma aplicação/tarefa

Nesse tipo de avaliação, a ontologia é testada com base no seu desempenho na realização de uma tarefa específica. Em uma proposta de avaliação desse tipo, Porzel e Malaka (2004) propõem uma abordagem quantitativa para avaliação de ontologias.

Os autores se baseiam na questão fundamental: quão efetiva é uma ontologia na execução de uma tarefa bem definida? Eles defendem que, se uma ontologia foi construída para realizar determinada tarefa, a melhor forma de avaliar é testando a ontologia na execução dessa tarefa específica. Na proposta dos autores, primeiro escolhe-se uma ou mais ontologias que serão utilizadas; segundo, escolhe-se a tarefa que deseja que a ontologia escolhida execute (os autores frisam que a tarefa escolhida precisa ser suficientemente complexa para que sirva como uma referência adequada na avaliação); depois, escolhe-se o “padrão ouro” (esse padrão define a medida de desempenho perfeita), ou seja, um modelo com as respostas adequadas que servirá de comparação para a ontologia avaliada. Assim, um algoritmo específico (construído pela equipe de avaliação) usa a ontologia para desempenhar a tarefa proposta. O resultado da avaliação mostra três tipos de erros: 1) erros de inserção, que indicam conceitos supérfluos encontrados na ontologia; 2) erros de eliminação, que indicam conceitos faltantes; 3) erros de substituição, que indicam conceitos ambíguos. Ainda de acordo com Porzel e Malaka (2004), as ontologias podem ser averiguadas em três níveis básicos: o escopo do vocabulário, a qualidade da taxonomia e a adequabilidade das relações semânticas.

Fernández, Cantador e Castells (2006) apresentam uma ferramenta nomeada CORE (Collaborative Ontology Reuse and Evaluation) para identificar qual ontologia de um repositório é a mais adequada para representar um domínio. Para a realização da tarefa proposta, o ambiente da ferramenta é dividido em três componentes: o primeiro recebe a descrição do problema baseada em um conjunto de termos; o segundo aplica critérios automáticos de avaliação às ontologias, com o objetivo de indicar qual se adequa melhor ao problema; o terceiro faz uso de avaliações manuais em ontologias, as quais incorporam a avaliação colaborativa de usuários. Com base nesses três componentes, o sistema elenca qual ontologia representa melhor o domínio proposto.

A proposta de Strasunskas e Tomassen (2008) visa identificar quais recursos da ontologia influenciam na tarefa de busca na Web. Elaborou-se um modelo teórico para medir o desempenho da ontologia na realização da tarefa de busca, com o objetivo de identificar a importância da qualidade da ontologia para a realização da tarefa. Em

seguida, o modelo teórico foi testado experimentalmente em quatro ontologias de diferentes domínios, o que permitiu identificar como a variação de qualidade de uma ontologia pode influenciar no desempenho de uma busca na Web.

Clarke et al. (2013) fazem um experimento de avaliação com base em tarefa da Gene Ontology (GO). Na proposta dos autores, avalia-se a qualidade e a utilidade da GO e suas anotações por meio do seu desempenho em uma tarefa comum, ou seja, utilizam-se análises enriquecidas em conjunto de genes com o objetivo de identificar quais termos da ontologia estão bem representados e quais não estão. Isso determina a efetividade das anotações da Gene Ontology fornecendo resultados biologicamente precisos. Além disso, permite entender os pontos fortes e fracos da ontologia avaliada.

3.3 Avaliações com base na comparação a uma fonte de dados

As avaliações desse tipo são feitas por meio da comparação da ontologia a uma fonte de dados, que geralmente se trata de textos do domínio coberto pela ontologia avaliada. Entre as propostas de avaliação encontradas na literatura, estão a de Patel et al. (2003), Brewster et al. (2004), Elhadad, Gabay e Netzer (2010), Ouyang et al. (2011), Hlomani e Stacey (2013) e Hlomani e Stacey (2014), que são abordadas a seguir.

Patel et al. (2003) desenvolveram um protótipo denominado Ontokhoj para buscar, classificar e ranquear ontologias com base em dados extraídos da Web semântica. As ontologias são carregadas em um software de classificação que determina se a ontologia se refere a um tópico particular.

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 coleção de textos tem-se 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 Net²¹ 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.

Elhadad, Gabay e Netzer (2010) propõem um método de avaliação de ontologias que conta com o mapeamento de instâncias de uma ontologia a documentos textuais. O processo que orienta a avaliação é o seguinte: a partir de uma determinada instância da ontologia, coleta-se grande quantidade de documentos textuais que têm relação com a instância. Em seguida, faz-se uma análise para determinar se a ontologia reflete os documentos encontrados. Com base nesse mapeamento, pode-se avaliar a adequação das relações das ontologias medindo o seu potencial de classificação por meio dos documentos textuais. De acordo com os autores, o método fornece um *feedback* concreto aos ontologistas e uma estimativa quantitativa da adequação funcional das relações de ontologias para uma melhoria na experiência da pesquisa.

Ouyang et al. (2011) apresentam uma proposta para avaliar a qualidade da ontologia em comparação a um corpus de dados utilizando três critérios: cobertura, coesão e acoplamento. Para a realização dos experimentos empíricos, utilizaram-se cinco ontologias, três delas relacionadas ao domínio de projetos, do Departamento de Ciência da Computação da Universidade de Maryland e duas relacionadas ao meio ambiente, da NASA. Os autores apresentaram expressões para o cálculo dos critérios

²¹Word Net: é um banco de dados lexical para o idioma Inglês, que agrupa palavras em conjuntos de sinônimos e fornece definições curtas. <http://wordnetweb.princeton.edu/perl/webwn>

mencionados e fizeram os cálculos de cada ontologia em comparação com dois corpus de dados da American National Corpus (OANC). Em seguida, os resultados foram tabelados e comparados em gráficos. A proposta dos autores vai além e envolve a avaliação de usuários, que, com base nos resultados da comparação, decidem qual dos critérios (cobertura, coesão e acoplamento) possui mais peso e assim qual ontologia atende melhor aos seus requisitos.

Hlomani e Stacey (2013) se baseiam na proposta de Brewster et al. (2004) para avaliar ontologias no domínio de gerenciamento de fluxos de trabalho. Os principais passos seguidos para a avaliação foram: definição do corpus, ou seja, selecionaram-se os documentos (artigos acadêmicos revisados); cálculo de similaridade entre as palavras de cada artigo e os conceitos das ontologias avaliadas; por fim, avaliação estatística, ou seja, análise de qual ontologia possuía mais similaridade com os documentos do domínio.

Em outro trabalho, Hlomani e Stacey (2014) propõem uma visão multidimensional para avaliação de ontologias com base na comparação a uma fonte de dados. De acordo com os autores, as propostas até então de avaliação de ontologias com base em uma fonte de dados consideram o conhecimento do domínio como sendo estático. Criticando essa forma de ver o conhecimento do domínio, eles propõem um esquema teórico para avaliação de ontologias que considera o conhecimento como sendo dinâmico, que pode variar com base em várias dimensões. Eles discutem que se o conhecimento do domínio é dinâmico, então as ontologias deveriam ser avaliadas com base nas mesmas dimensões que definem as características do domínio.

3.4 Avaliações feitas por humanos

As propostas de avaliação que utilizam essa abordagem geralmente são realizadas intelectualmente, de maneira que uma pessoa avalia a ontologia com base em alguns critérios preestabelecidos. Exemplos desse tipo de avaliação podem ser vistos em

Lozano-Tello e Gómez-Pérez (2004), Almeida (2009) e Orme, Yao e Etzkorn (2007).

Lozano-Tello e Gómez-Pérez (2004) adotam uma abordagem para avaliação de ontologias com base em uma taxonomia de 160 características, denominada “Estrutura Multinível de Características”, que são segmentadas em 5 dimensões. São elas: conteúdo, linguagem, metodologia, ferramentas e custos. Essas dimensões são definidas por meio de um conjunto de fatores, que são elementos requeridos para análise para obter o valor da dimensão adequada. Os fatores são os elementos fundamentais que devem ser analisados para obter o valor de cada dimensão. Dimensões, fatores e características são nomeados ‘critério’ e reunidos em uma ontologia referência. O método chamado Ontometric tenta obter o valor de cada critério por meio de fórmulas que avaliam os pesos atribuídos aos termos. O objetivo é selecionar a melhor opção entre alternativas e assim o usuário pode decidir qual ontologia é mais adequada para determinado projeto.

Orme, Yao e Etzkorn (2007) apresentam uma forma de avaliar ontologias ancorada no *feedback* de avaliadores humanos com experiência na área de avaliação de softwares. Eles propõem um conjunto de métricas, com base em conceitos da área de avaliação de software, para medir a complexidade e coesão de ontologias. Entre as métricas estão: validade do atributo, validade da unidade, validade instrumental, validade do protocolo, escada da unidade, etc. Após uma avaliação teórica de cada métrica proposta, é realizado um teste empírico com avaliadores humanos que determinam a qualidade das ontologias com base na atribuição de valores para cada métrica. Cada métrica reflete diferentes tipos de mudanças necessárias na ontologia.

Almeida (2009) propõe uma abordagem de avaliação baseada no conteúdo da ontologia. Ele apresenta uma pesquisa qualitativa para avaliar o conteúdo de uma ontologia de uma grande companhia energética brasileira. O processo de avaliação consiste de duas fases: a primeira trata-se do uso de um protótipo pelos especialistas do domínio; a segunda fase trata-se de três questionários apoiados nas seguintes dimensões: questões de competência, qualidade da informação e objetivos educacionais da taxonomia. O protótipo trata-se de um simples motor de busca, que

tem o objetivo de buscar termos correspondentes a conceitos na ontologia. Ele foi implementado com quatro funcionalidades: uma interface de busca e navegação que permite aos usuários busca por conceitos e relações; uma hierarquia de conceitos que possibilita acessar conceitos e relações por meio de *hyperlinks* organizados na taxonomia; uma visão parabólica dos conceitos que oferece aos usuários uma melhor visualização da estrutura como um todo e assim permite entender o contexto dos conceitos durante uma busca; por fim, uma interface de gerenciamento, em que é possível carregar diferentes ontologias no protótipo. O processo funcionou da seguinte forma: primeiro os especialistas do domínio utilizaram o protótipo por uma semana e depois responderam aos três questionários, constituídos de 81 perguntas (20 perguntas no questionário da dimensão questões de competência; 36 perguntas no questionário da dimensão qualidade da informação e 25 perguntas no questionário da dimensão objetivos educacionais). Cada pergunta do questionário possuía uma escala de 1 a 5 (sendo 1 'não corresponde às necessidades' e 5 'corresponde totalmente às necessidades'). Assim, com base no cálculo das escalas atribuídas pelos especialistas do domínio, é feita a avaliação da ontologia.

3.5 Considerações do capítulo

O QUADRO 3 apresenta quais níveis da ontologia cada tipo de avaliação consegue analisar. Nele é possível observar a relação entre as propostas e esses níveis.

QUADRO 3
Visão global das propostas de avaliação

Proposta de avaliação				
Nível	Padrão ouro	Baseada em aplicação	Baseada em uma fonte de dados	Feita por humanos
Léxico, vocabulário, conceito, dados	X	X	X	X
Hierárquico, taxonômico	X	X	X	X
Outras relações semânticas	X	X	X	X
Contexto, aplicação		X		X
Sintático	X			X
Estrutura, arquitetura, desenho				X

Fonte: Adaptado de Brank, Grobelnik, Mladenić, 2006, p. 2.

Constata-se, com base no quadro apresentado, que as propostas de avaliação feitas por humanos são as únicas que conseguem avaliar todos os seis níveis de uma ontologia. Já as propostas baseadas na comparação a um *Gold Standard* e as propostas baseadas no uso da ontologia em uma aplicação/tarefa permitem a avaliação de quatro níveis de uma ontologia, com a diferença em um nível de avaliação. Enquanto as propostas baseadas na comparação a um *Gold Standard* conseguem avaliar o nível 'sintático' e não permitem avaliar o nível do 'contexto e aplicação', as propostas baseadas no uso da ontologia em uma aplicação/tarefa possibilitam a avaliação do nível 'contexto e aplicação', porém, não possibilitam a avaliação do nível 'sintático'. Por fim, as propostas baseadas na comparação da ontologia a uma fonte de dados permitem a avaliação de três níveis de uma ontologia.

É válido frisar que o fato de um tipo de proposta cobrir todos os níveis não garante a

sua qualidade ou supremacia em relação aos outros tipos de propostas, visto que é necessário levar em conta outros critérios na avaliação de ontologias, por exemplo, os objetivos da avaliação e o que exatamente se pretende avaliar na ontologia. Portanto, se o objetivo é apenas avaliar o desempenho da ontologia na representação de conceitos, que se encaixa no nível 'léxico, vocabulário e conceito', qualquer um dos tipos de propostas podem ser utilizadas. Assim, para a escolha da proposta mais adequada levam-se em consideração os outros objetivos e possibilidade de aplicação de determinada proposta.

Encontram-se na literatura algumas críticas com relação a cada um dos tipos de propostas para avaliação de ontologias, que podem ajudar na escolha da proposta para avaliação de determinada ontologia. Com relação às propostas baseadas na comparação da ontologia a um *Gold Standard*, alguns pesquisadores (como HLOMANI; STACEY, 2013) dizem que é complexo estimar até que ponto o modelo é ideal, além de ser difícil encontrar um modelo que englobe o mesmo domínio da ontologia que está sendo avaliada. Ademais, quando os resultados da ontologia avaliada diferem dos resultados do *Gold Standard*, é difícil descobrir a origem do problema. No que se refere às propostas de avaliações baseadas em tarefas e aplicações, os críticos dizem que elas são muito específicas, pois testam a ontologia na realização de uma única tarefa. Portanto, a ontologia pode obter ótimos resultados em determinada tarefa, mas quando testada em outra, os resultados podem ser diferentes. Dessa forma, dependendo dos objetivos da avaliação, esse tipo de proposta não se aplica. Com relação às propostas de avaliação realizadas por humanos, as dificuldades são estabelecer quem são os usuários certos e quais os melhores critérios para a avaliação da ontologia. Além disso, o fato da avaliação depender da intuição do especialista faz com que se torne incerto a verificação do rigor da avaliação. No que tange às propostas de avaliação de ontologias baseadas na comparação a uma fonte de dados, Hlomani e Stacey (2014) apontam para o fato da maioria das propostas considerarem o conhecimento de determinado domínio como algo constante, ou seja, que não se altera, porém não é isso que acontece, o conhecimento de um domínio é algo dinâmico. Assim, se uma ontologia foi avaliada em comparação ao conhecimento de um domínio em determinada época, se essa mesma ontologia for avaliada em comparação ao mesmo domínio de conhecimento

em uma época diferente, os resultados da avaliação podem ser diferentes, posto que com o passar do tempo o domínio de conhecimento tenha sofrido alterações. Essa dinamicidade do conhecimento pode variar de domínio para domínio.

