

UNIVERSIDADE DO EXTREMO SUL CATARINENSE – UNESC

CURSO DE CIÊNCIA DA COMPUTAÇÃO

JOEL DA ROSA

**CRITÉRIOS DE QUALIDADE SEMÂNTICA PARA AVALIAÇÃO DE
USO DE RECURSOS EM ONTOLOGIAS DE FUNDAMENTAÇÃO E
TESAUROS CONCEITUAIS**

CRICIÚMA

2012

JOEL DA ROSA

**CRITÉRIOS DE QUALIDADE SEMÂNTICA PARA AVALIAÇÃO DE
USO DE RECURSOS EM ONTOLOGIAS DE FUNDAMENTAÇÃO E
TESAUROS CONCEITUAIS**

Trabalho de Conclusão de Curso, apresentado para
obtenção do Grau de Bacharel em Ciência da
Computação na Universidade do Extremo Sul
Catarinense, UNESC.

Orientadora: Prof.^a MSc. Leila Laís Gonçalves.

CRICIÚMA

2012

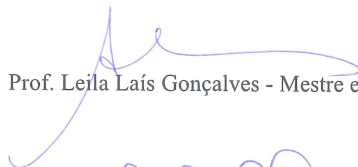
JOEL DA ROSA

**CRITÉRIOS DE QUALIDADE SEMÂNTICA PARA AVALIAÇÃO DE USO DE
RECURSOS EM ONTOLOGIAS DE FUNDAMENTAÇÃO E TESAuros
CONCEITUAIS**

Trabalho de Conclusão de Curso, aprovado pela
Banca Examinadora para obtenção do Grau de
Bacharel em Ciência da Computação na
Universidade do Extremo Sul Catarinense, UNESC.

Criciúma, 27 de novembro de 2012.

BANCA EXAMINADORA



Prof. Leila Laís Gonçalves - Mestre em ciências - (UNESC) – Orientador



Prof. Merisandra Côrtes de Mattos Garcia - Mestre em ciências – (UNESC)



Prof. Fabrício Giordani - Especialista - (UNESC)

Aos meus empregadores e ao Artigo 170 que proporcionaram o pagamento desta graduação.

AGRADECIMENTOS

Aos meus pais, que nunca me forçaram a estudar ou mesmo me graduar, mas que também nunca me disseram para desistir, que me ensinaram coisas valiosas para a minha vida toda, entre estas, obediência, honra, lealdade, simplicidade e respeito.

A todos os meus professores, que por meio de seus ensinamentos facilitaram a conclusão desta graduação, iniciando-se pela minha professora do pré-escolar, dona Sirlei.

A minha orientadora, Leila, por aceitar me orientar neste trabalho de conclusão de curso me guiando de forma paciente e didática nos momentos fáceis e difíceis.

Aos meus colegas que iniciaram a graduação comigo e que infelizmente não conseguiram estar presentes, entre estes, o Emanuel, vulgo Sasso, que o homem lá de cima solicitou a sua presença tão precocemente.

A minha veterana Camila Cechinel Silvestri, que no início desta graduação escreveu em minha testa a palavra “calouro”, me iniciando em um novo mundo, que separa as crianças dos adultos.

A minha amiga Fernanda Nunes Peruchi, que pegou o primeiro ônibus comigo rumo a esta graduação e que esteve comigo até os momentos finais deste trabalho de conclusão de curso, aturando minhas mensagens de celular sobre a finalização do mesmo.

Aos meus amigos, que me chamaram para o bar objetivando afogar as lágrimas, que me convidaram para as festas para curtir a vida adoidado, que estiveram comigo na vida louca, que entenderam os meus “não”, que não respeitaram os meus “não” e obrigaram a dizer “sim”, que me carregaram quando eu não podia andar, que brigaram ou fugiram ao meu lado.

Por último e não menos importante ao Facebook e ao Google, que estiveram sempre on-line, o primeiro fonte de distração durante os momentos difíceis deste trabalho e o segundo que me corrigiu quando necessário, que me forneceu conhecimentos extras, que me traduziu palavras incompreensíveis, que me disse onde estava o que eu procurava, ambos companheiros inseparáveis.

“I have a dream for the Web in which computers become capable of analyzing all the data on the Web.”

Tim Berners-Lee.

RESUMO

A Web Semântica objetiva a significação de conteúdos informacionais na Web, onde máquinas e humanos possam trabalhar em cooperação na interpretação dos dados, gerando interoperabilidade e semântica. Apesar da vasta publicação sobre este assunto, há pouca pesquisa no que se refere à qualidade da representação de conhecimento em páginas da Web. Este trabalho objetiva uma validação comparativa entre as formas de representação de conhecimento possibilitadas por ontologias de fundamentação e tesouros conceituais. Tanto a pesquisa bibliográfica quanto a validação prática buscaram apontar as características de qualidade semântica que estes objetos proporcionam para a redução de ambiguidade semântica na representação de conhecimento. A validação comparativa foi baseada em estudo de caso aplicado no domínio de conhecimento Nepomuk (indexador de área de trabalho do KDE). As etapas de validação consistiram na definição de tipos de ambiguidade semântica, identificação das características de qualidade semântica para redução da ambiguidade em ontologias de fundamentação e tesouros conceituais. Com os resultados da validação pode-se concluir a validade da comparação entre ontologias e tesouros visto que ambos apresentaram características ímpares para a qualidade semântica e conseqüentemente na redução de ambiguidade na representação do conhecimento. A partir dos resultados pode-se elencar critérios de qualidade semântica para uso em ontologias de fundamentação e tesouros conceituais no que tange a estruturação de informações. Foi possível também observar que os resultados na utilização de suas estruturas dependem da finalidade de representação, haja vista que o primeiro foca mais nas máquinas e o segundo mais em humanos, entretanto ambos realizam uma representação cooperativa entre máquinas e humanos, porém de maneiras diferentes.

Palavras-chave: Web Semântica. Ambiguidade. Qualidade Semântica. Ontologias de Fundamentação. Tesouros Conceituais.

ABSTRACT

The Semantic Web aims at the meaning of informational content to the Web, where machines and humans can work cooperatively in interpreting the data, generating interoperability and semantic. Despite the extensive publication on this subject, there is little research related to the quality of knowledge representation on Web pages. This work aims at a comparative validation forms of knowledge representation made possible by foundational ontology and conceptual thesauri. Both the literature as validation practice sought to point the semantics quality that these objects provide to reduce semantic ambiguity in the representation of a knowledge. The validation was based on comparative case study in the field of applied knowledge Nepomuk (indexer KDE desktop). The validation steps consisted in defining types of semantic ambiguity, identification of quality features to reduce semantic ambiguity in foundational ontology and conceptual thesauri. With the validation results we can conclude the validity of the comparison between ontologies and thesauri since both had unique characteristics to the semantic quality and consequently in reducing ambiguity in knowledge representation. From the results it can be to list semantic quality criteria for use in foundational ontologies and conceptual thesauri with respect to structuring information. It was also possible to observe that the results in the use of their structures depend on the purpose of the representation, given that the first focuses more in machines and the second more in humans, however both carry a representation cooperative between machines and humans, but of different ways.

Keywords: Semantic Web. Ambiguity. Semantic Quality. Foundational Ontologies. Conceptual Thesauri.

LISTA DE ILUSTRAÇÕES

Figura 1 - Arquitetura da Web Semântica.....	22
Figura 2 - Estrutura RDF em formato de grafo.....	25
Figura 3 - Estrutura RDF em XML.....	25
Figura 4 - Exemplo de RDF <i>Schema</i>	26
Figura 5 - Instância de RDF <i>Schema</i>	27
Figura 6 - Exemplo de ontologia em OWL 1.....	29
Figura 7 - Exemplo de ontologia em OWL 2.....	31
Figura 8 - RDF <i>Graph</i> de um conceito em SKOS.....	32
Figura 9 - Conceito com seus rótulos preferidos e alternativos.....	32
Figura 10 - Estrutura RDF com conceitos e seus rótulos preferidos e alternativos.....	33
Figura 11 - Ligações hierárquicas e associativas de um conceito.....	34
Figura 12 - Exemplo de consulta em SPARQL.....	35
Figura 13 - Consulta em SPARQL abreviada.....	35
Figura 14 - Diagrama da família de dialetos RIF.....	37
Figura 15 - Exemplo de RIF <i>Core</i>	37
Figura 16 - Tela do Protégé.....	39
Figura 17 - Tela do TemaTres.....	40
Figura 18 - Editor de consulta SPARQL.....	42
Figura 19 - Código implementado em SWP e seu resultado.....	43
Figura 20 - Utilização de RDFa em código HTML.....	46
Figura 21 - Utilização de schema.org em uma página da Web.....	47
Figura 22 - Busca no Wolfram Alpha baseada em um cálculo geográfico.....	48
Figura 23 - Resposta de uma pergunta no Evi.....	48
Figura 24 - Busca no Google utilizando <i>Knowledge Graph</i>	49
Figura 25 - Ontologia na publicação de uma notícia.....	52
Figura 26 - Exemplo e estrutura de um Mapa Conceitual.....	56
Figura 27 - Forma gráfica da definição de Ontologia proposta por Gruber (1993).....	57
Figura 28 - Parte do tesouro TRT.....	61
Figura 29 - Estrutura da ontologia SUMO.....	65
Figura 30 - Distinção ontológica em uma tipologia de universais.....	68

Figura 31 - Exemplo de ontologia de fundamentação com kinds, subkinds e phases.....	69
Figura 32 - Parte do Tesouro de Folclore e Cultura Popular Brasileira.....	71
Figura 33 - Camadas da qualidade da informação.....	78
Figura 34 - Relações entre conceitos.....	83
Figura 35 - Escopo de pastas do usuário delimitado.....	93
Quadro 1 - Recursos em uma página Web.....	24
Quadro 2 - Comparação de construções das linguagens de OWL 1.....	28
Quadro 3 - Comparativo das características das formas de representação de conhecimento...	63
Quadro 4 - Comparativo entre as características das ontologias a partir de sua evolução.....	67
Quadro 5 - Comparativo entre as características dos tesouros a partir de sua evolução.....	73
Quadro 6 - Requisitos de qualidade de dados importantes aos consumidores de dados.....	79
Quadro 7 - Critérios de qualidade dos dados para a Web Semântica.....	80
Quadro 8 - Solução para ambiguidade em ontologias de fundamentação e tesouros conceituais.....	94
Quadro 9 - Critérios de qualidade semântica para uso em ontologias de fundamentação e tesouros conceituais.....	98

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface</i>
BLD	<i>Basic Logic Dialect</i>
BT	<i>Broader Term</i>
BTP	<i>Broader Term Partitive</i>
CNFCP	Centro Nacional de Folclore e Cultura Popular
DC	<i>Dublin Core</i>
DL	<i>Description Logic</i>
DLP	<i>Description Logic Programs</i>
DOLCE	<i>Descriptive Ontology for Linguistic and Cognitive Engineering</i>
DTB	<i>Datatypes and Built-ins</i>
DTD	<i>Document Type Definition</i>
EL	<i>Expression Language</i>
FLD	<i>Framework for Logic Dialects</i>
FOAF	<i>Friend of a Friend</i>
GFO	<i>General Formal Ontology</i>
GRDDL	<i>Gleaning Resource Descriptions from Dialects of Languages</i>
HTML	<i>HyperText Markup Language</i>
IA	Inteligência Artificial
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
ISO	<i>International Organization for Standardization</i>
JSON	<i>JavaScript Object Notation</i>
KOS	<i>Knowledge Organization System</i>
NFO	<i>Nepomuk File Ontology</i>
NT	<i>Narrower Term</i>
NTP	<i>Narrower Term Partitive</i>
OODBMS	<i>Object-Oriented Database Management Systems</i>
OWL	<i>Web Ontology Language</i>
PHP	<i>PHP Hypertext Preprocessor</i>
PRD	<i>Production Rules Dialect</i>
QL	<i>Query Language</i>

RDBMS	<i>Relational Database Management System</i>
RDF	<i>Resource Description Framework</i>
RDFa	<i>Resource Description Framework in Attributes</i>
RIA	<i>Rich Internet Applications</i>
RIF	<i>Rule Interchange Format</i>
RL	<i>Rule Language</i>
RT	<i>Related Term</i>
SGML	<i>Standard Generalized Markup Language</i>
SKOS	<i>Simple Knowledge Organization System</i>
SMW	<i>Semantic MediaWiki</i>
SOA	<i>Service Oriented Architecture</i>
SOC	Sistemas de Organização do Conhecimento
SPARQL	<i>SPARQL Protocol and RDF Query Language</i>
SPIN	<i>SPARQL Inferencing Notation</i>
SUMO	<i>Suggested Upper Merged Ontology</i>
SWP	<i>SPARQL Web Pages</i>
TA	Termo Associado
TE	Termo Específico
TEP	Termo Específico Partitivo
TG	Termo Genérico
TGP	Termo Genérico Partitivo
TOVE	<i>Toronto Virtual Enterprise</i>
TRB	<i>Transportation Research Board</i>
TRT	<i>Transportation Research Thesaurus</i>
U	Uso
UF	<i>Used For</i>
UFO	<i>Unified Foundational Ontology</i>
UP	Usado Para
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
URN	<i>Uniform Resource Name</i>
W3C	<i>Wide Web Consortium</i>

WIQA	<i>Information Quality Assessment Framework</i>
XHTML	<i>eXtensible HyperText Markup Language</i>
XML	<i>eXtensible Markup Language</i>
XSD	<i>XML Schema Definition</i>
XSL	<i>eXtensible Stylesheet Language</i>

SUMÁRIO

1 INTRODUÇÃO.....	16
1.1 OBJETIVO GERAL.....	17
1.2 OBJETIVOS ESPECÍFICOS.....	17
1.3 JUSTIFICATIVA.....	18
1.4 ESTRUTURA DO TRABALHO.....	19
2 WEB SEMÂNTICA.....	20
2.2 ARQUITETURA DA WEB SEMÂNTICA.....	21
2.2.1 Camada de estruturação.....	23
2.2.2 Camada de sintaxe.....	23
2.2.3 Camada de interoperabilidade.....	24
2.2.4 Camada de semântica.....	25
2.2.5 Camada de lógica, prova, confiança e criptografia.....	38
2.2.6 Camada de interação.....	38
2.3 FERRAMENTAS PARA A WEB SEMÂNTICA.....	39
2.4 EXEMPLOS DE APLICAÇÕES NA WEB SEMÂNTICA.....	47
3 REPRESENTAÇÃO DE CONHECIMENTO NA WEB.....	53
3.1 FORMAS DE REPRESENTAÇÃO DE CONHECIMENTO NA WEB.....	54
3.1.1 Metadados.....	54
3.1.2 Mapas conceituais.....	56
3.1.3 Ontologias.....	57
3.1.4 Tesouros.....	59
3.1.5 Comparativo das formas de representação de conhecimento.....	62
3.2 ONTOLOGIAS DE FUNDAMENTAÇÃO.....	63
3.2.1 Exemplos de ontologias de fundamentação.....	64
3.2.2 Ferramentas para ontologias de fundamentação.....	65
3.2.3 Representação de conhecimento em ontologias de fundamentação.....	66
3.3 TESAUROS CONCEITUAIS.....	69
3.3.1 Exemplos de tesouros conceituais.....	70
3.3.2 Ferramentas para tesouros conceituais.....	72
3.3.3 Representação de conhecimento em tesouros conceituais.....	73

3.4 AMBIGUIDADE.....	74
3.5 QUALIDADE SEMÂNTICA.....	78
3.6 AMBIGUIDADE E QUALIDADE SEMÂNTICA EM ONTOLOGIAS DE FUNDAMENTAÇÃO E TESAUROS CONCEITUAIS.....	81
4 TRABALHOS CORRELATOS.....	84
4.1 TESAUROS CONCEITUAIS E ONTOLOGIAS DE FUNDAMENTAÇÃO: ANÁLISE COMPARATIVA ENTRE AS BASES TEÓRICO- METODOLÓGICAS UTILIZADAS EM SEUS MODELOS DE REPRESENTAÇÃO DE DOMÍNIOS.....	84
4.2 TESAUROS E ONTOLOGIAS SOB A LUZ DA TEORIA COMUNICATIVA DA TERMINOLOGIA.....	85
4.3 REPRESENTAÇÃO DE INFORMAÇÃO E CONHECIMENTO: ESTUDO DAS DIFERENTES ABORDAGENS ENTRE A CIÊNCIA DA INFORMAÇÃO E A CIÊNCIA DA COMPUTAÇÃO.....	86
4.4 TESAUROS E A WORLD WIDE WEB.....	87
4.5 DETECÇÃO E CORREÇÃO DE PROBLEMAS DE QUALIDADE DOS DADOS: MODELO, SINTAXE E SEMÂNTICA.....	87
4.6 CARACTERÍSTICAS DA QUALIDADE DE DADOS RELACIONADOS PUBLICADOS EM FONTES DE DADOS.....	88
5 VALIDAÇÃO COMPARATIVA DE QUALIDADE SEMÂNTICA NA REPRESENTAÇÃO DE CONHECIMENTO EM PÁGINAS DA WEB: UM ESTUDO DE CASO.....	90
5.1 METODOLOGIA.....	90
5.1.1 Levantamento bibliográfico.....	90
5.1.2 Estudo de caso.....	91
5.1.2.1 Domínio de conhecimento.....	92
5.1.2.2 Tipos de ambiguidade semântica.....	93
5.1.2.3 Características de qualidade semântica.....	93
5.1.2.4 Validação comparativa.....	94
5.1.2.5 Resultados da validação comparativa.....	95
5.1.3 Critérios de qualidade semântica para uso em ontologias de fundamentação e tesauros conceituais.....	98
6 CONCLUSÃO.....	99

REFERÊNCIAS.....	101
APÊNDICE A – Código fonte Ontologias de Fundamentação.....	110
APÊNDICE B – Código fonte Tesouros Conceituais.....	112
APÊNDICE C – Artigo.....	114

1 INTRODUÇÃO

A concepção da Web Semântica, idealizada por Tim Berners-Lee (2000, tradução nossa), é fundamentada na representação do conhecimento introduzindo estrutura e significado ao conteúdo Web, tendo como objetivo transformar uma rede de documentos e informações em uma rede de dados que potencialmente conecta conhecimento, passível de processamento e interpretação por máquinas. Para atender este objetivo, é necessário tecnologias para tratar a informação de forma detalhada, objetivando a descrição, representação e interpretação (DEVEDZIC, 2006, tradução nossa).