Este trabalho não pretendeu esgotar todos os assuntos relacionados à avaliação de ontologias. Existem muitos trabalhos que abordam a questão de vários pontos de vista. Muitas propostas, hipóteses e alguns experimentos são percebidos na literatura; porém, pouca coisa está consolidada. E como afirmam vários autores, entre eles Gómez-Pérez (2001) e Vrandečić (2009), a avaliação de ontologias ainda é um campo emergente, que precisa de um aprofundamento de ideias e de diretrizes.

Com base no levantamento feito na literatura, percebe-se que as diversas abordagens para avaliação de ontologias variam de acordo com o tipo e o propósito da ontologia que está sendo avaliada. A temática ainda está em aberto e são necessários estudos para o desenvolvimento do campo da avaliação de ontologias. Outras discussões sobre o assunto podem ser encontradas em Gruninger, Fox (1995); Corcho et al. (2004); Guarino, Welty (2002); Tartir et al. (2006); Yu, Thom, Tam (2007); e outros.

4 METODOLOGIA

Nesta seção, apresentam-se os procedimentos seguidos para realização da pesquisa. Está estruturada da seguinte forma: na subseção 4.1, caracterização da pesquisa, classifica-se a pesquisa de acordo com sua natureza, objetivo e abordagem; na subseção 4.2, método, apresenta-se o método escolhido, com base na revisão de literatura realizada sobre avaliação de ontologias; e por fim, a subseção 4.3 apresenta o percurso metodológico.

A aplicação dos procedimentos metodológicos usados para a avaliação da ontologia será apresentada no capítulo 5 – Descrição dos procedimentos e resultados. Com base em outros trabalhos que apresentam semelhanças com a presente pesquisa (MACULAN, 2015), optou-se por essa forma de apresentação dos procedimentos metodológicos. Acredita-se que essa estrutura de apresentação facilitará a leitura e o entendimento do caminho realizado para se chegar aos resultados da pesquisa.

4.1 Caracterização da pesquisa

A caracterização da presente pesquisa se baseia nos parâmetros para classificação de pesquisas apresentados no trabalho de Silveira e Córdova (2009). Logo, é caracterizada como de natureza aplicada por ter o objetivo de gerar conhecimento para aplicação prática direcionada a um problema específico, visto que os resultados da avaliação da ontologia podem apontar para a necessidade de possíveis mudanças na ontologia avaliada.

No que se refere aos objetivos, trata-se de uma pesquisa exploratória, pois tem o propósito de habituar-se ao problema e torná-lo transparente. A pesquisa visa identificar as possíveis falhas da ontologia na representação de conceitos. Quanto à forma de abordagem do problema, propõe-se uma pesquisa qualitativa. A pesquisa

caracteriza-se ainda como um estudo de caso do tipo instrumental, pois visa aplicar uma única proposta para avaliação de ontologias.

O universo desta pesquisa é a ontologia da Rede AgroHidro da EMBRAPA, denominada OntoAgroHidro. Como descrito na seção 2.5.4, a OntoAgroHidro foi desenvolvida dentro de um projeto da Rede Agro-Hidro, 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.

4.2 Método

Como apresentado no capítulo 3 (Avaliação de ontologias), ainda não existem métodos consolidados para avaliação de ontologias; o que existe são propostas construídas individualmente por cada grupo com base em objetivos e problemas bem específicos. Mesmo as metodologias para desenvolvimento de ontologias que apresentam uma seção sobre avaliação de ontologias não abordam um padrão de avaliação estabilizado. Dessa forma, esta pesquisa se baseia nas propostas relacionadas à avaliação de ontologias direcionada a uma fonte de dados.

As etapas para realização da pesquisa se baseiam em um trabalho desenvolvido por Brewster et al. (2004), em que os autores apresentam uma proposta para avaliação de ontologia com base na comparação a uma fonte de dados. (Os detalhes da proposta foram apresentados na seção 3.3 – Avaliações com base na comparação a uma fonte de dados). Porém, para sua utilização nesta pesquisa, foi necessário fazer adaptações, visto que a proposta de avaliação não apresenta detalhamentos para a sua execução. As adaptações dos principais passos da proposta de Brewster et al. (2004) para a presente pesquisa estão explicitadas no QUADRO 4.

QUADRO 4

Adaptações da proposta de avaliação de Brewster et al. (2004) para a proposta da presente pesquisa

Proposta de avaliação de Brewster et al. (2004)	Proposta de avaliação adaptada para esta pesquisa
..	1) Definição das questões de competência.
..	2) Seleção da coleção de documentos.
1) Identificação de palavras-chave/termos na coleção de textos.	3) Seleção das expressões relacionadas aos conceitos das QCs.
2) Expansão da consulta.	
3) Mapeamento dos termos identificados na ontologia.	4) Avaliação e inclusão das expressões na ontologia.
4) Comparação da ontologia com a coleção e contagem dos termos. ²²	

Fonte: Elaborado pelo autor.

Basicamente, as adaptações da proposta de avaliação apresentada por Brewster et al. (2004) estão relacionadas às duas primeiras etapas, que não existem na proposta dos autores. Outras adaptações são com relação à unificação das etapas 1 e 2 de Brewster et al. (2004) em uma só etapa (etapa 3) e das etapas 3 e 4 na etapa 4.

O motivo da escolha da proposta de Brewster et al. (2004) se deve ao fato de ser uma proposta com etapas simples 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 (como pode ser visto no Quadro 3), o que já é suficiente para o escopo desta pesquisa. Ademais, as críticas com relação aos tipos de propostas para avaliação de ontologias discutidas na seção 3.5 auxiliaram na escolha. Embora todos os quatro tipos de propostas para

²² A 4ª etapa não aparece de modo explícito na proposta de Brewster et al. (2004).

avaliação de ontologias tenham suas limitações, acredita-se que as propostas baseadas na comparação a uma fonte de dados podem garantir a qualidade da ontologia avaliada.

4.3 Percurso metodológico

Antes de abordar as etapas executadas para a avaliação da ontologia, cabe mencionar o levantamento bibliográfico realizado para o desenvolvimento da pesquisa. Foram pesquisados os seguintes temas: Organização e Representação do Conhecimento, Ontologias e Avaliação de Ontologias. Para isso, foram utilizadas grandes bases de dados específicas da BCI e outras multidisciplinares, como: *Annual Review of Information Science and Technology* (ARIST); Portal CAPES, *Google Scholar*, *Scielo*, *CiteSeer*, *Elsevier Science Direct*, Portal ACM, Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), Bibliotecas Digitais de Teses e Dissertações de diferentes instituições, a Biblioteca Professora Etelvina Lima, da ECI-UFMG e periódicos nacionais e internacionais.

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 FIG. 5 apresenta-se uma visão geral das etapas de avaliação e cada uma de suas subetapas. O detalhamento de cada uma das etapas encontra-se no capítulo 5 – Descrição dos procedimentos e resultados – no qual também são apresentados os

resultados obtidos em cada etapa, ao mesmo tempo em que são descritos os procedimentos metodológicos.

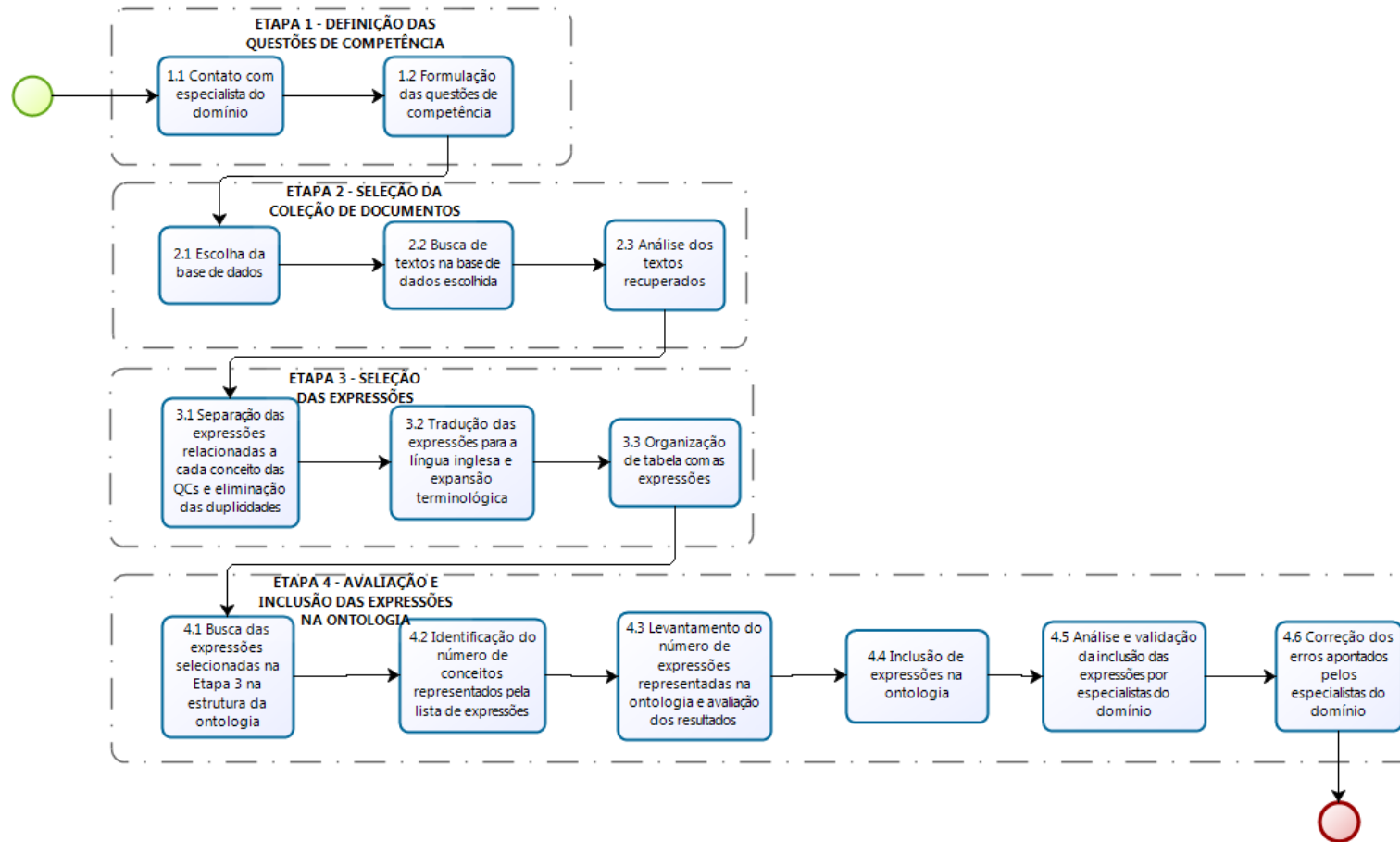


FIGURA 5 – Visão geral das etapas de avaliação da ontologia
 Fonte: Elaborada pelo autor.

5 DESCRIÇÃO DOS PROCEDIMENTOS E RESULTADOS

Neste capítulo, 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.

5.1 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.

(1) Contato com um especialista do domínio:

Contou-se com a contribuição de um especialista que compreendesse o domínio modelado pela OntoAgroHidro. Assim, teve-se o auxílio do Pesquisador Doutor da EMBRAPA, Ivo Pierozzi Júnior. A participação desse especialista do domínio nesta etapa foi primordial, visto que ele possuía o conhecimento necessário para a formulação das questões de competência.

(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:

- a) Os cultivos orgânicos ocasionam algum impacto na qualidade da água de nossas bacias?
- b) Faltará água nos reservatórios de abastecimento de água para a população paulista em 2016?
- c) 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.

Para facilitar a apresentação dos resultados nas tabelas, as questões de competência, acima mencionadas, serão apresentadas, no decorrer deste trabalho, da seguinte forma: “Questão 1” referir-se-á a: Os cultivos orgânicos ocasionam algum impacto na qualidade da água de nossas bacias?; “Questão 2”: Faltará 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?

5.2 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 – BDPA²³, 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:

Utilizando os conceitos principais presentes em cada questão de competência, realizaram-se as buscas na base de dados BDPA. Exemplo: Os **cultivos orgânicos** ocasionam algum impacto na **qualidade de água** de nossas **bacias**? As expressões²⁴ em negrito foram consideradas como os conceitos principais da questão de competência. O que estamos denominando de conceito principal são os processos, fenômenos ou entidades do mundo real, cuja forma verbal é uma unidade terminológica do domínio modelado e que deveria estar representada na estrutura da ontologia.

Foram realizadas buscas na BDPA utilizando os conceitos das questões de competência, individualmente. Os critérios empregados foram os seguintes: (1) utilização dos conceitos no singular; (2) utilização de aspas para os conceitos que continham mais de uma palavra (houve algumas exceções, devido às particularidades de cada conceito buscado). O resultado dessa pesquisa, discriminado por questão de competência, está detalhado na TAB.1 – Resultados da busca dos documentos na base BDPA.

²³ Disponível em: <https://www.bdpa.cnptia.embrapa.br/consulta/>

²⁴ Expressão: toda enunciação linguística (Dicionário de Biblioteconomia).

TABELA 1
Resultados da busca dos documentos na base BDPA

Questão	Conceito	Expressão de busca	Quantidade de registros recuperados	Observação
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.
	* ²⁵ População (população paulista)	#	0	Não foram extraídos documentos da BDPA.
	*Tempo (2016)	#	0	Não foram extraídos documentos da BDPA.
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 pelo autor.

²⁵ 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 e NAL. De onde foram extraídos os descritores relacionados aos dois conceitos.

(3) Análise dos textos recuperados:

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 (*Portable Document Format*).
- ii. Realização de leitura de um documento por vez, iniciando no primeiro registro recuperado que possuía o texto completo no formato PDF.
- 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.

A lista dos documentos analisados que foram usados para extração das expressões relacionadas aos conceitos das questões de competência encontra-se no APÊNDICE A.

A TAB. 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	Quantidade de docs. analisados
Questão 1	Qualidade de água	10
	Cultivo orgânico	8
	Bacia	6
Questão 2	Água	5
	Reservatório de abastecimento de água	7
	²⁶ População (população paulista)	0
	Tempo (2016)	0
Questão 3	Desmatamento	7
	Nascente	5
	Quantidade de água	6
	Rio	5

Fonte: Elaborado pelo autor.

Como pode ser observado na TAB. 2, a quantidade de documentos analisados para os conceitos “população” e “tempo” foi igual a zero, isso por serem conceitos genéricos, não sendo, portanto, buscados documentos relacionados a esses conceitos na base de dados BDPA. Assim, as expressões relacionadas aos dois conceitos foram selecionadas nos tesouros AGROVOC e NAL.

²⁶ 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 relacionados aos dois conceitos foram buscados nos tesouros AGROVOC e NAL. De onde foram extraídos os descritores relacionados aos dois conceitos.

5.3 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.

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 inglesa²⁷, utilizando o tradutor *online* Google Tradutor e os tesouros AGROVOC²⁸ e NAL²⁹.