Um dos desafios da Web Semântica é criar uma estrutura que consiga expressar o significado de seu conteúdo e ao mesmo tempo estabelecer regras para processar esse significado de forma a inferir novos dados e regras. Algumas soluções têm sido desenvolvidas para representar conhecimento e anotar a informação contida na Web envolvendo linguagens documentárias e/ou modelos conceituais. Estas soluções são constituídas de sistemas simbólicos que visam descrever sinteticamente conteúdos documentais, e são utilizadas nos sistemas informacionais para indexação, armazenamento e recuperação da informação (GUIZZARDI, 2005, tradução nossa).

A qualidade semântica dos sistemas de representação do conhecimento é o grau de correspondência entre o modelo conceitual gerado e o domínio de conhecimento ao qual se refere, requerendo que o modelo seja ao mesmo tempo completo e válido. Portanto, além de estruturar a informação, é necessário que estes sistemas resolvam, entre outros problemas, a ambiguidade da linguagem natural que gera por exemplo a polissemia, onde, uma palavra ou termo que possui vários significados distintos resulta em dois conceitos com relação semântica entre os mesmos (MEDEIROS, 2011).

Dentre os sistemas para representação e organização do conhecimento com maior reconhecimento, destacam-se: tesouros e ontologias. O uso de ontologias em conteúdo Web, amplamente discutido na literatura, permite expressar regras e compartilhar representações conceituais. Entretanto, a criação de ontologias é um processo complexo e demorado, o que dificulta sua adoção (ISOTANI et al, 2008). Ontologias tem uma semântica formalizada, dirigidas a sistemas computacionais, mas ainda são pouco legíveis por humanos. Os tesouros têm aplicação na Web no contexto de estruturação e recuperação da informação, com a função de evitar ou diminuir a flexibilidade da linguagem e descrever um conceito de maneira

unívoca em um sistema de informação, orientando assim qual o termo mais adequado para representar um conceito (SALES, 2008). O emprego dos tesouros tem sido muito bem-sucedido, mas a ambiguidade, riqueza e capacidade de inovação constante das linguagens nas quais se expressam os documentos Web ainda ocasionam representação e recuperação de informação com a pertinência no domínio do conhecimento modelado (MURAKAMI, 2011).

Assim sendo, ontologias e tesouros por mais que sejam amplamente reconhecidos em suas respectivas áreas, Ciência da Computação e Ciência da Informação, não conseguem atender totalmente o objetivo quando o assunto é redução de ambiguidade na representação do conhecimento, ou seja, pouca concisão ou clareza, fazendo-se necessário utilizar outras formas de representação de conhecimento. Neste sentido apresentam-se ontologias de fundamentação e tesouros conceituais, evoluções de ontologias e tesouros respectivamente, que tem em sua natureza uma representação de conhecimento mais alinhada entre o modelo conceitual e o domínio de conhecimento empregado (MEDEIROS, 2011).

1.1 OBJETIVO GERAL

Estruturar critérios de qualidade semântica para avaliação de uso de recursos em ontologias de fundamentação e tesouros conceituais.

1.2 OBJETIVOS ESPECÍFICOS

O desenvolvimento deste trabalho baseia-se nos seguintes objetivos específicos:

- a) contextualizar a Web Semântica e sua contribuição na estruturação e representação do conhecimento em páginas da Web;
- b) identificar formas de representação de conhecimento em páginas da Web;
- c) realizar um estudo de caso para uma validação comparativa de características de qualidade semântica em ontologias de fundamentação e tesouros conceituais na estruturação de informações que objetivam a representação de conhecimento em páginas da Web.

1.3 JUSTIFICATIVA

Segundo Flemming (2011, tradução nossa) as pesquisas e trabalhos relacionados a Web Semântica desenvolvidos por pesquisadores e instituições destinaram-se basicamente a arquitetura, tecnologias e linguagens para a Web Semântica. Buscou-se o aprimoramento da arquitetura para a interoperabilidade entre agentes de *software* e nas próprias páginas da Web, porém, pouco objetivando na qualidade semântica na representação e recuperação do conhecimento nestes ambientes.

A representação e recuperação de conhecimento em páginas da Web é um dos problemas enfrentados pelos motores de busca, que recuperam informações de modo léxico, percorrendo a Web, indexando e aplicando algoritmos em busca de resultados, entretanto ignorando determinadas páginas da Web, páginas estas, pertencentes a Web “oculta”, páginas sem *links*, metadados e relacionamentos (BRANSKI, 2004), problema este que a Web Semântica visa resolver. Entretanto, as tecnologias e linguagens propostas possuem características e modos de aplicação distintas, dificultando na decisão e formalização de uma forma de representação, assim como na definição de uma qualidade semântica.

Entre as qualidades semânticas mais importantes nas formas de representação de conhecimento, destaca-se a diminuição na ocorrência de ambiguidade conceitual, principalmente em páginas da Web, onde tanto humanos quanto máquinas interpretam as informações. Precisa-se formalizar o conhecimento representando, de modo que tanto humanos quanto máquinas possam interpretar e compartilhar o mesmo conhecimento, trabalhando em conjunto na recuperação da informação (BERNERS-LEE et al, 2001, tradução nossa).

Segundo Silva (2006) a ambiguidade é algo que precisa ser resolvido e até mesmo eliminado, faltando estudos e soluções que satisfaçam este objetivo, principalmente por ser um problema que afeta várias áreas e de diferentes maneiras, portanto, necessita-se focar em resolver partes específicas da ambiguidade. Para Bräscher (2002) a ambiguidade semântica está diretamente ligada a representação de conhecimento, ainda mais quando há um contato constante entre máquinas e humanos na interação deste, onde a falta de informações na representação de conhecimento não a deixa concisa e clara quanto aos seus significados.

Alguns trabalhos como o de Moreira (2003) e Sales (2008) apresentam objetivos referentes a comparação entre ontologias e tesouros como formas de representação de

conhecimento, trabalhos como o de Bräscher (2002), Pagani (2009) e Silva (2006) relatam a ambiguidade na representação e recuperação de informação em páginas da Web, os trabalhos de Flemming (2011) e Oliveira (2008) trabalham com a qualidade semântica dos dados e o trabalho de Medeiros (2011) compara as metodologias de construção entre ontologias de fundamentação e tesouros conceituais, sendo que nenhum destes trabalhos estudam a qualidade semântica que reduz a ambiguidade na representação de conhecimento quando se trata de ontologias de fundamentação e tesouros conceituais.

Ontologias de fundamentação e tesouros conceituais são formas de representação de conhecimento que apresentam características de qualidade semântica diferenciadas, que ajudam a reduzir a ambiguidade conceitual, principalmente em virtude de possuírem sua representação orientada a categorias e conceitos e não somente ao termo. Entretanto, essas características se apresentam de maneiras diferentes nas duas formas de representação de conhecimento, necessitando-se assim de uma validação comparativa entre as mesmas afim de validar as características de qualidade semântica que reduzem a ambiguidade conceitual na estruturação de informações com foco na representação de conhecimento.

1.4 ESTRUTURA DO TRABALHO

Este trabalho é composto por seis capítulos, iniciando-se pela introdução do mesmo, problema, objetivos e justificativa.

No capítulo 2 é conceituada a Web Semântica, sua arquitetura, ferramentas e aplicações.

A representação de conhecimento na Web é apresentada no capítulo 3, definindo-se as suas formas de representação, ambiguidade e qualidade semântica.

O capítulo 4 é composto pelos trabalhos correlatos do levantamento bibliográfico.

O estudo de caso e a análise dos resultados deste trabalho referentes a validação comparativa entre ontologias de fundamentação e tesouros conceituais, assim como a definição de critérios de qualidade semântica são realizados no capítulo 5.

A conclusão deste trabalho e sugestões para trabalhos futuros está inserida no capítulo 6.

2 WEB SEMÂNTICA

A Web Semântica é o projeto que visa concretizar a Web 3.0, a Web que está sucedendo a Web 2.0, que por sua vez sucedeu a Web 1.0. A primeira versão da Web surgiu no início dos 90 e constituía-se de páginas da Web simples, desenvolvidas e publicadas por programadores, contendo apenas textos, figuras, arquivos e *hiperlinks*, estes últimos, ligando páginas da Web, figuras e/ou arquivos (BREITMAN, 2005). Ainda segundo Breitman (2005), a Web 2.0 veio para popularizar a internet e conseqüentemente as páginas da Web, fazendo com que pessoas sem conhecimento em programação pudessem publicar informações por meio de páginas administrativas e até criarem suas próprias páginas, publicando-as na internet.

Com o advento das redes sociais a quantidade de publicações aumentou exponencialmente, explodindo o uso da segunda versão da Web, gerando segundo o *World Wide Web Consortium* (W3C) Brasil (2011), a “Web dos Documentos”. Berners-Lee et al (2001, tradução nossa) classificam a Web 2.0 como a “Web Sintática”, onde computadores recuperam e mostram informações e as pessoas as interpretam. Breitman (2005) comenta que os conteúdos semânticos das páginas sintáticas são codificados de maneira acessível apenas as pessoas e cita alguns problemas que motores de buscas encontram dentro da Web Sintática:

- a) muitas páginas encontradas, porém pouco precisas;
- b) resultados são sensíveis ao termo procurado;
- c) páginas encontradas não se relacionam com outras páginas.

A Web 2.0 com uso de *tags* para marcar informação em conjunto com algoritmos dos motores de buscas tentam melhorar a recuperação da informação, entretanto as páginas continuam com pouca semântica. Segundo Berners-Lee et al (2001, tradução nossa) a Web Semântica é a extensão da Web Sintática, onde as páginas da Web são complementadas por metadados que organizam e concedem significado, ou seja, semântica as informações publicadas. Essas informações podem ser transformadas em uma rede de dados que potencialmente conecta conhecimento, este, podendo ser compartilhado e interpretado por humanos e máquinas (DEVEDZIC, 2006, tradução nossa).

O conhecimento obtido em páginas semânticas pode nos ajudar com tarefas rotineiras e/ou complexas, como eliminar informações inúteis em um resultado de uma busca ou até mesmo descobrir o caixa eletrônico mais próximo. Berners-Lee et al (2001, tradução

nossa) ilustram o seguinte cenário possibilitado pelo uso da Web Semântica:

- a) Lucy precisa marcar uma consulta médica com um ortopedista e uma série de sessões de fisioterapia para sua mãe. Como ela vai ter de levar sua mãe às consultas, é necessário que estas sejam marcadas em um horário em que Lucy esteja livre, de preferência em um local perto da casa de sua mãe. Tanto o médico quanto os fisioterapeutas devem ser qualificados e fazer parte do plano de saúde da família;
- b) Lucy requisita ao seu agente (que funciona na Web Semântica) uma melhor solução de horário para a consulta;
- c) o agente recupera o tratamento prescrito à mãe de Lucy por meio do agente do médico que está cuidando dela;
- d) o agente procura em várias listas de provedores de serviços médicos e verifica aqueles que fazem parte do plano de saúde da família, que ficam dentro de um raio de 2 Km da casa da mãe de Lucy e que estão classificados como bons profissionais em um serviço de classificação de profissionais de saúde;
- e) o agente tenta achar combinações entre os horários disponíveis da agenda de Lucy e os horários vagos dos profissionais, disponibilizados por seus agentes e/ou páginas da Web e realiza a marcação da consulta.

Ramalho et al (2007) comentam que desde 1998 Tim Berners-Lee impulsiona estudos sobre a Web Semântica, assim como liderando o consórcio W3C, responsável por desenvolver padrões para a criação e interpretação de conteúdo na Web. A visão do W3C sobre a Web Semântica é uma “Web dos Dados *Linkados*”, onde tecnologias realizam a marcação semântica nas páginas e garantem interoperabilidade e poder de raciocínio na Web (W3C Brasil, 2011).

2.2 ARQUITETURA DA WEB SEMÂNTICA

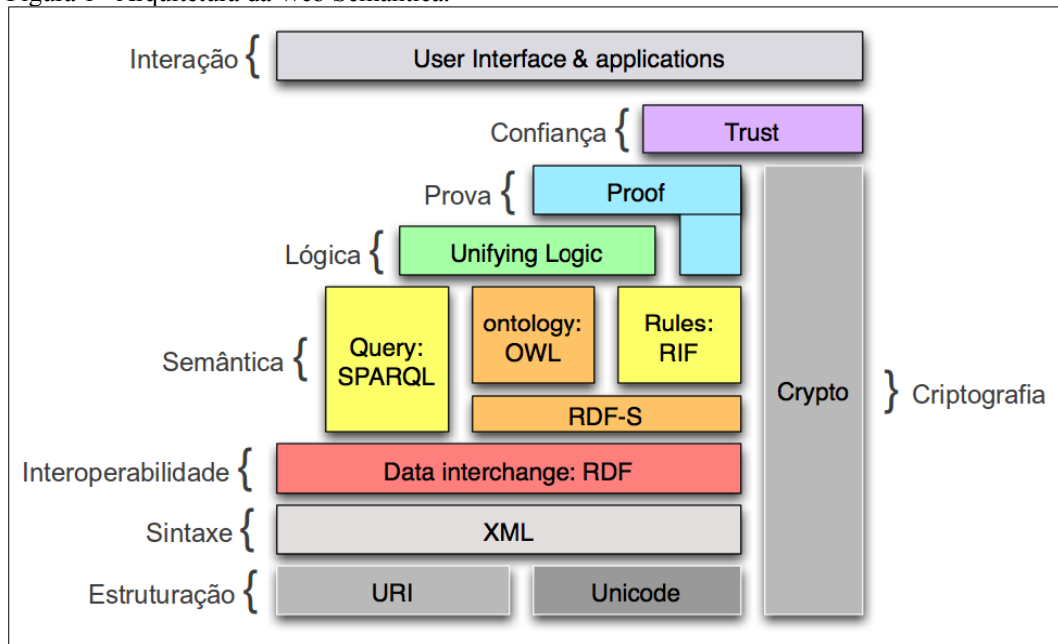
O funcionamento da Web Semântica depende de um conjunto de tecnologias que se relacionam entre si, porém de maneira escalonável, estruturadas de modo que uma ou mais tecnologias contemplem uma determinada camada. Cada camada é responsável por um passo ou requisito da Web Semântica (RAMALHO et al, 2007). A primeira proposta de arquitetura para a Web Semântica foi demonstrada em 2000 por Tim Berners-Lee, sendo que a ideia não

foi construir algo totalmente novo e sim uma complementação das tecnologias base da internet, listadas por Breitman (2005) a seguir:

- a) *Uniform Resource Identifier* (URI);
- b) *UNICODE*;
- c) *eXtensible Markup Language* (XML);
- d) *HyperText Markup Language* (HTML).

Desde então outras três propostas foram apresentadas pelo próprio Berners-Lee com a finalidade de contemplar todos os conceitos e responsabilidades que a Web Semântica possui, onde entre a primeira e última proposta no ano de 2009 houveram várias e importantes modificações, como a inclusão, união, separação e expressividade de camadas, assim como a inclusão de tecnologias recomendadas pelo W3C. A figura 1 apresenta a arquitetura proposta por Berners-Lee em 2009, com as camadas definidas e tecnologias já recomendadas oficialmente pelo W3C.

Figura 1 - Arquitetura da Web Semântica.



Fonte: Adaptado de Berners-Lee (2009).

Segundo Berners-Lee (2009), cada camada da arquitetura da Web Semântica pode ser vista como um conceito de uma função da Web Semântica, divididas em estruturação, sintaxe, interoperabilidade, semântica, lógica, prova, confiança, criptografia e interação.

2.2.1 Camada de estruturação

Unicode e URI são os padrões responsáveis pela base e estruturação da arquitetura da Web Semântica, garantindo uma maneira segura de armazenamento e transmissão de informações (RAMALHO et al, 2007). *Unicode* é um padrão industrial que fornece um número único para cada caractere, possibilitando assim máquinas manipularem-os não importando a plataforma, programa ou língua (*THE UNICODE CONSORTIUM*, 2011).

URI é um padrão internacional que padroniza e identifica unicamente um conjunto de caracteres (*Unicode*), ou seja, um recurso em uma página da web por meio de um endereço (COATES, 2001, tradução nossa). Mealling e Denenberg (2002) classificam URI em duas classes, *Uniform Resource Locator* (URL) e *Uniform Resource Name* (URN), onde os endereços são iniciados com “http:” e ”isbn:” respectivamente.