²⁷ Foi necessário fazer essa tradução porque a OntoAgroHidro é bilíngue.

²⁸ Disponível em: <http://aims.fao.org/standards/agrovoc/functionalities/search>

²⁹ Disponível em: <http://agclass.nal.usda.gov/dne/search.shtml>

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 inglês³⁰ 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 resultado dessa etapa foi representado na Tabela do APÊNDICE B contendo todas as expressões extraídas da coleção de documentos e os descritores selecionados nos tesouros, com os seus correspondentes no idioma inglês.

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.

5.4 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

³⁰ Foi 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.

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 e com o uso do recurso “Ctrl + F” buscou-se uma expressão de cada vez (Tabela do APÊNDICE B) na estrutura da ontologia. Além 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. O APÊNDICE C – Representação das expressões pela OntoAgroHidro – apresenta as expressões representadas e não representadas pela 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, por meio de análise da tabela do APÊNDICE C. 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 TAB. 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 (percebe-se que a quantidade de conceitos é menor, pois como foi descrito anteriormente, um conceito pode ser representado por várias expressões diferentes); 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	Número de expressões ³¹	Número de conceitos repres. pelas expressões ³²	Número 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 pelo autor.

Percebeu-se, com base nos dados apresentados na TAB. 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”, pois, como pode ser constatado no APÊNDICE C, 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 TAB. 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.

³¹ O número de expressões por questão de competência foi extraído do Apêndice B.

³² O número de conceitos se baseia em uma análise de quantos conceitos estão representados na lista de expressões (Apêndice B).

TABELA 4

Representatividade da OntoAgroHidro discriminada por cada conceito das questões de competência

Questão	Conceito buscado	Nº de expressões	N.º de conceitos repres. pelas expressões	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%
Questão 3	Desmatamento	10	1	0	0%
	Nascente	5	1	0	0%
	Quantidade de água	4	1	1	100%
	Rio	5	4	2	50%

Fonte: Elaborada pelo autor.

Em uma abordagem analítica da TAB. 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 (descrito na seção 2.5.4). 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. Portanto, a representatividade da OntoAgroHidro no que tange aos conceitos relacionados à água (recurso hídrico), mostra a necessidade de melhorias na representação de alguns conceitos para que se possa atingir a representatividade total da amostra selecionada nesta pesquisa.

Também, pode-se notar que a TAB. 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.

Ainda em análise da TAB. 4, constata-se que os conceitos “cultivo orgânico”, “população”, “desmatamento” e “nascente” não tiveram representatividade (0%). Isso demonstra que não havia na estrutura da ontologia nenhuma expressão que representasse esses conceitos. O conceito “cultivo orgânico”, que trata de um tipo de sistema de cultivo em que não se utilizam agrotóxicos, por exemplo, deveria aparecer como uma subclasse de “sistemas de cultivo”, como pode ser observado na FIG. 6, que apresenta um recorte da estrutura de classes em que o conceito “cultivo orgânico” deveria estar representado.



FIGURA 6 – Recorte da estrutura de classes em que o conceito cultivo orgânico deveria estar representado.

Fonte: Captura de tela do software Protégé.

O conceito “população”, que, nesse caso específico, refere-se à população humana, deveria estar representado dentro da classe “Dimension”. A criação de uma subclasse (HumanDimension) dentro da classe “Dimension” e a inclusão do conceito de população dentro dessa subclasse poderia tornar a representação mais clara. Na FIG. 7 apresenta-se a estrutura de classes em que o conceito poderia ser inserido.

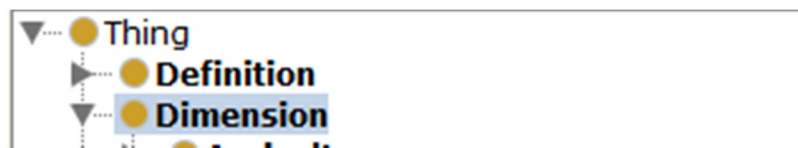


FIGURA 7 – Recorte da estrutura de classes em que o conceito população deveria estar representado.

Fonte: Captura de tela do software Protégé.

Com relação ao conceito “desmatamento”, que é um agente de transformação ambiental causado pela ação do homem, portanto, antrópico, deveria estar representado na subclasse “AnthropicEnvironmentalTransformation”. A FIG. 8 apresenta a estrutura de classes em que o conceito desmatamento deveria estar representado.



FIGURA 8 – Recorte da estrutura de classes em que o conceito desmatamento deveria estar representado.

Fonte: Captura de tela do software Protégé.

Por fim, o conceito “nascente”, que se trata do local onde emerge água subterrânea naturalmente, seria representado na ontologia dentro da subclasse “Água Superficial”, além de ser também um corpo de água; portanto poderia estar representado como

subclasse de “BodyOfWater”. Na FIG. 9, apresenta-se o recorte da estrutura de classes em que o conceito “nascente” seria representado.

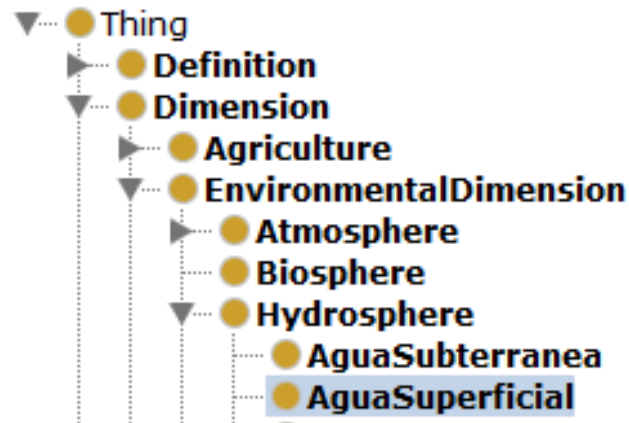


FIGURA 9 – Recorte da estrutura de classes em que o conceito nascente deveria estar representado.

Fonte: Captura de tela do software Protégé.

(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 agricultura³³ (como glossários, artigos, etc.), incluíram-se na estrutura da ontologia, as expressões consideradas relevantes pelo autor desta dissertação, 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

³³ Essas fontes de informação serviram de guia para a inclusão das expressões da OntoAgroHidro.

em um 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 definições³⁴. Algumas definições foram retiradas da base lexical Onto.PT³⁵. 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 também 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*).

³⁴ 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.

³⁵ 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>

(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. O Relatório das expressões incluídas na estrutura da OntoAgroHidro (APÊNDICE D) apresenta todas as expressões incluídas e seus respectivos rótulos, todos devidamente validados pelos especialistas do domínio.

As principais correções apontadas pelos especialistas do domínio referiam-se às expressões que demandavam um conhecimento mais aprofundado do domínio, que não é o domínio de especialidade do autor desta dissertação. Entre essas expressões, estão os parâmetros para medir a qualidade de água, tais como: cloreto dissolvido, oxigênio consumido, sólidos dissolvidos voláteis, sólidos em suspensão e sólidos totais. A grande dificuldade foi entender que algumas dessas expressões, além de serem incluídas como classes ou subclasses na ontologia, também deveriam ser incluídas como propriedade de dados, pois no caso específico da OntoAgroHidro, elas representam quantidades (um número) que podem atestar o nível de qualidade da água.

5.5 Considerações do capítulo

Este capítulo apresentou os procedimentos metodológicos seguidos para a realização da presente pesquisa e os resultados gerados com base na aplicação de cada um dos procedimentos. A proposta de avaliação da OntoAgroHidro apresentada neste trabalho perpassou por quatro etapas, subdivididas em 14 procedimentos, como pode ser observado na FIG. 5 – Visão geral das etapas de avaliação da ontologia – do capítulo 4, que apresenta a proposta de modo esquematizado.

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. Essa 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 podem contribuir no enriquecimento terminológico e conceitual da OntoAgroHidro e eliminar essa falta de conceitos específicos do domínio.

6 CONSIDERAÇÕES FINAIS

As ontologias permitem criar um modelo formal de mundo, que pode facilitar o entendimento de como as coisas estão organizadas, como elas se relacionam e funcionam. Dessa forma, possuem várias finalidades, podem ser usadas para facilitar a comunicação entre especialistas de um domínio específico, uma vez que permitem a criação de um vocabulário comum do domínio. As ontologias podem ser utilizadas para interoperabilidade entre diferentes sistemas, criando uma ponte entre esses sistemas e fazendo com que eles compartilhem o mesmo vocabulário. Ademais, elas podem ser aplicadas no contexto da recuperação da informação, possibilitando que as buscas possuam mais sentido. Mas, para que possa realizar as atividades descritas e várias outras as quais elas tenham o potencial para desempenhar, é necessário garantir a eficiência delas.

A presente 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. A avaliação de ontologias tem o propósito de comprovar a conformidade do modelo de mundo com o mundo modelado formalmente na ontologia.

Para tanto, baseou-se em uma proposta de avaliação de ontologias que visa à comparação entre a ontologia que está sendo avaliada e uma coleção de documentos textuais sobre o domínio modelado pela ontologia. Assim, adaptou-se a proposta apresentada por Brewster et al. (2004) e avaliou-se a OntoAgroHidro, uma ontologia de domínio da EMBRAPA.

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 seguem 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.

A proposta de avaliação aplicada neste trabalho se mostrou útil para a avaliação da OntoAgroHidro, pois permitiu atingir todos os objetivos propostos. No que concerne ao objetivo geral, que foi avaliar a OntoAgroHidro com base em uma proposta de avaliação ancorada na comparação a um corpus de texto, foi possível aplicar, mediante adaptações, uma proposta existente na literatura que se baseia na comparação a um corpus de textos. Os objetivos específicos também foram executados, com relação ao primeiro, avaliar o desempenho da OntoAgroHidro na representação de conceitos do domínio modelado com base em questões de competência; realizou-se uma contagem de quantos conceitos do corpus a ontologia representava e calculou-se a porcentagem. Para a execução do segundo objetivo, que se tratou de sugerir mudanças na OntoAgroHidro para melhorar a modelagem do domínio, a partir dos resultados da avaliação; apontaram-se algumas falhas na representação de conceitos da ontologia e recomendou-se a inclusão de alguns conceitos na estrutura da OntoAgroHidro. Por fim, para a realização do terceiro objetivo: contribuir para o enriquecimento conceitual e terminológico da ontologia, foram atribuídas anotações às classes da ontologia, sugerindo a inclusão de mais de um termo como rótulo para a representação dos conceitos, além da inclusão de novos conceitos na estrutura da OntoAgroHidro.

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. A extração manual dos termos dos documentos também foi uma dificuldade, pois é um processo demorado e que exige uma leitura atenta.

Por fim, a pesquisa apresentou uma proposta de avaliação para uma ontologia dos domínios da Agricultura e Recursos Hídricos, desenvolvida por uma Empresa Brasileira. Assim, 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.

6.1 Limitações da pesquisa

Este trabalho se restringiu a aplicar apenas uma proposta de avaliação de ontologias em uma ontologia de domínio da EMBRAPA. Isso fornece uma perspectiva muito específica da avaliação, ou seja, por ser um estudo de caso, a generalização dos resultados pode não ser adequada, embora a pesquisa apresente um esquema de avaliação que pode vir a ser aplicado em outros contextos. Além disso, algumas variáveis podem interferir na avaliação da ontologia, como o tamanho da ontologia avaliada e o domínio que a ontologia representa.

Dentro dos objetivos da pesquisa, não foi possível avaliar as relações, hierarquia e outras dimensões (níveis) da ontologia. A avaliação focou apenas no nível dos conceitos, por acreditar que essas unidades do conhecimento são uma das partes principais em um sistema de organização do conhecimento, como as ontologias.

A pesquisa também se restringiu a um número de conceitos para avaliação da ontologia, devido ao fato de ser uma avaliação quase em sua totalidade manual/intelectual. Dessa forma, não foi possível expandir o escopo de conceitos da ontologia para avaliação. Porém, uma abordagem totalmente automática poderá expandir o escopo conceitual e possibilitar uma avaliação mais ampla e de toda a estrutura da ontologia.

6.2 Trabalhos futuros

No decorrer desta pesquisa, foram identificadas direções que podem culminar em trabalhos futuros. Uma dessas direções é a expansão do escopo conceitual para avaliação da ontologia. Acredita-se que a execução da metodologia proposta com um grande número de conceitos da ontologia ou talvez a ontologia por completo, possa trazer resultados mais significativos, possibilitando um diagnóstico mais específico da OntoAgroHidro. Outra direção que pode ser explorada em uma pesquisa futura é a extração automática de expressões em documentos textuais do domínio e a comparação dessas expressões com a ontologia avaliada e com outras ontologias do domínio.

Em suma, a presente pesquisa abre um leque de opções para pesquisas futuras, tais como: comparação entre duas ou mais propostas de avaliação e medição do desempenho da ontologia em cada uma delas; utilização de bases lexicais, tais como a Onto.PT para enriquecimento terminológico da OntoAgroHidro; comparação entre as duas versões da OntoAgroHidro, ou seja, a ontologia antes da inclusão das expressões e após a inclusão das expressões – pode-se comparar o desempenho na recuperação da informação; avaliação de outras dimensões (níveis) da OntoAgroHidro, como relacionamentos, axiomas e estrutura.

REFERÊNCIAS

- ALMEIDA, Maurício Barcellos; BAX, Marcello Peixoto. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, Brasília, DF, v. 32, n. 3, p. 7-20, set./dez. 2003.
- ALMEIDA, Mauricio Barcellos. A proposal to evaluate ontology content. *Applied Ontology*, Birmingham, v. 4, n. 3-4, p. 245-265, 2009. Disponível em: <<http://content.iospress.com/articles/applied-ontology/ao070>>. Acesso em: 05 abr. 2014.
- ALVARENGA, Lídia. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 8, n. 15, p. 18-40, 2003. Disponível em: <10.5007/1518-2924.2003v8n15p18>. Acesso em: 05 abr. 2014.
- AMERICAN NATIONAL INFORMATION STANDARDS ORGANIZATION. *ANSI/NISO Z39.19 : Guidelines for the construction, format and management of monolingual thesauri*. Bethesda, MD: Niso Press, 2003.
- ARANALDE, Michel Maya. Reflexões sobre os sistemas categoriais de Aristóteles, Kant e Ranganathan. *Ciência da Informação*, Brasília, DF, v. 38, n. 1, p. 86-108, jan./abr. 2009. Disponível em: <<http://dx.doi.org/10.1590/S0100-19652009000100006>>. Acesso em: 05 abr. 2014.
- ARANHA, Maria Lucia de Arruda; MARTINS, Maria Helena Pires. *Filosofando: introdução a filosofia*. 2. ed. São Paulo: Moderna, 1993. 395 p.
- BONACIN, Rodrigo; NABUCO, Olga Fernanda; PIEROZZI, Ivo. Modeling the Impacts of Agriculture on Water Resources: Semantic Interoperability Issues. In: 23rd INTERNATIONAL WORKSHOPS ON ENABLING TECHNOLOGIES: INFRASTRUCTURES FOR COLLABORATIVE ENTERPRISE (WETICE), 23., 2014 Parma. [*Proceedings*]... Parma: IEEE, 2014. p. 447-452. Disponível em: <10.1109/WETICE.2014.17>. Acesso em: 22 jan. 2015.
- BONACIN, R.; NABUCO, O. F.; PIEROZZI JÚNIOR, I. Ontology models of the impacts of agriculture and climate changes on water resources: Scenarios on interoperability and information recovery. *Future Generation Computer Systems*, [S.l.], v. 54, p. 423-434, jan. 2016. 2015. Disponível em: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/137810/1/Ontology-Bonacin.pdf>>.