2.2.2 Camada de sintaxe

O padrão XML é o responsável pela sintaxe da formalização e descrição dos recursos de uma página da Web, sendo um simples e flexível formato de estruturação de textos, derivado do *Standard Generalized Markup Language* (SGML), desenvolvido originalmente para estruturar publicações eletrônicas, se tornando muito importante também na troca de informações na Web e em outras tecnologias computacionais (QUIN, 2012, tradução nossa).

O XML é composto por elementos e atributos, transformados em *tags* HTML por meio da linguagem *eXtensible Stylesheet Language* (XSL), estas *tags* são descritas em um documento *Document Type Definition* (DTD), isto torna o XML uma linguagem flexível, já que uma determinada comunidade de desenvolvedores podem estruturar suas informações de acordo com seus propósitos (SOUZA; ALVARENGA, 2004).

Souza e Alvarenga (2004) ainda enfatizam que essa flexibilidade do XML pode trazer ambiguidades de conceitos entre os elementos e atributos, sendo a solução para isto o uso de *Namespaces*, que são uma identificação única de elementos e atributos de um documento XML por meio de uma referência URI, tornando-se um método simples para a qualificação e validação dos mesmos (BRAY et al, 2009, tradução nossa).

2.2.3 Camada de interoperabilidade

Segundo o W3C Brasil (2011) a Web Semântica é um repositório de dados na Web, onde pessoas e máquinas poderão interagir com estes dados como em um banco de dados normal com tabelas e relacionamentos. Para isso ser possível é preciso de uma tecnologia que possua uma forma de modelar e relacionar os dados, sendo esta, a *Resource Description Framework* (RDF), que fornece uma estrutura para representar informações sobre recursos em páginas da Web por meio de propriedades, estas, correspondentes aos tradicionais pares atributo-valor. Breitman (2005) esclarece que qualquer objeto em uma página da Web pode ser um recurso, desde que possua uma URI. O quadro 1 exemplifica possíveis recursos de uma página da web com objetivo de serem representados como propriedades em uma estrutura RDF.

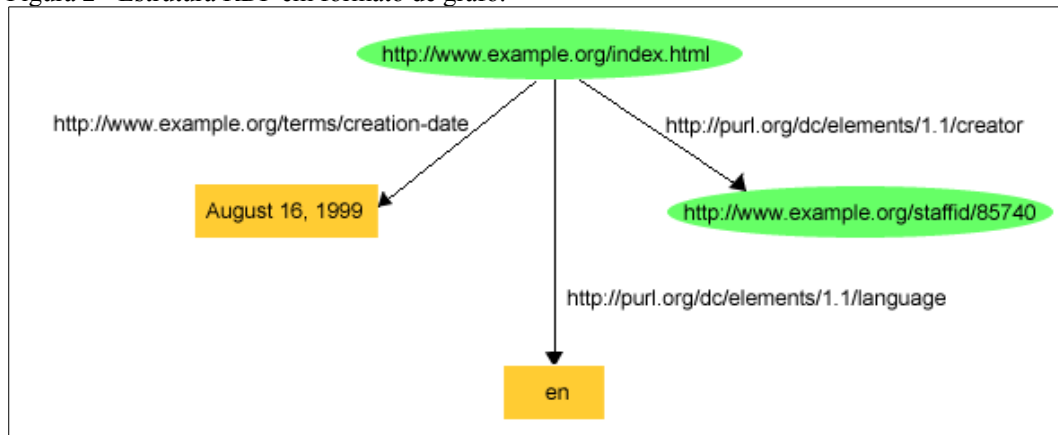
Quadro 1 - Recursos em uma página Web.

Recurso	Propriedade	Valor
http://www.example.org/index.html	http://www.example.org/terms/ creation-date	“August 16, 1999”
http://www.example.org/index.html	http://purl.org/dc/elements/1.1/ language	“en”
http://www.example.org/index.html	http://purl.org/dc/elements/1.1/ creator	“http://www.example.org/staffid/ 85740”
http://www.example.org/staffid/85740	http://www.example.org/terms/ name	“John Smith”
http://www.example.org/staffid/85740	http://www.example.org/terms/ age	“27”

Fonte: Do Autor.

Estruturas RDF são usadas principalmente nos casos em que informações precisam ser processadas e interpretadas por máquinas ao invés de serem somente exibidas para humanos (MANOLA; MILLER, 2004, tradução nossa). Brickley e Guha (2004, tradução nossa) enfatizam que a RDF amplia a ligação de recursos, permitindo que aplicações diferentes compartilhem a mesma informação, haja vista que essa é uma das fortes características da RDF, a interoperabilidade entre sistemas. Na figura 2 tem-se um exemplo de uma estrutura RDF representada por meio de um grafo RDF, contendo as informações de uma página da Web por meio das propriedades: data de criação, língua e autor.

Figura 2 - Estrutura RDF em formato de grafo.



Fonte: Manola e Miller (2004).

A figura 3 demonstra a mesma estrutura RDF, porém representada em linguagem XML.

Figura 3 - Estrutura RDF em XML.

```

1 <?xml version="1.0"?>
2 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3     xmlns:dc="http://purl.org/dc/elements/1.1/"
4     xmlns:exterm="http://www.example.org/terms/">
5
6   <rdf:Description rdf:about="http://www.example.org/index.html">
7     <exterm:creation-date>August 16, 1999</exterm:creation-date>
8     <dc:language>en</dc:language>
9     <dc:creator rdf:resource="http://www.example.org/staffid/85740" />
10  </rdf:Description>
11
12 </rdf:RDF>

```

Fonte: Adaptado de Manola e Miller (2004).

2.2.4 Camada de semântica

Nas camadas anteriores se tem uma base para uma construção semântica, mas é nesta camada que necessita-se de um vocabulário que permita descrever os aspectos semânticos dos recursos, criando relações de conceitos entre os mesmos, possibilitando a geração de significado para cada recurso em uma página da Web (RAMALHO et al, 2007).

Segundo Breitmam (2005), RDF fornece apenas um modo padronizado de utilizar a linguagem XML na representação de recursos e suas relações em uma página da Web, com isso se tem a necessidade de algo a mais, este sendo o RDF-S ou RDF *Schema*, uma especificação que define a sintaxe de um vocabulário para uma RDF (BRICKLEY; GUHA, 2004, tradução nossa), possibilitando a definição de taxonomias de recursos em termos de

uma hierarquia de classes (RAMALHO et al, 2007).

Brickley e Guha (2004, tradução nossa) descrevem a *RDF Schema* como uma extensão semântica da RDF, pois somente a RDF não é capaz de descrever as propriedades de uma estrutura RDF, sendo este o papel da *RDF Schema*, que define classes e propriedades que podem ser usados para descrever outras classes, propriedades e recursos e a própria relação entre os recursos de uma página. Manola e Miller (2004) exemplificam nas figuras 4 e 5 a definição e utilização de uma *RDF Schema*.

Na figura 4 se tem a estrutura da página, com uma hierarquia de classes sobre veículos e pessoas, assim como propriedades que interagem com estas classes.

Figura 4 - Exemplo de RDF Schema.

```

1 <?xml version="1.0"?>
2 <!DOCTYPE rdf:RDF [

```

Fonte: Adaptado de Manola e Miller (2004).

A figura 5 exemplifica a utilização da *RDF Schema* por meio de uma instância de um veículo representada em uma estrutura RDF.

Figura 5 - Instância de RDF Schema.

```

1 <?xml version="1.0"?>
2 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3     xmlns:ex="http://example.org/schemas/vehicles#"
4     xml:base="http://example.org/things">
5
6 <ex:PassengerVehicle rdf:ID="johnSmithsCar">
7   <ex:driver rdf:resource="http://www.example.org/staffid/85740"/>
8   <ex:rearSeatLegRoom rdf:datatype="xsd:integer">52</ex:rearSeatLegRoom>
9 </ex:PassengerVehicle>
10
11 </rdf:RDF>

```

Fonte: Adaptado de Manola e Miller (2004).

Ainda na camada semântica tem-se as ontologias, representadas pela linguagem *Web Ontology Language* (OWL), uma extensão da *RDF Schema* que possibilita uma forte marcação semântica para publicação e compartilhamento de ontologias na web (DEAN; SCHREIBER, 2004). OWL vem logo acima da *RDF Schema* na pilha de tecnologias da arquitetura da Web Semântica, adicionando um vocabulário mais completo para descrever classes e propriedades, assim como suas relações de disjunção, cardinalidade, igualdade, entre outras (MCGUINNESS; HARMELEN, 2004, tradução nossa). Breitmam (2005) resume OWL nas seguintes funções:

- a) criar ontologias;
- b) explicitar conceitos e propriedades fornecendo informações sobre ambos;
- c) explicitar fatos sobre um determinado domínio;
- d) racionalizar sobre ontologias e fatos.

A OWL 1, recomendada oficialmente pelo W3C em 2004, pode ser utilizada por meio de uma de suas três linguagens, da mais expressiva a menos expressiva: *OWL Full*, *OWL Description Logic* (DL) e *OWL Lite*, comentadas a seguir por Dean e Schreiber (2004, tradução nossa).

OWL Full é a utilização por completo de OWL 1 contendo todos os recursos e proporcionando uma utilização livre, sem restrições de construções RDF, isto é, não impondo uma separação de classes, propriedades e valores em um documento, entretanto não possuindo garantias computacionais, ou seja, não é possível realizar inferências.

OWL DL tem a mesma expressividade da *OWL Full*, porém impõe restrições em suas construções, mantendo uma completude computacional. OWL DL utiliza-se da lógica descritiva para conceder poder de raciocínio a Web Semântica, neste caso, capacidade de inferência.

OWL Lite é a utilização mais básica de OWL 1, possuindo pouca expressividade e

utilizando-se de apenas um sub-conjunto de construções da linguagem. O objetivo da *OWL Lite* é criar ontologias que possuam uma hierarquia de classificação e restrições simples, sendo um caminho mais curto na migração para tesouros e outras taxonomias. O resultando do uso da *OWL Lite* é uma baixa complexidade de desenvolvimento e ao mesmo tempo a mesma capacidade de restrições e raciocínio da *OWL DL*.

O quadro 2 compara as três linguagens de OWL 1 no que se refere a suas construções.

Quadro 2 - Comparação de construções das linguagens de OWL 1.

Construções	OWL Full	OWL DL	OWL Lite
Características RDF Schema: <i>Class (Thing, Nothing), rdfs:subClassOf, rdf:Property, rdfs:subPropertyOf, rdfs:domain, rdfs:range, Individual.</i>	X	X	X
Igualdade: <i>equivalentClass, equivalentProperty, sameAs, differentFrom, AllDifferent, distinctMembers.</i>	X	X	X
Características de propriedades: <i>ObjectProperty, DatatypeProperty, inverseOf, TransitiveProperty, SymmetricProperty, FunctionalProperty, InverseFunctionalProperty.</i>	X	X	X
Restrições de propriedades: <i>Restriction, onProperty, allValuesFrom, someValuesFrom.</i>	X	X	X
Informações de cabeçalho: <i>Ontology, imports.</i>	X	X	X
Intersecção de classes: <i>intersectionOf.</i>	X	X	X
Versionamento: <i>versionInfo, priorVersion, backwardCompatibleWith, incompatibleWith, DeprecatedClass, DeprecatedProperty.</i>	X	X	X
Propriedades de anotação: <i>rdfs:label, rdfs:comment, rdfs:seeAlso, rdfs:isDefinedBy, AnnotationProperty, OntologyProperty.</i>	X	X	X
Tipos de dados: <i>xsd datatypes.</i>	X		
Cardinalidade: <i>minCardinality, maxCardinality, cardinality.</i>	X (≥ 0)	X (≥ 0)	X (0 ou 1)
Axiomas de classe: <i>oneOf, dataRange, disjointWith, equivalentClass, rdfs:subClassOf.</i>	X	X	
Combinações booleanas de expressões de classe: <i>unionOf, complementOf, intersectionOf.</i>	X	X	
Informação de preenchimento: <i>hasValue.</i>	X	X	

Fonte: Adaptado de McGuinness e Harmelen (2004).

McGuinness et al (2004, tradução nossa) esclarecem que a estrutura de uma ontologia em OWL 1 possui alguns elementos básicos, divididos em *namespaces*, cabeçalho, classes, indivíduos, propriedades, restrições simples e restrições existenciais. A figura 6 exemplifica um trecho de uma ontologia desenvolvida em OWL 1 representada em uma estrutura RDF dentro de um documento XML.

Figura 6 - Exemplo de ontologia em OWL 1.

```

1 <?xml version="1.0"?>
2 <rdf:RDF xmlns = "http://www.w3.org/TR/2003/PR-owl-guide-20031209/wine#"
3     xmlns:vin = "http://www.w3.org/TR/2003/PR-owl-guide-20031209/wine#"
4     xml:base = "http://www.w3.org/TR/2003/PR-owl-guide-20031209/wine#"
5     xmlns:food= "http://www.w3.org/TR/2003/PR-owl-guide-20031209/food#"
6     xmlns:owl = "http://www.w3.org/2002/07/owl#"
7     xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
8     xmlns:rdfs = "http://www.w3.org/2000/01/rdf-schema#"
9     xmlns:xsd = "http://www.w3.org/2001/XMLSchema#">
10
11 <owl:Ontology rdf:about="">
12   <rdfs:comment>An example OWL ontology</rdfs:comment>
13   <owl:priorVersion>
14     <owl:Ontology rdf:about="http://www.w3.org/TR/2003/CR-owl-guide-20030818/wine"/>
15   </owl:priorVersion>
16   <owl:imports rdf:resource="http://www.w3.org/TR/2003/PR-owl-guide-20031209/food"/>
17   <rdfs:label>Wine Ontology</rdfs:label>
18 </owl:Ontology>
19
20 <owl:Class rdf:ID="Wine">
21   <rdfs:subClassOf rdf:resource="&food;PotableLiquid" />
22   <rdfs:subClassOf>
23     <owl:Restriction>
24       <owl:onProperty rdf:resource="#hasMaker" />
25       <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
26     </owl:Restriction>
27   </rdfs:subClassOf>
28   <rdfs:label xml:lang="en">wine</rdfs:label>
29   <rdfs:label xml:lang="fr">vin</rdfs:label>
30 </owl:Class>
31 ...
32 <owl:ObjectProperty rdf:ID="hasMaker">
33   <rdfs:type rdf:resource="&owl;FunctionalProperty" />
34 </owl:ObjectProperty>
35 ...
36 </rdf:RDF>

```

Fonte: Adaptado de McGuinness et al (2004).

OWL 2, recomendada oficialmente pelo W3C em 2009, é uma extensão e uma revisão de OWL 1, mais precisamente de OWL DL e OWL *Full*, possuindo construções bastante semelhantes, modificando-se basicamente algumas nomenclaturas, mantendo total compatibilidade com OWL 1 e agregando novas funcionalidades, principalmente nas construções de propriedades (W3C OWL *Working Group*, 2009, tradução nossa).

OWL 2 é definida a partir de tipos de dados descritos na linguagem XML *Schema Definition* (XSD), sendo esta uma dependência para OWL 2 diferentemente de OWL 1 (HITZLER et al, 2009, tradução nossa). Entre as novas funcionalidades disponíveis em OWL 2, Golbreich et al (2009) destacam as que aumentam a expressividade em OWL:

- a) definição de chaves;
- b) definição de cadeias de propriedades;
- c) tipos de dados mais ricos;
- d) intervalos de dados;
- e) restrições de cardinalidade qualificadas;
- f) propriedades assimétricas, reflexivas e disjuntas;
- g) capacidade de anotação melhorada.

Fokoue et al (2009, tradução nossa) explicam que assim como em OWL 1, OWL 2 também é dividida em três linguagens, classificadas nos perfis OWL 2 *Expression Language* (EL), OWL 2 *Query Language* (QL) e OWL 2 *Rule Language* (RL), que podem ser adotados dependendo do problema a ser resolvido.

OWL 2 EL é um perfil útil em aplicações que empregam ontologias que contêm várias propriedades e/ou classes e que requerem poder de expressividade. OWL 2 QL é voltado para aplicações que utilizam um vasto volume de dados de instância, e onde o atendimento de consultas é a tarefa mais importante de raciocínio, projetado para que os dados (afirmações) que são armazenados em um sistema de banco de dados possam ser consultados por meio de uma ontologia.

OWL 2 RL é útil para aplicações que exigem raciocínio escalável sem sacrificar demais a força expressiva, inspirado em *Description Logic Programs* (DLP). A figura 7 exemplifica um trecho de uma ontologia desenvolvida em OWL 2.