Acesso em: 25 maio 2015.

BONACIN, Rodrigo; NABUCO, Olga Fernanda; PIEROZZI JÚNIOR, Ivo. In: III SEMINÁRIO DA REDE AGROHIDRO E I WORKSHOP DO PROJETO “OS IMPACTOS DA AGRICULTURA E DAS MUDANÇAS CLIMÁTICAS NOS RECURSOS HÍDRICOS” ,3., 2015, Brasília, DF. *Resumo dos trabalhos apresentados...*, Brasília, DF: EMBRAPA, 2015. p. 56. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/134803/1/Agrohidro-representacao-Bonacin.pdf>>. Acesso em: 10 dez. 2015.

BRANK, J.; GROBELNIK, M.; MLADENIĆ, D. A survey of ontology evaluation techniques. In: PROCEEDINGS OF THE CONFERENCE ON DATA MINING AND DATA WAREHOUSES, Ljubljana, Sloveni, 2006, [*Proceedings...*]. Ljubljana, Sloveni, 2006. Disponível em: <http://lorcancoyle.org/files/Ye2007Ontology-.pdf>>. Acesso em: 03 set. 2014.

BRÄSCHER, Marisa; CAFÉ, Lígia. Organização da informação ou organização do conhecimento. In: IX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. *Anais...* São Paulo: ECA/USP, 2008. p. 1-14.

BREWSTER, C. *et al.* Data driven ontology evaluation. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 2004, Lisboa. *Anais...* Lisboa: [s.n], 2004. Disponível em: <http://www.cbrewster.com/papers/BrewsterLREC.pdf> . Acesso em: 22 fev. 2014.

BURKE, Peter. *Uma história social do conhecimento I: de Gutenberg a Diderot*. Rio de Janeiro: Jorge Zahar, 2003. 241 p.

CABRÉ, M. T. *La terminología: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Linguística Aplicada, 1999.

CARLAN, Eliana. *Sistemas de organização do conhecimento: uma reflexão no contexto da ciência da informação*. 2010. 195 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, DF, 2010.

CHANDRASEKARAN, Balakrishnan; JOSEPHSON, John R.; BENJAMINS, V. Richard. What are ontologies, and why do we need them?. *IEEE Intelligent Systems*, Piscataway, New Jersey, n. 1, p. 20-26, 1999.

CHOI, JooYoung *et al.* A semantic web ontology for small molecules and their biological targets. *Journal of Chemical Information and Modeling*, Brisbane, Australia, v. 50, n. 5, p. 732-741, 2010.

CLARKE, Erik L. *et al.* A task-based approach for Gene Ontology evaluation. *Journal of Biomedical Semantics*, Londres, v. 4, n. 1, p. 1, 2013. Disponível em: <<http://jbiomedsem.biomedcentral.com/articles/10.1186/2041-1480-4-S1-S4>>. Acesso em: 10 jan. 2015.

CORCHO, Óscar *et al.* ODEval: a tool for evaluating RDF (S), DAML+ OIL, and OWL concept taxonomies. In: ARTIFICIAL INTELLIGENCE APPLICATIONS AND INNOVATIONS. 2004, Springer, USA. *Proceedings...* Springer, USA: [s.n], 2004. p. 369-382.

CORRÊA, Paulo José Melo Gomes. *Uma ontologia para representação do conhecimento do domínio da química analítica com adição de novos agentes e funcionalidades para análise e monitoramento de combustíveis*. 2009. 128 f. Dissertação (Mestrado em Engenharia de Eletricidade) - Centro de Ciências Exatas e Tecnologia, Universidade Federal do Maranhão, São Luís, 2009.

CUNHA, Murilo Bastos da.; CAVALCANTI, Cordélia R. *Dicionário de biblioteconomia e arquivologia*. Brasília: Brique de Lemos/Livros, 2008. 451 p.

DAHLBERG, I. Teoria do conceito. *Ciência da Informação*, Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978.

DAHLBERG, I. Knowledge organization: A new science. *Knowledge organization*, Frankfurt, v. 33, n. 1, p. 11-19, 2006. Disponível em: <http://www.db.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_Dahlberg.htm>. Acesso em: 03 abr. 2014.

DELLSCHAFT, K.; STAAB, S. On How to Perform a Gold Standard Based Evaluation of Ontology Learning. In: PROC. FIFTH INT'L CONF. SEMANTIC WEB, 2006, Heidelberg. *Anais...* Heidelberg: [s.n], 2006. p. 228-241. Disponível em: <<http://www.socialdynamics.it/tagora/wp-content/2007/05/Dellschaft2006fy.pdf>>. Acesso em: 03 abr. 2014.

ELHADAD, D. GABAY, M.; NETZER, Y. Automatic evaluation of search ontologies in

the entertainment domain using text classification. In: _____. *Applied Semantic Technologies: Using Semantics in Intelligent Information Processing*. [S.l.]: Taylor and Francis, 2010. Disponível em: <<https://www.cs.bgu.ac.il/~elhadad/papers/OntologyEvaluation.pdf>>. Acesso em: 10 jun. 2014.

FERNÁNDEZ, M.; GOMEZ-PÉREZ, A.; JURISTO, H. *Methontology: from ontological art towards ontological engineering*. 1997. Disponível em: <http://oa.upm.es/5484/1/METHONTOLOGY_.pdf>. Acesso em: 04 ago. 2014

FERNÁNDEZ, M.; CANTADOR, Iván; CASTELLS, Pablo. CORE: A tool for collaborative ontology reuse and evaluation. In: 4TH INTERNATIONAL WORKSHOP ON EVALUATION OF ONTOLOGIES FOR THE WEB, 4., 2006, Edinburgh, UK. *Proceedings...* Edinburgh, UK, 2006. Disponível em: <<http://oro.open.ac.uk/28592/1/eon06.pdf>>. Acesso em: 03 abr. 2014.

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: 10 jun. 2014.

FERNÁNDEZ-LÓPEZ, M. *et al.* Building a chemical ontology using methodology and the ontology design environment, *Intelligent Systems & their Applications*, [S.l.], v. 14, n.1, p.37-46, jan./fev. 1999.

FERREIRA, Aurélio Buarque de Holanda. *Míni Aurélio: o dicionário da língua portuguesa*. Rio de Janeiro: Positivo, 2010.

FURGERI, Sérgio. *Representação de informação e conhecimento: estudo das diferentes abordagens entre Ciência da Informação e a Ciência da Computação*. 2006. 161 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Ciência da Informação, Pontifícia Universidade Católica de Campinas, Campinas, 2006.

GANGEMI, A. *et al.* A metaontology-based framework for ontology evaluation and selection. In: 4TH INTERNATIONAL WORKSHOP ON EVALUATION OF ONTOLOGIES FOR THE WEB, 4., Edinburgh, UK, 2006, Edinburgh, UK. *Proceedings...* Edinburgh, UK, 2006. Disponível em: <<http://ceur-ws.org/Vol->

179/eon2006gangemietal.pdf>. Acesso em: 03 abr. 2014.

GILCHRIST, Alan. Thesauri, taxonomies and ontologies: an etymological note. *Journal of Documentation*, [S.l.], v. 59, n. 1, p. 7-18, 2003.

GOMES, Hagar Espanha. *Manual de elaboração de tesouros monolíngues*. Brasília, DF: CNPq/PNBU, 1990. 78 p.

GÒMEZ-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: 03 abr. 2014.

GOMÉZ-PÉREZ, A. PAZOS, Juristo N, J. Evaluation and assessment of the knowledge sharing technology. In: MARS, N. J. L. *Towards very large knowledge bases: knowledge building & knowledge sharing*. Amsterdam: IOS, 1995. p. 289-296. Disponível em: <https://www.researchgate.net/publication/243773900_Evaluation_and_Assessment_of_the_Knowledge_Sharing_Technology>. Acesso em: 03 abr. 2014.

GÓMEZ-PÉREZ, A.; FERNÁNDEZ, M.; VICENTE, A.J. de. Towards a method to conceptualize domain ontologies. In: 12TH EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE (ECAI'96), 12., 1996, Budapest, Romênia. *Proceedings...* Budapest, Romênia: Facultad de Informática (UPM), 1996. Disponível em: <http://www.academia.edu/13104696/Towards_a_Method_to_Conceptualize_Domain_Ontologies>. Acesso em: 12 ago. 2014.

GRUBER, Thomas R. Toward principles for the Design of Ontologies Used for Knowledge Sharing. In: FORMAL ONTOLOGY IN CONCEPTUAL ANALYSIS AND KNOWLEDGE REPRESENTATION, 1993, Padova. Italy. *Proceedings...* Padova. Italy: [s.n], 1993.

GRUNINGER, M.; FOX, M. S. Methodolgy for the design and evaluation of ontologies. In: PROCEEDINGS OF THE IJCAI WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWELDGE SHARING, Toronto. *Proceedings...* Montreal, Canadá: [s.n], 1995. Disponível em: <<http://stl.mie.utoronto.ca/publications/gruninger-ijcai95.pdf>>. Acesso em: 12 ago. 2014.

GUARINO, N. (Ed.). Formal ontology in information systems. In: PROCEEDINGS OF THE FIRST INTERNATIONAL CONFERENCE (FOIS'98), 1998, Trento, Italia. *Proceedings...* Trento, Italia: IOS Press, 1998. p. 3-15.

GUARINO, Nicola; WELTY, Christopher. Evaluating ontological decisions with OntoClean. *Communications of the ACM*, [S.l.], v. 45, n. 2, 2002. p. 61-65. Disponível em: <<http://www.loa.istc.cnr.it/old/Papers/CACM2002.pdf>>. Acesso em: 10 abr. 2014.

HJORLAND, B. Knowledge Organization Systems.[S.l]: [s.n], 2007. Disponível em:<http://www.db.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_systems.htm> . Acesso em: 10 set. 2014.

HLOMANI, Hlomani; STACEY, D. A. An Extension to the Data-driven Ontology Evaluation. In: 15TH INTERNATIONAL CONFERENCE ON INFORMATION REUSE AND INTEGRATION (IRI), 15., 2014, Ontário, Canadá. *Proceedings...* Ontário, Canadá: IEEE, 2014.

HLOMANI, Hlomani; STACEY, D. A. Contributing evidence to data-driven ontology evaluation: Workflow ontologies perspective. In: 5TH INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND ONTOLOGY DEVELOPMENT, 5., Vilamoura, Portugal. *Proceedings...* Vilamoura, Portugal: SCITEPRESS, 2013. p. 207-213.

HODGE, G. *Systems of Knowledge Organization for Digital Libraries: beyond traditional authority files*. Washington, DC: The Digital Library Federation, 2000.

INTERNATIONAL STANDARD ORGANIZATION ISO. *ISO 1087-1: Terminology work: vocabulary: Part 1: theory and application*. Genova, 2000. 42 p.

JAPIASSU, Hilton; MARCONDES, Danilo. *Dicionário básico de filosofia*. 5. ed. Rio de Janeiro: J. Zahar, 2008. 309 p.

JONES, D.; BENCH-CAPON, T.; VISSER, P. Methodologies for ontology development. 1998. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.52.2437&rep=rep1&type=pdf>>. Acesso em: 10 set. 2014.

LOZANO-TELLO, A.; GÓMEZ-PÉREZ, A. Ontometric: A method to choose the appropriate ontology. *Journal of Database Management*, Hershey, Pensilvânia, v.

15, n. 2, 18 p. 2004. Disponível em: <<http://www.igi-global.com/article/ontometric-method-choose-appropriate-ontology/3308>>. Acesso em: 10 set. 2014.

MACULAN, B. C. M. S. *Estudo e aplicação de metodologia para reengenharia de tesouro: remodelagem do THESAGRO*. 345f. 2015. Tese de Doutorado. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.

MAEDCHE, A.; STAAB, S. Measuring similarity between ontologies. In: EUROPEAN CONFERENCE ON KNOWLEDGE ACQUISITION AND MANAGEMENT, 2002, Madri, Espanha. *Proceedings...* Madri, Espanha: [s.n], 2002. Disponível em: <http://disi.unitn.it/~accord/RelatedWork/Matching/Staab2002.pdf>. Acesso em: 20 set. 2014.

MAEDCHE, Alexander. *Ontology learning for the semantic web*. Springer: Science & Business Media, 2012.

MIRANDA, Marcos Luiz Cavalcanti de. *Organização e representação do conhecimento: fundamentos teórico-metodológicos na busca e recuperação da informação em ambientes virtuais*. 2005. 353 f. Tese (Doutorado em Ciência da Informação) - Universidade Federal do Rio de Janeiro; Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2005.

MOREIRA, Alexandra; ALVARENGA, Lídia; OLIVEIRA, Alcione de Paiva. O nível do conhecimento e os instrumentos de representação: tesouros e ontologias. *DataGramaZero: Revista de Ciência da Informação*, Rio de Janeiro, v. 5, n. 6, 2004. Disponível em: <http://www.dgz.org.br/dez04/Art_01.htm>. Acesso em: 02 jun. 2014.

MYPOULOS, J. An overview of knowledge representation. In: WORKSHOP ON DATA ABSTRACTION, DATABASES AND CONCEPTUAL MODELLING. 1980, New York, NY. *Proceedings...* New York, NY: ACM, 1981. p. 5-12.

NOY, N. F.; MCGUINNESS, D. L. *Ontology Development 101: A Guide to Create Your First Ontology*. 2001. Disponível em: <http://protege.stanford.edu/publications/ontology_development/ontology101.pdf> . Acesso em: 11 set. 2014.

NOY, N. F., HAFNER C.D. The state of the art in ontology design: a survey and comparative review. *AI Magazine*, Palo Alto, CA, v. 18, n. 3, 1997. Disponível em: <<http://www.aaai.org/ojs/index.php/aimagazine/article/view/1306/1207>>. Acesso em:

02 jun. 2014.

ORME, A.; YAO, H.; ETZKORN, L. Indicating ontology data quality, stability, and completeness throughout ontology evolution. *Journal of Software Maintenance and Evolution: Research and Practice*, Orme, TN, v. 19, n. 1, p. 49-55, Jan./Feb. 2007. Disponível em: <<http://dx.doi.org/10.1002/smr.341>>. Acesso em: 20 jan. 2015.