Figura 7 - Exemplo de ontologia em OWL 2.

```

1 Prefix(:=<http://example.com/owl/families/>)
2 Prefix(otherOnt:=<http://example.org/otherOntologies/families/>)
3 Prefix(xsd:=<http://www.w3.org/2001/XMLSchema#>)
4 Prefix(owl:=<http://www.w3.org/2002/07/owl#>)
5 Ontology(<http://example.com/owl/families>
6   Import( <http://example.org/otherOntologies/families.owl> )
7   Declaration( NamedIndividual( :John ) )
8   ...
9   Declaration( Class( :Person ) )
10  AnnotationAssertion( rdfs:comment :Person "Represents the set of all people." )
11  ...
12  Declaration( Class( :Man ) )
13  ...
14  Declaration( ObjectProperty( :hasBrother ) )
15  ...
16  Declaration( DataProperty( :hasAge ) )
17  ...
18  Declaration( Datatype( :personAge ) )
19  ...
20  DataPropertyRange( :hasAge xsd:nonNegativeInteger )
21  ...
22  FunctionalDataProperty( :hasAge )
23  ...
24  SubClassOf(
25    :Father
26    ObjectIntersectionOf( :Man :Parent )
27  )
28  ...
29  EquivalentClasses(
30    :MyBirthdayGuests
31    ObjectOneOf( :Bill :John :Mary )
32  )
33  ...
34  ClassAssertion( :Father :John )
35  ...
36  DataPropertyAssertion( :hasAge :John "51"^^xsd:integer )
37  ...
38  DifferentIndividuals( :John :Bill )
39 )

```

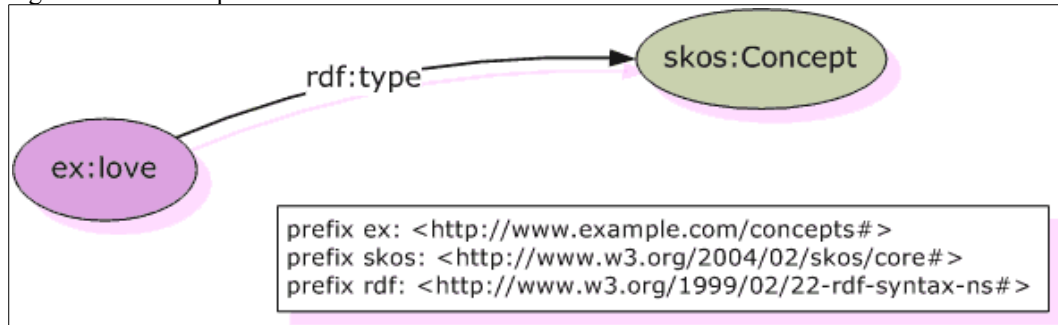
Fonte: Adaptado de Hitzler et al (2009).

A OWL 2 fez com que a expressividade de OWL aumentasse consideravelmente, entretanto perdeu-se a simplicidade de OWL Lite no que se diz respeito a representação de taxonomias, tesouros, esquemas de classificação e vocabulários controlados em páginas da Web. Para suprir esta lacuna o W3C recomendou oficialmente em 2009 o *Simple Knowledge Organization System* (SKOS), um padrão que define um modelo de dados legível por máquinas que facilita portar sistemas de organização de conhecimento simples para a Web Semântica.

SKOS é modelado em uma estrutura RDF que fornece uma linguagem leve e intuitiva, podendo ser usada por si só ou em conjunto com a OWL (BECHHOFFER; MILES, 2009, tradução nossa). Segundo Isaac e Summers (2009, tradução nossa) o principal elemento de um vocabulário SKOS é o conceito, representado por uma unidade de pensamento, ideia, significado ou objeto e evento, todos considerados como entidades abstratas na organização do conhecimento, sendo estes representados por meio de uma URI em uma estrutura RDF.

Brickley e Miles (2005, tradução nossa) afirmam que um recurso em uma página na Web é um recurso conceitual, ou seja o recurso em si é um conceito, conforme exemplificado na figura 8.

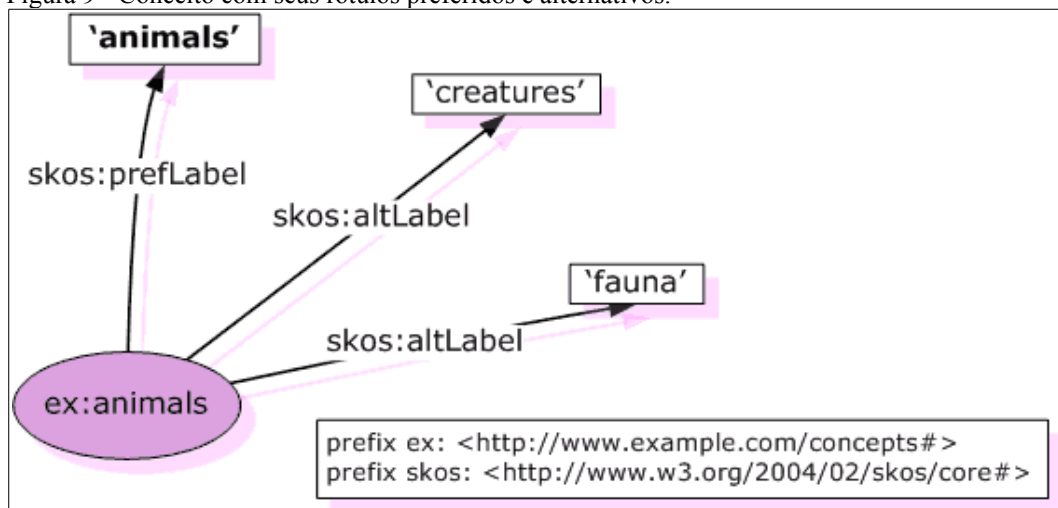
Figura 8 - RDF Graph de um conceito em SKOS.



Fonte: Brickley e Miles (2005).

Conforme Isaac e Summers (2009, tradução nossa) as primeiras caracterizações de um conceito são seus rótulos, expressões em linguagem natural que anexam e distinguem valor a um recurso conceitual. Existem três rótulos: *prefLabel* (rótulo preferido), *altLabel* (rótulo alternativo) e *hiddenLabel* (rótulo oculto), implicações que vão desde uma forte relação de denotação, unívoca, para uma informação que ajuda na pesquisa. Na figura 9 há um exemplo de um um conceito e seus rótulos preferidos e alternativos.

Figura 9 - Conceito com seus rótulos preferidos e alternativos.



Fonte: Brickley e Miles (2005).

Enquanto que na figura 10 o mesmo exemplo é demonstrado agora em uma estrutura RDF.

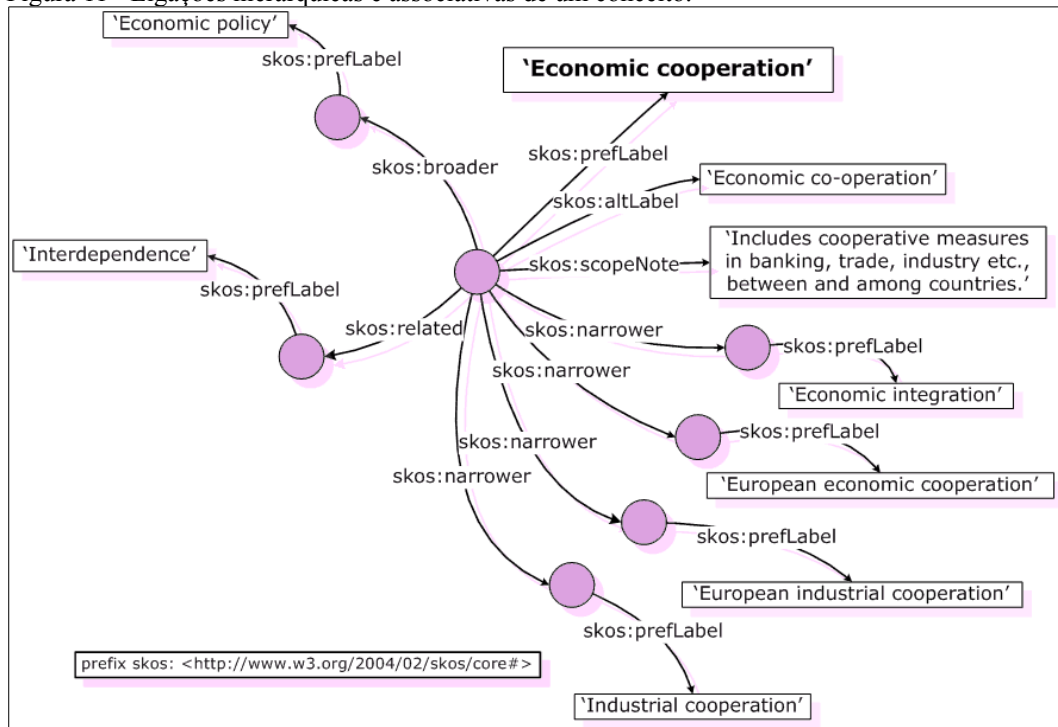
Figura 10 - Estrutura RDF com conceitos e seus rótulos preferidos e alternativos.

```
1 <rdf:RDF
2   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3   xmlns:skos="http://www.w3.org/2004/02/skos/core#">
4
5   <skos:Concept rdf:about="http://www.example.com/concepts#animals">
6     <skos:prefLabel>animals</skos:prefLabel>
7     <skos:altLabel>creatures</skos:altLabel>
8     <skos:altLabel>fauna</skos:altLabel>
9   </skos:Concept>
10
11 </rdf:RDF>
```

Fonte: Adaptado de Brickley e Miles (2005).