OUYANG, Liubo et al. A method of ontology evaluation based on coverage, cohesion and coupling. In: EIGHTH INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS AND KNOWLEDGE DISCOVERY. 8., 2011, Shanghai. *Proceedings...* Shanghai: IEEE, 2011. p. 2451-2455.

PATEL, Chintan et al. OntoKhoj: A semantic web portal for ontology searching, ranking and classification. In: INTERNATIONAL WORKSHOP ON WEB INFORMATION AND DATA MANAGEMENT. 5., 2003, Chintan. *Proceedings...* Chintan: ACM, 2003. p. 58–61.

PIEADADE, M. A. R. *Introdução à teoria da classificação*. 2. ed. Rio de Janeiro: Interciência, 1983.

PORZEL, R.; MALAKA, R. A task-based approach for ontology evaluation. In: 16TH 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: 10 set. 2014.

PRESUTTI, Valentina *et al.* Pattern-based ontology design. In: _____. *Ontology Engineering in a Networked World*. Springer; Berlin; Heidelberg: Springer, 2012. cap. 3. p. 35-64.

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: 27 ago. 2015.

REDE AGROHIDRO. 2015. Disponível em: <<https://www.agropediabrasilis.cnptia.embrapa.br/web/agrohidro-mp1>>. Acesso em: 02 fev. 2015.

REN, Y. *et al.* Towards competency-question-driven ontology authoring. In: 11TH EXTENDED SEMANTIC WEB CONFERENCE (ESWC). 11., 2014, Creta, Grécia. *Proceedings...* Creta, Grécia: Springer, 2014. Disponível em: <http://2014.eswc-conferences.org/sites/default/files/papers/paper_145.pdf>. Acesso em: 10 dez. 2015.

SABOU, M. *et al.* Ontology selection: Ontology evaluation on the real semantic web. In: 4TH INTERNATIONAL WORKSHOP ON EVALUATION OF ONTOLOGIES FOR THEWEB 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: 10 set. 2014.

SALES, Rodrigo de. *Tesaurus e ontologias sob a luz da teoria comunicativa da terminologia*. 2008. 164 f. Dissertação (Mestrado em Ciência da Informação) - Centro de Ciências da Educação, Universidade Federal de Santa Catarina, Florianópolis, 2008. Disponível em: <<http://repositorio.ufsc.br/xmlui/handle/123456789/92173>>. Acesso em: 20 jan. 2015.

SALES, Rodrigo de; CAFÉ, Lígia. Diferenças entre Tesaurus e Ontologias. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 14, n. 1, p. 99-116, jan./abr. 2009. Disponível em: <<http://dx.doi.org/10.1590/S1413-99362009000100008>>. Acesso em: 20 jan. 2015.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>>. Acesso em: 20 jan. 2015.

SARACEVIC, T. Information Science. *Journal of the American Society for Information Science*, [Sl.] v. 50, n. 12, p. 1051-1063, 1999. Disponível em: <[http://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-4571\(1999\)50:12%3C1051::AID-ASI2%3E3.0.CO;2-Z/abstract](http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(1999)50:12%3C1051::AID-ASI2%3E3.0.CO;2-Z/abstract)>. Acesso em: 20 jan. 2015.

SILVEIRA, D. T.; CÓRDOVA, F. P. A pesquisa científica. In: GERHARDT, T. E.; SILVEIRA, D. T. *Métodos de pesquisa*. Porto Alegre: UFRGS, 2009. p. 31-42.

SMITH, B. Ontology and informations systems. 2004. Disponível em: <http://ontology.buffalo.edu/smith/articles/ontology_pic.pdf>

. Acesso em: 15 set. 2014.

SMITH, Barry. Beyond concepts: ontology as a reality representation. Proceedings of FOIS 2004. In: 3RD INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS. 3., 2004, Torino, Italia. *Proceedings...* Torino, Italia: IOS Press, 2004.

SOERTEL, Dagobert. The rise of ontologies or the reinvention of classification. *Journal of the American Society of Information Science*, [S.l.], v. 50, n. 12, p. 1119-1120, Oct. 1999. Disponível em: <<http://www.dsoertel.com/cv/B70.pdf>>
. Acesso em: 15 set. 2014.

SOUZA, R. R.; ALVARENGA, L. A web semântica e suas contribuições para a ciência da informação. *Ciência da Informação*, Brasília, DF, v. 33, n. 1, p. 132-141, jan./abr. 2004. Disponível em: <<http://dx.doi.org/10.1590/S0100-19652004000100016>>. Acesso em: 15 set. 2014.

STRASUNSKAS, D.; TOMASSEN, S. L.; Empirical insights on a value of ontology quality in ontology-driven web search. In: MEERSMAN, R.; TARI, Z. (Eds.). *Lecture notes in Computer Science*. Berlin: Springer, 2008. v. 2, cap. 2, p. 1319-1337.

STUDER, R.; BENJAMINS, V.; FENSEL, D. Knowledge Engineering: Principles and Methods. *Data and Knowledge Engineering*, [S.l.] v. 25, n. 1-2, p. 161-197, 1998.

SUÁREZ-FIGUEROA, M. C.; PRADEL, C.; HERNANDEZ, N. Verifying ontology requirements with SWIP. In: 18TH INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT (EKAW) 18., 2012, Galway, Irlanda. *Proceedings...* Galway, Irlanda: [s. n], 2012.

TARTIR, S. *et al.* Ontoqa: Metric-based ontology quality analysis. In: WORKSHOP ON KNOWLEDGE ACQUISITION FROM DISTRIBUTED, AUTONOMOUS, SEMANTICALLY HETEROGENEOUS DATA AND KNOWLEDGE SOURCES. 2005, Houston, USA. *Proceedings...* Houston, USA: IEEE, 2006. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.7909&rep=rep1&type=pdf>>. Acesso em: 14 ago. 2014.

TRISTÃO, A. M. D.; Fachin, R. B. G.; Alarcon, O. E. Sistemas de classificação facetada e tesouros: instrumentos para a organização do conhecimento. *Ciência da Informação*, Brasília, DF, v. 33, n. 2, p. 161-171, maio/ago. 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n2/a17v33n2.pdf>>. Acesso em: 20 jun. 2015.

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://journals.cambridge.org/action/displayAbstract;jsessionid=273174E2C8795929808D17E3A1D7C27C.journals?fromPage=online&aid=4071848>>. Acesso em: 20 jun. 2015.

VICKERY, B. C. Faceted classification for the web. *Axiomathes*, Houten, Holanda, v. 18, n. 2, p. 145-160, 2008. Disponível em:

<<http://link.springer.com/article/10.1007%2Fs10516-007-9025-9>>. Acesso em: 20 jun. 2015.

VICKERY, B. C. Knowledge representation: a brief review. *Journal Documentation*, London, v. 42, n.3, p.145-159, Sept. 1986. Disponível em:

<<http://www.emeraldinsight.com/doi/abs/10.1108/eb026790>>. Acesso em: 20 jan. 2015.

VICKERY, B. C. Ontologies. *Journal of Information Science*, [S.l.], v. 23, n. 4, p. 272-286, 1997.

VRANDECIC, Denny. Ontology evaluation. In: STAAB, S.; STUDER, R. *Handbook on Ontologies*. Springer: Springer, 2009. p. 293-313. Disponível

em:<http://link.springer.com/chapter/10.1007/978-3-540-92673-3_13>. Acesso em: 12 ago. 2014.

WORLD WIDE WEB CONSORTIUM – W3C. OWL 2 Web Ontology Language Document Overview. 2012. Disponível em: <<https://www.w3.org/TR/owl2-overview/>>. Acesso em: 10 dez. 2015.

WORLD WIDE WEB CONSORTIUM – W3C. SKOS Simple Knowledge Organization System. 2012. Disponível em: <<https://www.w3.org/2004/02/skos/>>. Acesso em: 10 dez. 2015.

WORLD WIDE WEB CONSORTIUM – W3C. SKOS-XL Simple Knowledge Organization System eXtension for Labels. 2012. Disponível em:

<<https://www.w3.org/TR/skos-reference/skos-xl.html>>. Acesso em: 10 dez. 2015.

WORLD WIDE WEB CONSORTIUM – W3C. Web Ontology Language. 2012.

Disponível em: <<https://www.w3.org/2001/sw/wiki/OWL>>. Acesso em: 10 dez. 2015.

YU, J.; THOM, J. A.; TAM, A. Ontology evaluation using wikipedia categories for browsing. In: PROCEEDINGS OF THE SIXTEENTH ACM CONFERENCE ON CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 2007, New York. *Proceedings...*New York, NY: ACM, 2007. p. 223-232. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.184.780&rep=rep1&type=pdf>>. Acesso em: 12 ago. 2014.

ZAVITSANOS, Elias; PALIOURAS, Georgios; VOUIROS, George A. Gold standard evaluation of ontology learning methods through ontology transformation and alignment. *Transactions on Knowledge and Data Engineering*, [S.l.], v. 23, n. 11, p. 1635-1648, 2011.

ZENG, M. L. Knowledge Organization Systems. *Knowledge Organization*, Frankfurt, v. 35, n. 23, p. 160-182, 2008.

APÊNDICES

APÊNDICE A – Lista de documentos analisados para extração das expressões³⁶

Questão de competência	Conceito buscado	Referências dos documentos selecionados
Questão 1	Qualidade de água	<p>LIMA, G. T. N. P. de; BERTOLO, L. S.; SANTOS, R. F. dos; SILVA, J. dos S. V. da. A influência da composição dos elementos da paisagem na qualidade de água e nos serviços ecossistêmicos. Estudo de caso: bacia do rio Formoso, MS. In: SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 4., 2012, Bonito, MS. Anais... Brasília, DF: Embrapa, 2012. p. 1217-1227. 1 CD-ROM. Geopantanal, 2012.</p> <p>PRIMAVESI, O.; PRIMAVESI, A. C. P. A. A qualidade da água na microbacia hidrográfica do Ribeirão Canchim, São Carlos, SP, ocupada por atividade pecuária. In: CONGRESSO BRASILEIRO DE LIMINOLOGIA, 22., 1997, São Carlos. Resumos... São Carlos: UFSCar, 1997, p.391.</p> <p>FIGUEIREDO, R. de O. A sustentabilidade na agricultura e suas relações com a qualidade das águas em bacias hidrográficas na Amazônia Oriental. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 19., 2011, Maceió. Anais... Maceió: Associação Brasileira de Recursos Hídricos, 2011. 14 p.</p> <p>SILVEIRA, M. P.; QUEIROZ, J. F. de; RIBACINKO, D. B.; FERRAZ, J. M. G.; CARVALHO, M. P.; MARIGO, A. L. S.; SITTON, M.; ZAMBON, G.; SILVA, J. R. da. Avaliação biológica da qualidade da água em duas microbacias do rio Mogi Guaçu (SP) e sua relação com os impactos agrícolas. In: REUNIÃO ANUAL DO INSTITUTO BIOLÓGICO, 9., 2006, São Paulo, SP. O Biológico, v. 68, sup. 2, 2006. p. 1-10.</p> <p>BATALHA, S. S. A.; MARTORANO, L. G.; BIASE, A. G.; MORALES, G. P.; PONTES, A. N.; SANTOS, L. S. dos. Condições físico-químicas e biológicas em águas superficiais do Rio Tapajós e a</p>

³⁶ As referências apresentadas no APÊNDICE A foram extraídas como relatório da base de dados BDPA. Portanto, os dados que aparecem nas referências são os mesmos do relatório.

Questão de competência	Conceito buscado	Referências dos documentos selecionados
		<p>conservação de Floresta Nacional na Amazônia, Brasil. <i>Revista Ambiente & Água</i>, Taubaté, v. 9, n. 4, p. 647-663, 2014.</p> <p>AMORIM, M. C. C. de; PORTO, E. R. Considerações sobre controle e vigilância da qualidade de água de cisternas e seus tratamentos. In: SIMPÓSIO BRASILEIRO DE CAPTAÇÃO E MANEJO DE ÁGUA DE CHUVA, 4., 2003, Juazeiro. Anais... Juazeiro: ABCMAC; Petrolina: Embrapa Semiárido, 2003. 1 CD-ROM.</p> <p>CUNHA, G.R. da; SANTI, A.; DALMAGO, G.A.; PIRES, J.L.F.; PASINATO, A. Dinâmica do pH da água das chuvas em Passo Fundo, RS. <i>Pesquisa Agropecuária Brasileira</i>, Brasília, v.44, n. 4, p. 339-346, abr. 2009.</p> <p>AMORIM, J.R. de A.; FERNANDES, P.D.; GHEYI, H.R.; AZEVEDO, N.C. de. Efeito da salinidade e modo de aplicação da água de irrigação no crescimento e produção de alho. <i>Pesquisa Agropecuária Brasileira</i>, Brasília, v.37, nº 2, p. 167-176, fev. 2002.</p> <p>BRITO, L. T. de L.; ANJOS, J. B. dos; PORTO, E. R.; SILVA, A. de S.; SOUZA, M. A. de; XENOFONTE, G. H. S. Qualidade físico-química e bacteriológica das águas de cisternas no município de Ouricuri-PE. In: SIMPÓSIO BRASILEIRO DE CAPTAÇÃO E MANEJO DE ÁGUA DE CHUVA, 5., 2005, Teresina, PI. Anais... Teresina: ABCMAC; Governo do Estado do Piauí; Embrapa Semiárido; IRPAA; ASA, 2005. 1 CD-ROM.</p> <p>SILVA, F. N. dos S.; SOUZA, F. D. M. de; LOPES, Í. F.; PIRES, E. Í. S.; GOMES, R. B.; GIRAO, E. G. Vigilantes da água no Ceará: monitoramento participativo da qualidade da água da microbacia de Neblina como ferramenta de gestão ambiental. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA PARA O PROGRESSO DA CIÊNCIA, 62., 2010, Natal. Resumos... São Paulo: Sociedade Brasileira para o Progresso da Ciência, 2010.</p>
	Cultivo orgânico	<p>BORGES, L. de A. B.; SANTOS, I. G. dos; SOARES, R. A. B.; FERNANDES, P. M.; MADARI, B. E.; SILVA, M. A. S. da. Carbono e densidade do solo em uma cronosequência de canaviais cultivado em sistema orgânico. In: ENCONTRO BRASILEIRO DE SUBSTÂNCIAS HÚMICAS, 10., 2013, Santo</p>

Questão de competência	Conceito buscado	Referências dos documentos selecionados
		<p>Antônio de Goiás. Matéria orgânica e qualidade ambiental: anais. Brasília, DF: Embrapa, 2013. p. 149-152.</p>
		<p>SANTOS, T. L. S.; MELO, L. dos S.; NUNES, M. U. C. Comportamento de cultivares de morango em cultivo orgânico na região sul do Estado de Sergipe. In: SEMINÁRIO DE INICIAÇÃO CIENTÍFICA E PÓS-GRADUAÇÃO DA EMBRAPA TABULEIROS COSTEIROS, 4., 2014, Aracaju. Anais... Brasília, DF: Embrapa, 2014.</p>
		<p>PINTO, J. M.; GAVA, C. A. T.; LIMA, M. A. C.; SILVA, A. F.; RESENDE, G. M. de. Cultivo orgânico de meloeiro com aplicação de biofertilizantes e doses de substância húmica via fertirrigação. <i>Revista Ceres</i>, Viçosa, 55, n.4, p. 280-286, 2008</p>
		<p>FERREIRA, T. C.; ARAUJO, N. C. Cultivo orgânico de Zea mays L. com manipueira. CONGRESSO BRASILEIRO DE MAMONA, 5.; SIMPÓSIO INTERNACIONAL DE OLEAGINOSAS ENERGÉTICAS, 2.; FÓRUM CAPIXABA DE PINHÃO-MANSO, 1., 2012, Guarapari. Desafios e Oportunidades: anais. Campina Grande: Embrapa Algodão, 2012. p. 301</p>
		<p>AMARO, G. B.; RESENDE, F. V.; CARVALHO, A. D. F. de; LOPES, J. F.; LIMA, M. F.; MICHEREFF FILHO, M. Desempenho de cultivares de abóbora do tipo japonesa no cultivo orgânico. <i>Horticultura Brasileira</i>, Brasília, DF, v. 30, n. 2, p. S5518-S5523, jul. 2012. Suplemento. Trabalho apresentado no 52. Congresso Brasileiro de Olericultura, Salvador, 2012.</p>
		<p>COSTA, N. D.; ARAUJO, J. F.; SANTOS, C. A. F.; RESENDE, G. M. de; LIMA, M. A. C. de. Desempenho de cultivares de cebola em cultivo orgânico e tipos de solo no Vale do São Francisco. <i>Horticultura Brasileira</i>, Brasília, DF, v. 26, n. 4, p. 476-480, out./dez. 2008.</p>
		<p>OLIVEIRA, F. N. S.; LIMA, A. A. C.; AQUINO, A. R. L.; SANTOS, J. L. Manejo sustentável do cajueiro anão precoce enxertado em cultivo orgânico. Fortaleza: Embrapa Agroindústria Tropical, 2000. 3 p. (Embrapa Agroindústria Tropical. Pesquisa em Andamento, 79.).</p>
		<p>AMARIZ, A.; LIMA, M. A. C. de; COSTA, N. D.; TRINDADE, D. C. G. da; RIBEIRO, T. P. Qualidade e conservação pós-colheita de cultivares de cebola submetidas a doses de potássio em cultivo orgânico,</p>

Questão de competência	Conceito buscado	Referências dos documentos selecionados
	Bacia	<p>no Vale do São Francisco. In: JORNADA DE INICIAÇÃO CIENTÍFICA DA EMBRAPA SEMIÁRIDO, 3., 2008, Petrolina. Anais... Petrolina: Embrapa Semiárido, 2008. p. 173-179.</p> <p>COSTA, T. C. e C. da; GUIMARÃES, S. P. Delineamento e parametrização ambiental de microbacias hidrográficas para o Estado do Rio de Janeiro. In: CONGRESSO ACADÊMICO DE MEIO AMBIENTE E DESENVOLVIMENTO, 2., 2006, Niterói. Anais... Niterói: Instituto de Geociências-UFF, 2006. 1 CD-ROM.</p> <p>ROSA, M. B. S. da. Dinâmica do carbono em pequenas bacias de drenagem sob uso de agricultura familiar na Amazônia Oriental. 2007. 119 p. il. Dissertação (Mestrado) – Universidade Federal do Pará, Belém, PA. Orientada por Ricardo de Oliveira Figueiredo, CPATU.</p> <p>BARRETO, A. N.; NOGUEIRA, L. C.; CAMPOS, G. M. Irrigação por bacias em nível na cultura da bananeira. Parnaíba: EMBRAPA-CNPAL, 1992. 33 p. (EMBRAPA-CNPAL. Circular Técnica, 4).</p> <p>FRITZSONS, E.; MANTOVANI, L. E. Os substratos geológicos e os coeficientes morfométricos em bacias hidrográficas do carste dolomítico no Estado do Paraná. <i>Revista Brasileira de Geografia Física</i>, Recife, v. 3, p. 181-189, 2010.</p> <p>ESQUERDO, J. C. D. M.; SILVA, J. dos S. V. da. Redefinição dos limites das sub-bacias hidrográficas do estado do Mato Grosso do sul com o auxílio de técnicas computacionais. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 8., 2011, Bento Gonçalves. Anais... Florianópolis: UFSC; Pelotas: UFPel, 2011. Não paginado. SBIAgro 2011.</p> <p>ESQUERDO, J. C. D. M.; SILVA, J. dos S. V. da. Uso de geotecnologias na redefinição dos limites das sub-bacias hidrográficas do Estado de Mato Grosso do Sul. In: SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 4., 2012, Bonito, MS. Anais... Brasília, DF: Embrapa, 2012. p. 1033-1042. 1 CD-ROM. Geopantanal 2012.</p>
Questão 2	Água	RESENDE, A. V. de. <i>Agricultura e qualidade da água: contaminação da água por nitrato</i> . Planaltina: Embrapa Cerrados, 2002. 29 p. (Embrapa Cerrados. Documentos, 57). CRI6560

Questão de competência	Conceito buscado	Referências dos documentos selecionados
		<p>QUEIROZ, J. F. de; PESSOA, M. C. P. Y. <i>Água sempre presente na vida</i>. Jaguariúna: Embrapa Meio Ambiente, 2003. não paginado. il. color. (Cartilhas dos Jogos Ambientais da Ema, 1).</p> <p>SILVA, A. de S.; PORTO, E. R.; LIMA, L. T. de; GOMES, P. C. F. <i>Captação e conservação de água de chuva para consumo humano: cisternas rurais-dimensionamento, construção e manejo</i>. Petrolina: EMBRAPA-CPATSA, 1984. 103 p. il. (EMBRAPA-CPATSA. Circular Técnica, 12).</p> <p>REIS, G. da S.; MENEZES, E. S.; MARINS, D. de M.; LAITANO, O. Efeitos da ingestão prévia de água de coco sobre o balanço hídrico e desempenho aeróbio no calor. In: JORNADA DE INICIAÇÃO CIENTÍFICA DA EMBRAPA SEMIÁRIDO, 7.; JORNADA DE INICIAÇÃO CIENTÍFICA DA FACEPE/UNIVASF, 1., 2012, Petrolina. Anais... Petrolina: Embrapa Semiárido, 2012. p. 511-517. 1 CD-ROM. (Embrapa Semiárido. Documentos, 248).</p> <p>MARQUELLI, W. A.; OLIVEIRA, A. S. de; COELHO, E. F.; NOGUEIRA, L. C.; SOUSA, V. F. de. Manejo da água de irrigação. In: SOUSA, V. F. de; MARQUELLI, W. A.; COELHO, E. F.; PINTO, J. M.; COELHO FILHO, M. A. (Ed.). <i>Irrigação e fertirrigação em fruteiras e hortaliças</i>. Brasília, DF: Embrapa Informação Tecnológica: Embrapa Hortaliças; São Luís: Embrapa Cocais; Cruz das Almas: Embrapa Mandioca e Fruticultura; Petrolina: Embrapa Semiárido, 2011. p. 159-232.</p> <p>LANDAU, E. C.; BARROS, L. C. de; RIBEIRO, P. E. de A.; BARROS, I. de R. <i>Abrangência geográfica do Projeto Barraginhas no Brasil</i>. Sete Lagoas: Embrapa Milho e Sorgo, 2013. 45 p. (Embrapa Milho e Sorgo. Documentos, 159).</p> <p>BRITO, L. T. de L.; PORTO, E. R.; SILVA, D. F. da; HOLANDA JÚNIOR, E. V. DE; CAVALCANTI, N. de B. Água de chuva para consumo animal: estudo de caso com caprinos. In: SIMPÓSIO BRASILEIRO DE CAPTAÇÃO E MANEJO DE ÁGUA DE CHUVA, 5., 2005, Teresina, PI. Anais... Teresina: ABCMAC; Governo do Estado do Piauí; Embrapa Semiárido; IRPAA; ASA, 2005. 1 CD-ROM</p> <p>BARBOLA, I. F.; MORAES, M. F. P. G.; ANAZAWA, T. M.; NASCIMENTO, E. A.; SEPKA, E. R.; POLEGATTO, C. M.; MILLÉO, J.; SCHUHLI, G. S. e. Avaliação da comunidade de macroinvertebrados aquáticos como ferramenta para o monitoramento de um reservatório na bacia do Rio Pitangui, Paraná, Brasil, Iheringia, Série Zoologia, Porto Alegre, v.101, n. 1-2, p. 15-23, jun. 2011</p>

Questão de competência	Conceito buscado	Referências dos documentos selecionados
		<p>SILVA, A. de S.; PORTO, E. R.; LIMA, L. T. de; GOMES, P. C. F. <i>Captação e conservação de água de chuva para consumo humano: cisternas rurais-dimensionamento, construção e manejo</i>. Petrolina: EMBRAPA-CPATSA, 1984. 103 p. il. (EMBRAPA-CPATSA. Circular Técnica, 12).</p> <p>SILVA, A. de S.; SANTOS, E. D.; MAGALHÃES, A. A. <i>Introdução e avaliação do método de irrigação por sucção na região do Trópico Semiárido</i>. Petrolina: EMBRAPA-CPATSA, 1978 14 p. Trabalho preparado para o Congresso Nacional de Irrigação e Drenagem, Salvador, set. 1978.</p> <p>PRADO, R. B.; NOVO, E. M. L. M. Modeling pollution potential input from the drainage basin into Barra Bonita reservoir, São Paulo - Brazil. <i>Brazilian Journal of Biology</i>, São Carlos, SP, v. 75, n. 2, p. 314-323, 2015.</p> <p>RODRIGUES, L. N.; SANO, E. E.; AZEVEDO, J. A. de; SILVA, E. M. da. Pequenas barragens de terra na Bacia do Rio Preto: distribuição espacial e área do espelho d'água (2007). Planaltina, DF: Embrapa Cerrados, 2007. 1 folder.</p>
Questão 3	Desmatamento	<p>MIRANDA, E. E. de. Campeões de desmatamento. <i>Revista de Política Agrícola</i>, Brasília, DF, ano 15, n. 3, p. 83-84, jul./set. 2006. Ponto de Vista</p> <p>TERRA, T. N.; SILVA, J. dos S. V. da. Cenário de desmatamento da Bacia do Alto Paraguai (BAP) em 2022 e 2050. In: SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 4., 2012, Bonito, MS. Anais... Brasília, DF: Embrapa, 2012. p. 741-749. 1 CD-ROM. Geopantanal 2012.</p> <p>ALBUQUERQUE, S.G. de; SOARES, J.G.G.; OLIVEIRA, M.C. de; SALVIANO, L.M.C. Desempenho do capim-buffel sob vários métodos de estabelecimento no Sertão de Pernambuco. <i>Pesquisa Agropecuária Brasileira</i>, Brasília, v.29, n.8, p.1225-1230, ago. 1994.</p>

Questão de competência	Conceito buscado	Referências dos documentos selecionados
		<p>SILVA, J. dos S. V. da; ABDON, M. de M.; MORAES, J. A. de. Desmatamento na bacia do Alto Paraguai no Brasil. In: SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 3., 2010, Cáceres, MT. Anais... Campinas: Embrapa Informática Agropecuária; São José dos Campos: INPE, 2010. p. 458-467. Geopantanal 2010.</p> <p>ANJOS, J. B. dos; PATANCHON, J. L. <i>Destocador manual</i> - CPATSA. Petrolina: EMBRAPA-CPATSA, 1990. 5 p. il. (EMBRAPA-CPATSA. Comunicado Técnico, 42).</p> <p>ANGELO, H.; POMPERMAYER, R. S.; ALMEIDA, A. N. de; MOREIRA, J. M. M. A. P. O custo social do desmatamento da Amazônia brasileira: o caso da Castanha-do-Brasil (<i>Bertholletia excelsa</i>). <i>Ciência Florestal</i>, Santa Maria, RS, v. 23, n. 1, p. 183-191, jan./mar. 2013.</p> <p>SILVA, G. B. S. da; MELLO, A. Y. I.; STEINKE, V. A. Unidades de conservação no bioma Cerrado: desafios e oportunidades para a conservação no Mato Grosso. <i>Geografia</i>, Rio Claro, SP, v. 37, n. 3, p. 541-554, set./dez. 2012.</p>
	Nascente	<p>ARAGAO, R. de; CRUZ, M. A. S.; AMORIM, J. R. A. de; MENDONCA, L. C.; PANTALEAO, S. de M.; SCHNEIDER, E. H. M.; FIGUEIREDO, E. E. de; SRINIVASAM, V. S. Análise da influência da mata ciliar na bacia do rio Japarutuba-mirim, Sergipe, sobre o escoamento superficial e produção de sedimentos simulados com o modelo Swat. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 20., 2013, Bento Gonçalves. Água desenvolvimento econômico e socioambiental: anais. Bento Gonçalves, RS: ABRH, 2013.</p> <p>ARAGAO, R. de; CRUZ, M. A. S.; AMORIM, J. R. A. de; MENDONCA, L. C.; PANTALEAO, S. de M. Avaliação da Influência da Mata Ciliar na Bacia do Rio Siriri, Sergipe, sobre o Escoamento Superficial e Produção de Sedimentos via Modelo SWAT. In: ENCONTRO DE RECURSO HÍDRICO EM SERGIPE, 7., 2014, Aracaju. Anais... Brasília, DF: Embrapa, 2014.</p> <p>SILVA, A. de S.; SILVA, A. L. da; SILVA, C. M. M. de S.; BRITO, L. T. de L. Avaliação de acesso, uso e manejo de água para consumo humano no Semiárido brasileiro. In: SEABRA, G. (Org.). <i>Terra: qualidade de vida, mobilidade e segurança nas cidades</i>. João Pessoa:UFPB, 2013. v. 3, p. 810-821. 1 CD-ROM.</p>

Questão de competência	Conceito buscado	Referências dos documentos selecionados
		<p>SILVA, A. de S.; FAY, E. F.; BRITO, L. T. de L. Avaliação microbiológica das águas de uso domiciliar. In: SIMPÓSIO BRASILEIRO DE CAPTAÇÃO E MANEJO DE ÁGUA DE CHUVA, 9., 2014, Feira de Santana. Água de chuva: segurança hídrica para o século XXI. Feira de Santana: Universidade Estadual de Feira de Santana: Universidade Federal do Recôncavo da Bahia, 2014. 1 CD-ROM.</p>
		<p>MAGALHÃES, L. T. S.; GOMES, J. B. V.; VASCO, A. N. do; AGUIAR NETTO, A. de O.; FERREIRA, R. A. Caracterização geo-pedológica das áreas de nascentes na bacia hidrográfica do rio Piauitinga, Sergipe, Brasil. <i>Revista Ambiente & Água</i>, Taubaté, p. 169-181, v. 7, n. 1, 2012.</p>
	Quantidade de água	<p>SOUZA, A. G. dos S.; RESENDE, R. S.; CARVALHO, T. B. de. Formação do bulbo molhado do coco. In: SEMINÁRIO DE INICIAÇÃO CIENTÍFICA E PÓS-GRADUAÇÃO DA EMBRAPA TABULEIROS COSTEIROS, 5., 2015, Aracaju. Anais... Brasília, DF: Embrapa, 2015. p. 280, ref. 220-226. 1 CD-ROM. Editor Técnico: Marcelo Ferreira Fernandes.</p> <p>DANTAS NETO, J.; SILVA, F. de A.S. e; FURTADO, D.A.; MATOS, J. de A. de. Influência da precipitação e idade da planta na produção e composição química do capim-buffel. <i>Pesquisa Agropecuária Brasileira</i>, Brasília, v.35, n.9, p.1867-74, set.2000.</p> <p>NASCIMENTO, P. dos S.; BASSOI, L. H.; PAZ, V. P. da S. Planilha eletrônica para auxílio à tomada de decisão em manejo de irrigação. <i>Irriga</i>, Botucatu, v. 17, n. 1, p. 1-15, jan./mar. 2012</p> <p>SILVA, S. M. S. e ; GHEYI, H. R.; BELTRÃO, N. E. de M.; SEVERINO, L. S.; DIAS, J. M.; NASCIMENTO, R. T. do. Produtividade da cultivar BRS Energia em função da lâmina de irrigação e populações de plantas. In: CONGRESSO BRASILEIRO DE MAMONA, 3., 2008, Salvador. Energia e ricinoquímica: anais. Salvador: SEAGRI: Embrapa Algodão, 2008. 1 CD-ROM</p> <p>SILVA, S. M. S. e; GHEYI, H. R.; BELTRÃO, N. E. de M.; GONDIM, T. M. de S.; SEVERINO, L. S.; CONCEIÇÃO, J. L. A. Influência da lâmina de irrigação e população de plantas em componentes de produção da mamona cultivar BRS Energia. In: CONGRESSO BRASILEIRO DE MAMONA, 3., 2008, Salvador. Energia e ricinoquímica: resumos. Salvador: SEAGRI: Embrapa Algodão, 2008. p. 140</p>

Questão de competência	Conceito buscado	Referências dos documentos selecionados
		<p>COLOZZI FILHO, A.; SIQUEIRA, J.O.; OLIVEIRA, E. de. Desinfestação Superficial de Esporos de Fungos Micorrzicos Vesicular-Arbusculares. I. Efeitos da Concentração e Tempo de Exposição de Agentes Desinfestantes e Antibióticos. Pesquisa Agropecuária Brasileira, Brasília, v.29, n.7, p.1119-1127. jul. 1994.</p>
	Rio	<p>OLIVEIRA, M. D. de; CALHEIROS, D. F. <i>Aporte de nutrientes e sólidos suspensos no Rio Taquari</i>. Corumbá: Embrapa Pantanal, 2002. 6 p. il. (Embrapa Pantanal. Circular Técnica, 31).</p> <p>OLIVEIRA, H. de; URCHEI, M.A.; FIETZ, C.R. <i>Aspectos físicos e socioeconômicos da bacia hidrográfica do rio Ivinhema</i>. Dourados: Embrapa Agropecuária Oeste, 2000. 52p. (Embrapa Agropecuária Oeste. Documentos, 25).</p> <p>OLIVEIRA, M. D. de; FERREIRA, C. J. <i>Estudos limnológicos para o monitoramento da bacia hidrográfica do Rio Miranda, Pantanal Sul</i>. Corumbá: Embrapa Pantanal, 2003. 61 p. (Embrapa Pantanal. Boletim de Pesquisa e Desenvolvimento, 54).</p> <p>GALDINO, S. <i>Método probabilístico de previsão do nívelmínimo no Rio Paraguai, em Ladário, MS - Pantanal</i>. Corumbá: Embrapa Pantanal, 2001. 42p. (Embrapa Pantanal. Circular Técnica, 28).</p> <p>SOUZA, O. C.de ; ARAÚJO, M. R.; USSAMI, N.; FISSEHA, S.; KONZEN, L. <i>Mudanças hidrológicas no pantanal associadas a processos erosivos e tectônicos na Bacia do Rio Taquari, MS</i>. Campo Grande, MS: Embrapa Gado de Corte, 2004. 26 p. (Embrapa Gado de Corte. Documentos, 145).</p>

APÊNDICE B – Expressões (em português e inglês) extraídas da coleção de documentos e dos tesouros

Questão 1: Os cultivos orgânicos ocasionam algum impacto na qualidade de água de nossas bacias?	
Conceito: Qualidade de água	
Expressões relacionadas (em português)	Expressões relacionadas (em inglês)
alcalinidade	Alkalinity
amônia não ionizada	un-ionized ammonia (NH ₃)
amônio	Ammonium
carbono orgânico total	total organic carbon (TOC)
cloreto	Chloride
cloreto dissolvido	dissolved chloride
clorofila	Chlorophyll
coliformes fecais	fecal coliforms
Coliformes Termotolerantes	thermotolerant coliforms
Coliformes Totais	Total coliforms
condutividade elétrica	Electrical conductivity
cor aparente	apparent color
cor verdadeira	true Color
Demanda Bioquímica de Oxigênio	Biochemical Oxygen Demand
demanda química de oxigênio	chemical oxygen demand
dureza da água	water hardness
Escherichia coli (E. coli)	Escherichia coli (E. coli)
ferro total	total iron
fósforo dissolvido	dissolved phosphorus
Fósforo total	Total phosphorus
manganês total	Total manganese
N-amoniaco	Ammonium Nitrogen
Nitrato	nitrate

Nitrogênio Total	Total Nitrogen
N-nitrato	N-nitrate
N-nitrito	N-Nitrite
Ortofosfato	orthophosphate
oxigênio consumido	oxygen consumed
oxigênio dissolvido	Dissolved oxygen
Pesticidas	pesticides
pH (potencial hidrogeniônico)	pH (potential hydrogenionic)
Qualidade de água	Water quality
salinidade	salinity
Sódio	Sodium
sólidos dissolvidos fixos	fixed solids dissolved
sólidos dissolvidos totais	total dissolved solids
sólidos dissolvidos voláteis	solid volatile dissolved
sólidos em suspensão	solid in suspension
sólidos fixos	solid fixed
sólidos sedimentados	solid sedimented
sólidos suspensos fixos	fixed suspended solids
sólidos suspensos voláteis	volatile suspended solids
Sólidos Totais	Total Solids
sólidos voláteis	volatile solids
Temperatura	Temperature
temperatura da água	water temperature
Turbidez	Turbidity
Conceito: Cultivo orgânico	
Expressões relacionadas (em português)	Expressões relacionadas (em inglês)
Agricultura alternativa	alternative agriculture
Agricultura orgânica	organic agriculture
Agroecologia	agroecology
Cultivo orgânico	organic crop
Cultura orgânica	Organic culture

Manejo orgânico	organic management
Manejo sustentável	sustainable management
Produção orgânica	organic production
Sistema de cultivo orgânico	organic crop system
Sistema de produção alternativo	Alternative production system
Sistema de produção orgânico	organic system
Sistema orgânico	organic system
Sistema orgânico de cultivo	organic system of farming

Conceito: Bacia

Expressões relacionadas (em português)	Expressões relacionadas (em inglês)
Bacia	basin
Bacia de drenagem	Drainage basin
Bacia hidrográfica	Watershed
Rede hidrográfica	hydrographic network
Região hidrográfica	river basin district
Sub bacia	sub basin

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

Conceito: Água

Expressões relacionadas (em português)	Expressões relacionadas (em inglês)
Açude	Dam
Água	Water
Água doce	Fresh water
Água em estado gasoso	Water in gaseous state
Água em estado líquido	Liquid water
Água em estado sólido	Water solid state
Água subterrânea	Subterranean water
Água superficial	Surface water

Aquífero	Aquifer
Barragem	Barrage
Cachoeira	waterfall
Corpo de água	Body of water
Lago	Lake
Lagoa	Lake
Manancial/Nascente	Water springs
Recurso hídrico	Water resource
Represa	Dam
Rio	River
Conceito: Reservatório de abastecimento de água	
Expressões relacionadas (em português)	Expressões relacionadas (em inglês)
Armazenamento de água	Water storage
Barragem	Dam
Cisterna/Tanque de água	Water tank
Lago	Lake
Lagoa artificial	Artificial Lake
Lençol freático	Groundwater table
Represa	Dam
Reserva hídrica	water reserve
Reservatório	Reservoir
Reservatório de abastecimento	Reservoir supply
Reservatório de água	Water reservoir
Reservatório subterrâneo	Underground reservoir
Sistema de armazenamento de água	Water Storage System
Tanque de armazenamento	Storage tank
Conceito: População (paulista)	
Expressões relacionadas (em português)	Expressões relacionadas (em inglês)

Densidade populacional	Population density
População	Population
População animal	Animal population
População de planta	Plant population
População humana	Human population

Conceito: Tempo (Ano 2016)

Expressões relacionadas (em português)	Expressões relacionadas (em inglês)
Duração	Duration
Periodicidade	Periodicity
Tempo	Time

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

Conceito: Desmatamento

Expressões relacionadas (em português)	Expressões relacionadas (em inglês)
Área degradada	Degraded area
Área desmatada	Deforested área
Degradação ambiental	Environmental degradation
Degradação de floresta	Forest degradation
Desarborização	Deforestation
Desflorestamento	Deforestation
Desmatamento	Deforestation
Desmatamento manual	Manual deforestation
Desmatamento mecânico	Mechanical deforestation
Desmate	Deforestation

Conceito: Nascente

Expressões relacionadas (em português)	Expressões relacionadas (em inglês)
Área de nascente	Spring área
Cacimba	Water hole
Manancial	Wellspring

Nascente	Spring/Wells
Nascente de água	Water spring
Conceito: Quantidade de água	
Expressões relacionadas (em português)	Expressões relacionadas (em inglês)
Água disponível	Available water
Balanço hídrico	Water balance
Quantidade de água	Water quantity
Volume de água	Water volume
Conceito: Rio	
Expressões relacionadas (em português)	Expressões relacionadas (em inglês)
Afluente	Affluent/ Tributary
Corpo de água	Body of water
Curso d'água	Watercourse
Rio	River
Sistema fluvial	River system

APÊNDICE C – Representação das expressões pela OntoAgroHidro

Questão 1: Os cultivos orgânicos ocasionam algum impacto na qualidade de água de nossas bacias?		
Conceito: Qualidade de água		
Expressões relacionadas	Expressões em inglês	Representação do conceito na OntoAgroHidro
alcalinidade	alkalinity	Sim
amônia não ionizada	un-ionized ammonia (NH ₃)	Sim
amônio	ammonium	Sim
carbono orgânico total	total organic carbon (TOC)	Sim
cloreto	chloride	Sim
cloreto dissolvido	dissolved chloride	Não
clorofila	chlorophyll	Sim
coliformes fecais	fecal coliforms	Sim
Coliformes Termotolerantes	thermotolerant coliforms	Sim
Coliformes Totais	Total coliforms	Sim
condutividade elétrica	Electrical conductivity	Sim
cor aparente	apparent color	Não
cor verdadeira	true Color	Não
Demanda Bioquímica de Oxigênio	Biochemical Oxygen Demand	Sim
demanda química de oxigênio	chemical oxygen demand	Sim
dureza da água	water hardness	Sim
Escherichia coli (E. coli)	Escherichia coli (E. coli)	Sim
ferro total	total iron	Sim
fósforo dissolvido	dissolved phosphorus	Sim
Fósforo total	Total phosphorus	Sim
manganês total	Total manganese	Sim

N-amoniaco	Ammonium Nitrogen	Não
nitrato	nitrate	Sim
Nitrogênio Total	Total Nitrogen	Sim
N-nitrato	N-nitrate	Sim
N-nitrito	N-Nitrite	Sim
ortofosfato	orthophosphate	Sim
oxigênio consumido	oxygen consumed	Não
oxigênio dissolvido	Dissolved oxygen	Sim
pesticidas	pesticides	Sim
pH (potencial hidrogeniônico)	pH (potential hydrogenionic)	Sim
Qualidade de água	Water quality	Sim
salinidade	salinity	Sim
Sódio	Sodium	Sim
sólidos dissolvidos fixos	fixed solids dissolved	Sim
sólidos dissolvidos totais	total dissolved solids	Sim
sólidos dissolvidos voláteis	solid volatile dissolved	Não
sólidos em suspensão	solid in suspension	Sim
sólidos fixos	solid fixed	Sim
sólidos sedimentados	solid sedimented	Sim
sólidos suspensos fixos	fixed suspended solids	Sim
sólidos suspensos voláteis	volatile suspended solids	Sim
Sólidos Totais	Total Solids	Não
sólidos voláteis	volatile solids	Não
Temperatura	Temperature	Sim
temperatura da água	water temperature	Sim
Turbidez	Turbidity	Sim
Conceito: Cultivo orgânico		

Expressões relacionadas	Expressões em inglês	Representação do Conceito na OntoAgroHidro
Agricultura alternativa	alternative agriculture	Não
Agricultura orgânica	organic agriculture	Não
Agroecologia	agroecology	Não
Cultivo orgânico	organic crop	Não
Manejo orgânico	organic management	Não
Manejo sustentável	sustainable management	Não
Produção orgânica	organic production	Não
Sistema de cultivo orgânico	organic crop system	Não
Sistema de produção alternativo	Alternative production system	Não
Sistema de produção orgânico	organic system	Não
Sistema orgânico	organic system	Não
Sistema orgânico de cultivo	organic system of farming	Não
Cultura orgânica	Organic culture	Não
Conceito: Bacia		
Expressões relacionadas	Expressões em inglês	Representação do Conceito na OntoAgroHidro
Bacia	basin	Não
Bacia hidrográfica	Watershed	Não
Rede hidrográfica	hydrographic network	Não
Sub bacia	sub basin	Não
Bacia de drenagem	Drainage basin	Não
Região hidrográfica	river basin district	Sim
Questão 2: Faltará água nos reservatórios de abastecimento de água para a população paulista em 2016?		

Conceito: Água		
Expressões relacionadas	Expressões em inglês	Representação do Conceito na OntoAgroHidro
Açude	Dam	Não
Água	Water	Sim
Água doce	Fresh water	Não
Água em estado gasoso	Water in gaseous state	Sim
Água em estado líquido	Liquid water	Sim
Água em estado sólido	Water solid state	Sim
Água subterrânea	Subterranean water	Sim
Água superficial	Surface water	Sim
Aquífero	Aquifer	Não
Lago	Lake	Sim
Manancial/Nascente	Water springs	Não
Represa	Dam	Não
Rio	River	Sim
Cachoeira	waterfall	Não
Recurso hídrico	Water resource	Sim
Lagoa	Lake	Sim
Barragem	Barrage	Sim
Corpo de água	Body of water	Sim
Conceito: Reservatório de abastecimento de água		
Expressões relacionadas	Expressões em inglês	Representação do Conceito na OntoAgroHidro
Armazenamento de água	Water storage	Não
Barragem	Dam	Sim
Lago	Lake	Sim
Lençol freático	Groundwater table	Não
Represa	Dam	Não
Reserva hídrica	water reserve	Não

Reservatório	Reservoir	Não
Reservatório de abastecimento	Reservoir supply	Não
Reservatório de água	Water reservoir	Não
Reservatório subterrâneo	Underground reservoir	Não
Sistema de armazenamento de água	Water Storage System	Não
Tanque de armazenamento	Storage tank	Não
Lagoa artificial	Artificial Lake	Não
Cisterna/Tanque de água	Water tank	Não
Conceito: População (paulista)		
Expressões relacionadas	Expressões em inglês	Representação do Conceito na OntoAgroHidro
População	Population	Não
População humana	Human population	Não
População animal	Animal population	Não
População de planta	Plant population	Não
Densidade populacional	Population density	Não
Conceito: Tempo (Ano 2016)		
Expressões relacionadas	Expressões em inglês	Representação do Conceito na OntoAgroHidro
Tempo	Time	Sim
Duração	Duration	Não
Periodicidade	Periodicity	Sim
Questão 3: Como o desmatamento de nascentes afeta a quantidade e a qualidade de água de nossos rios?		
Conceito: desmatamento		

Expressões relacionadas	Expressões em inglês	Representação do Conceito na OntoAgroHidro
Desmatamento	Deforestation	Não
Área desmatada	Deforested area	Não
Área degradada	Degraded area	Não
Desmate	Deforestation	Não
Desmatamento manual	Manual deforestation	Não
Desmatamento mecânico	Mechanical deforestation	Não
Desflorestamento	Deforestation	Não
Degradação de floresta	Forest degradation	Não
Desarborização	Deforestation	Não
Degradação ambiental	Environmental degradation	Não
Conceito: nascente		
Expressões relacionadas	Expressões em inglês	Representação do Conceito na OntoAgroHidro
Nascente	Spring/wells	Não
Cacimba	Water hole	Não
Manancial	Wellspring	Não
Área de nascente	Spring area	Não
Nascente de água	Water spring	Não
Conceito: quantidade de água		
Expressões relacionadas	Expressões em inglês	Representação do Conceito na OntoAgroHidro
Quantidade de água	Water quantity	Sim
Volume de água	Water volume	Não
Água disponível	Avaliable water	Não
Balanço hídrico	Water balance	Não
Conceito: rio		

Expressões relacionadas	Expressões em inglês	Representação do Conceito na OntoAgroHidro
Rio	River	Sim
Sistema fluvial	River system	Não
Afluente	Affluent/ Tributary	Não
Curso d'água	Watercourse	Não
Corpo de água	Body of water	Sim

APÊNDICE D – Relatório das expressões incluídas na OntoAgroHidro

1. Expressão: Afluente

Object Properties

TopObjectProperty

ehAfluente

Anotações

Label (pt): Afluente

Label (en): Affluent

AltLabel (pt): Tributário

AltLabel (en): Tributary

Comment (pt): rio que deságua noutro rio maior. Corrente de água, que se vai lançar em outra.

Comment (en): river that flows into another larger river. Stream, which will release another.

Descrição

Domain (Intersection): Rio

Range (Intersection): BodyOfWater

2. Expressão: Aquífero

Dimension

EnvironmentalDimension

Hydrosphere

WaterResources

Aquífero

Anotações

Label (pt): Aquífero

Label (en): Aquifers

Comment (pt): Uma formação geológica ou estrutura que transmite a água em quantidade suficiente para suprir as necessidades para o desenvolvimento de água, tal como um poço. Aquíferos são de areia, brita, rocha fraturada, ou de rochas fragmentas e vesicular normalmente saturadas.

Comment (en): A geological formation or structure that transmits water in sufficient quantity to supply the needs for a water development, such as a well. Aquifers are usually saturated sands, gravel, fractured rock, or cavernous and vesicular rock.

3. Expressão: Bacia

Dimension

EnvironmentalDimension

Hydrosphere

RegiaoHidrografica

Bacia Hidrográfica

Anotações

PrefLabel (pt): Bacia Hidrográfica

PrefLabel (en): Watershed

AltLabel (en): Basin

Comment (pt): A área de terra que drena a água para um determinado córrego, rio ou lago. É uma característica da terra que pode ser identificada por rastreio de uma linha ao longo das elevações mais altas entre duas áreas no mapa, muitas vezes, um cume. (pt)

Comment (en): The land area that drains water to a particular stream, river, or lake. It is a land feature that can be identified by tracing a line along the highest elevations between two areas on a map, often a ridge.

4. Expressão: Cachoeira

Dimension

EnvironmentalDimension

Hydrosphere

AguaSuperficial

Cachoeira

Anotações

Label (pt): Cachoeira

Label (en): Waterfall

AltLabel (pt): Cascata

AltLabel (en): Cascade

Comment (pt): formação geomorfológica na qual os cursos de água correm por cima de uma rocha de composição resistente à erosão, formando degraus com desnível acentuado.

Comment (en): geomorphological formation in which water courses run over a rock composition resistant to erosion, forming steps an accentuated gap.

5. Expressão: Cisterna

Dimension

EnvironmentalDimension

Hydrosphere

WaterResources

ReservatóriodeAgua

Cisterna

Anotações

Label (pt): Cisterna; tanque de água

Label (en): Watertank

Comment (pt): poço, receptáculo, abaixo do nível da terra, para conservar águas pluviais.

Comment (en): Well, receptacle, below the level of the land, to conserve rainwater.

6. Expressão: Cloreto Dissolvido

Como conceito:

Definition

ChemicalDefinition

ChemicalCompound

Inorganic

Dissolved_Solids

Cloreto_Dissolvido

Como quantidade:

Data Properties

TopDataProperty

Cloreto_Dissolvido

Anotações

Label (pt): Cloreto Dissolvido

Label (en): DissolvedChloride

Comment (pt): Cloreto em química inorgânica é uma espécie iônica formada por um átomo de cloro carregado negativamente, com estado de oxidação -1. Por extensão, é um composto iônico formado por este ânion ou com uma estrutura formalmente similar, ou seja, com ligação covalente entre um átomo de cloro e um elemento menos eletronegativo. Substância que pode ser utilizada como parâmetro para medir a qualidade de água.

Comment (en): Chloride in inorganic chemistry is an ionic species formed from a negatively charged chlorine atom, with a -1 oxidation state. By extension is an ionic compound formed by this anion with a formal or similar structure, that is, covalent bond between a chlorine atom and a less electronegative element. Substance which can be used as a parameter to measure the quality of water

7. Expressão: Cor aparente

Data Properties

TopDataProperty

waterDescriptive

colorDescriptive

CorAparente

Anotações

Label (pt): Cor Aparente

Label (en): Apparent Color

Comment (pt): Parâmetro para medir a qualidade de água que se refere à determinação de cor em amostras com turbidez (com material coloidal ou em suspensão).

Comment (en): Parameter to measure the quality of water that refers to the determination of color samples with turbidity (with or colloidal material in suspension).

8. Expressão: Cor verdadeira

Data Properties

TopDataProperty

waterDescriptive

colorDescriptive

CorVerdadeira

Anotações

Label (pt): Cor Verdadeira

Label (en): True Color

Comment (pt): Parâmetro para medir a qualidade de água que se refere à determinação de cor em amostras sem turbidez (Ex: após filtração ou centrifugação).

Comment (en): Parameter to measure the quality of water that refers to the determination color without turbidity in samples (eg, after filtration or centrifugation).

9. Expressão: Cultivo orgânico

EnvironmentalTransformationAgent

ManejoAgricola

Fitotecnia

SistemaDeCultivo

CultivoOrganico

Anotações

Label (pt): Cultivo Orgânico

Label (en): OrganicCrop

PrefLabel (pt): Agricultura Orgânica

PrefLabel (en): OrganicAgriculture

AltLabel (pt): Produção Orgânica; Cultura orgânica

AltLabel (en): OrganicProduction

Comment (pt): Métodos agrícolas sem o uso de produtos químicos.

Comment (en): Agricultural methods without use of chemical products

10. Expressão: Curso de Água

Dimension

EnvironmentalDimension

Hydrosphere

BodyOfWater

CursoDeAgua

Anotações

Label (pt): Curso de Água

Label (en): Watercourse

AltLabel (pt): Curso d'água

Comment (pt): qualquer corpo de água fluente, como rios, córregos, riachos, regatos, ribeiros, etc.

Comment (en): any body of flowing water as rivers, streams, creeks, streams, rivers, etc.

11. Expressão: Desmatamento

EnvironmentalTransformation

AnthropicEnvironmentalTransformation

Desmatamento

Anotações

PrefLabel (pt): Desmatamento

PrefLabel (en): Deforestation

AltLabel (pt): Desflorestação; Desflorestamento

AltLabel (en): Environmental Degradation

HiddenLabel (pt): Desarborização

Comment (pt): é o processo de destruição das florestas por meio da ação do homem.

Comment (en): is the process of destruction of forests by human action

12. Expressão: Lagoa Artificial

Dimension

EnvironmentalDimension

Hydrosphere

WaterResources

ReservatórioDeAgua

LagoaArtificial

Anotações

Label (pt): Lagoa Artificial

Label (en): Artificial Lake

Comment (pt): é um lago de origem artificial, normalmente concebido para armazenar água numa região seca, para dar mais vida a um ambiente ou que resulta da construção de uma represa ou uma hidrelétrica, cobrindo as áreas costeiras às margens de um rio. Os lagos artificiais são também feitos para a criação de peixes e outros organismos aquáticos, em regime de aquacultura extensiva.

Comment (en): It is a lake of artificial origin, usually designed to store water in a dry region, to give more life to an environment or resulting from the construction of a dam or a hydroelectric plant, covering the coastal areas on the banks of a river. The artificial lakes are also made to the farming of fish and other aquatic organisms in extensive aquaculture regime.

13. Expressão: Lençol Freático

Dimension

EnvironmentalDimension

Hydrosphere

AguaSubterranea

LencolFreatico

Anotações

Label (pt): Lençol Freático

Label (en): Groundwatertable

Comment (pt): A superfície superior das águas subterrâneas ou a profundidade abaixo do qual o solo é saturado com água.

Comment (en): The upper surface of the groundwater or that depth below which the soil is saturated with water.

14. Expressão: Nascente

Dimension

EnvironmentalDimension

Hydrosphere

AguaSuperficial /BodyOfWater

Nascente

Anotações

Label (pt): Nascente

*Label (en):*Watersprings

AltLabel (pt): Manancial

Comment (pt): local onde a água subterrânea brota à superfície.

Comment (en): place where the groundwater flows to the surface.

15. Expressão: Nitrogênio Amoniacal (NH₃-N)

Definition

ChemicalDefinition

ChemicalCompound

Nutrient

Nitrogen

Ammonia

Nitrogenio_Amoniacal_(NH3-N) (Instância)

Anotações

Label (pt): Nitrogênio Amoniacal (NH3-N)

Label (en): AmmoniumNitrogen

Comment (pt): uma das formas do nitrogênio no meio aquático. Pode ser utilizado como parâmetro para medir a qualidade de água.

Comment (en): one of the forms of nitrogen in the aquatic environment. It can be used as a parameter to measure the quality of water.

16. Expressão: Oxigênio Consumido

DataProperties

TopDataProperty

OxigenioConsumido

Anotações

Label (pt): Oxigênio Consumido

Label (en): Consumed Oxygen

AltLabel (pt): Consumo de Oxigênio

AltLabel (en): Oxygen Consumption

Comment (pt): medida empírica da quantidade de oxigênio consumido por microorganismos, na decomposição da matéria orgânica presente na água. Parâmetro para medir a qualidade de água.

Comment (en): empirical measure of the amount of oxygen consumed by microorganisms in the decomposition of organic matter present in water. A parameter to measure the quality of water.

17. Expressão: Barragem/Represa/Açude

Dimension

EnvironmentalDimension

Hydrosphere

BodyOfWater

Barragem

Anotações

Label (pt) Barragem

Label (en): Barrage

PrefLabel (pt): Represa

PrefLabel (en): Dam

AltLabel (pt): Açude

PrefLabel (en): Dam

Comment (pt): Uma barreira feita pelo homem construída por meio de um corpo de água com o propósito de captação de água para reservatório.

Comment (en): A manmade barrier built across a body of water for the purpose of impounding water in a reservoir.

18. Expressão: Reservatório de Água

Dimension

EnvironmentalDimension

Hydrosphere

WaterResources

ReservatorioDeAgua

Anotações

Label (pt): Reservatório de Água

Label (en): Waterreservoir

Comment (pt): unidades hidráulicas de acumulação e passagem de água, situadas em pontos estratégicos do sistema de modo a atenderem as seguintes situações: garantia da quantidade de água (demandas de equilíbrio, de emergência e de anti-incêndio); garantia de adução com vazão e altura manométrica constantes; menores diâmetros no sistema; e melhores condições de pressão.

Comment (en): hydraulic units of accumulation and transfer of water, located at strategic points of the system to meet the following conditions: ensuring the amount of water (balance of demands, emergency and fire prevention); adduction assurance with constant flow gauge and height; smaller diameters in the system; and best pressure conditions.

19. Expressão: Sólidos Dissolvidos Voláteis

Como conceito:

Definition

ChemicalDefinition

ChemicalCompound

Dissolved_Solids

Solidos_Dissolvidos_Volateis

Como quantidade:

Data Properties

TopDataProperty

totalDissolvedSolids / totalWaterSolids

Sólidos_Dissolvidos_Voláteis

Anotações

Label (pt): Sólidos Dissolvidos Voláteis

Label (en): SolidsVolatileDissolved

Comment (pt): parcela dos sólidos totais volatilizada no processo de calcinação, em mg/L;); Parâmetro para medir a qualidade de água.

Comment (en): volatilized portion of the total solids in the calcination process in mg / l;); A parameter to measure the quality of water.

20. Expressão: SólidosTotais

Data Properties

TopDataProperty

Sólidos_Totais

Anotações

Label (pt): Sólidos Totais

Label (en): Total Solids

Comment (pt): Matéria sólida que é suspensa, dissolvida, ou fixa num líquido, tal como água, águas residuais, leite, e são deixadas após evaporação e secagem de uma amostra.

Comment (en): Solid matter which is suspended, dissolved, or settled in a liquid, such as water, wastewater, Milk, and are left after evaporation and drying of a sample.

21. Expressão: População

Dimension

DimensaoHumana

Populacao

Anotações

Label (pt): População

Label (en): Population

Comment (pt): conjunto de seres vivos da mesma espécie que habitam uma certa região num certo momento.

Comment (en): set of organisms of the same species inhabiting a certain area at a certain time.