Bechhofer e Miles (2009, tradução nossa) explicam que os conceitos em um SKOS podem ser ligados a outros conceitos via propriedades de relações semânticas de forma hierárquica e/ou associativa, onde a ligação é inerente ao significado dos conceitos vinculados. A relação hierárquica entre dois conceitos indica que se está, de alguma forma mais amplo (*broader*) do que outro, mais restrito (*narrower*). Enquanto que em uma relação associativa os dois conceitos são inerentemente ligados (*related*), porém não sendo mais amplo ou mais restrito que o outro. Isaac e Summers (2009, tradução nossa) esclarecem que as relações semânticas de um SKOS desempenham um papel crucial para a definição de conceitos, onde o significado não é apenas definido pelas palavras em linguagem natural e seus rótulos e sim também por suas ligações com outros conceitos do vocabulário. A figura 11 demonstra um exemplo de ligações hierárquicas e associativas na representação de um conceito.

Figura 11 - Ligações hierárquicas e associativas de um conceito.



Fonte: Brickley e Miles (2005).

No entanto além de relações semânticas necessita-se também de uma documentação informal, tal como notas de escopo e definições. Bechhofer e Miles (2009, tradução nossa) descrevem que notas são usadas para fornecer informações relativas ao conceito não havendo nenhuma restrição, podendo ser um texto simples, hipertexto ou até uma imagem.

Isaac e Summers (2009, tradução nossa) destacam que conceitos podem ser criados e utilizados como entidades autônomas, no entanto, nas práticas de indexação os conceitos geralmente vêm em vocabulários cuidadosamente compilados, como os tesouros ou esquemas de classificação, necessitando estarem dentro de um esquema de conceitos e classificam as notas a descrições nos seguintes itens:

- a) *note* (notas de documentação);
- b) *scopeNote* (significado parcial pretendido do conceito);
- c) *definition* (significado completo pretendido do conceito);
- d) *example* (exemplo de aplicação do conceito);
- e) *historyNote* (histórico de modificações de significado do conceito);
- f) *editorialNote* (avisos, lembretes e outras informações administrativas);
- g) *changeNote* (notas de modificações).

Brickley e Miles (2005, tradução nossa) complementam que conceitos são definidos por outros conceitos, como parte de um esquema de conceitos, incluindo opcionalmente declarações sobre as relações semânticas entre os mesmos. Bechhofer e Miles (2009, tradução nossa) finalizam que esta definição é destinada a ser sugestiva e não restritiva, possuindo certa flexibilidade no modelo de dados.

SPARQL *Protocol and RDF Query Language* (SPARQL) é a linguagem de manipulação e consulta para a Web Semântica. Uma linguagem que permite efetuar consultas em fontes de dados estruturadas em RDF, possuindo recursos para consultas em padrões de triplos, conjunções, disjunções, padrões obrigatórios e opcionais (PRUD'HOMMEAUX; SEABORNE, 2008, tradução nossa).

Segundo Harris e Seaborne (2012, tradução nossa) uma consulta em SPARQL geralmente é baseada em um padrão triplo, onde tanto o recurso quanto propriedade e valor podem ser uma variável na consulta. A figura 12 demonstra um exemplo de consulta em SPARQL baseada no quadro 1 e figura 2, onde se tem na linha 1 o valor a ser retornado, linha 3 o recurso a ser consultado e linha 4 a restrição a ser aplicada.

Figura 12 - Exemplo de consulta em SPARQL.

```

1 SELECT ?name
2 WHERE
3 { <http://www.example.org/staffid/85740> <http://www.example.org/terms/name> ?name
4   FILTER regex(?name, "^John")
5 }

```

Fonte: Adaptado de Harris e Seaborne (2012).

Prud'hommeaux e Seaborne (2008, tradução nossa) ressaltam que recursos em consultas em SPARQL podem ser abreviados para facilitar o desenvolvimento. A figura 13 demonstra a abreviação da consulta exemplificada na figura 12.

Figura 13 - Consulta em SPARQL abreviada.

```

1 BASE <http://www.example.org/staffid/>
2 PREFIX terms: <http://www.example.org/terms/>
3
4 SELECT ?name
5 WHERE
6 { ?x <85740> terms:name ?name
7   FILTER regex(?name, "^John")
8 }

```

Fonte: Adaptado de Prud'Hommeaux e Seaborne (2008).

Harris e Seaborne (2012, tradução nossa) complementam que em uma consulta

SPARQL podem haver vários tipos de restrições aplicados a valores de números, caracteres, booleanos e data/hora, assim como uma abrangente biblioteca de funções para se trabalhar com os mesmos, resumindo a sintaxe em:

- a) funções agregadas (*count, avg, sum, min, max*);
- b) projeções (*as*);
- c) restrições (*filter, distinct, having, not exists, exists, minus, in, not in*);
- d) expressões regulares (*regex*);
- e) uniões (*union*);
- f) agrupamentos (*group by*);
- g) ordenações (*order by*);
- h) limite de resultados (*limit, offset*).

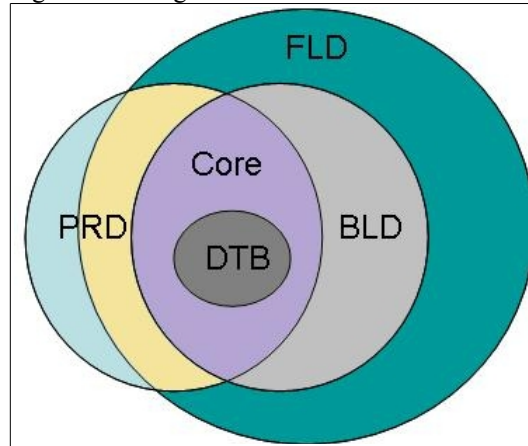
Rule Interchange Format (RIF) é um formato de intercâmbio de regras para a Web Semântica definido por um documento XSL, que permite que sistemas de regras diferentes possam se comunicar de forma transparente (BOLEY; KIFER, 2005, tradução nossa). Boley e Kifer (2005, tradução nossa) complementam ainda que um único sistema de regras é inviável, pois existem muitos paradigmas populares no uso de regras de representação de conhecimento e modelagem de negócio, contidas em regras de primeira ordem, lógica de programação e regras de ação. RIF foi projetado pelo W3C com uma sintaxe e semântica rigorosamente especificada, sendo capaz de representar os paradigmas atuais por meio de dialetos uniformes e extensíveis.

RIF *Framework for Logic Dialects* (FLD), não é de fato um dialeto e sim uma extensibilidade lógica com o objetivo de diminuir a quantidade de esforço necessária para definir e verificar lógicas que expandem as capacidades de RIF *Basic Logic Dialect* (BLD). RIF BLD é uma especialização da RIF FLD capaz de representar as regras de *Horn*, reforçada com uma série de extensões sintáticas para suportar recursos expressivos. RIF *Production Rules Dialect* (PRD) especifica um dialeto de produção de regras para permitir a interoperabilidade de regras, definido no núcleo como um subconjunto comum de RIF BLD e RIF PRD.

RIF *Core* é um subconjunto comum de RIF BLD e RIF PRD, incluindo também RIF *Datatypes and Built-ins* (DTB) que tem como propósito definir tipos de dados e funções construídas por meio de predicados semânticos (GINSBERG et al, 2005, tradução nossa).

A figura 14 demonstra um diagrama contendo toda a família de dialetos de RIF de acordo com sua expressividade.

Figura 14 - Diagrama da família de dialetos RIF.



Fonte: Ginsberg et al (2005).

Boley et al (2010, tradução nossa) exemplificam uma regra definida por meio de uma RIF *Core*, representando uma transação de compra e venda:

- a) um comprador adquire um item de um vendedor se o vendedor vende o item para o comprador;
- b) John vende LeRif a Maria.

A figura 15 demonstra o exemplo da estrutura RIF *Core* a partir do exemplo anterior.

Figura 15 – Exemplo de RIF Core.

```

1 Document(
2   Prefix(cpt <http://example.com/concepts#>)
3   Prefix(ppl <http://example.com/people#>)
4   Prefix(bks <http://example.com/books#>)
5
6   Group
7   (
8     Forall ?Buyer ?Item ?Seller (
9       cpt:buy(?Buyer ?Item ?Seller) :- cpt:sell(?Seller ?Item ?Buyer)
10    )
11
12    cpt:sell(ppl:John bks:LeRif ppl:Mary)
13  )
14 )

```

Fonte: Boley et al (2010).

2.2.5 Camada de lógica, prova, confiança e criptografia

Desde os primeiros esboços da arquitetura da Web Semântica os pesquisadores do W3C sugerem e também realizam recomendações oficiais de tecnologias para cada camada da arquitetura, entretanto as camadas de lógica, prova, confiança e por fim a de criptografia nunca foram esclarecidas quanto as suas tecnologias, mas sim quanto aos seus objetivos. Ramalho et al (2007) comentam que a camada de lógica possibilitará às máquinas realizarem inferências automáticas a partir das relações entre recursos com base em regras mais abrangentes do que as definidas na camada semântica e assim como também inferir novas informações. A camada de prova será responsável por verificar e garantir uma coerência lógica dos recursos, onde os aspectos semânticos das informações devem estar corretamente descritos, de forma que atenda os requisitos das camadas inferiores. Enquanto a camada de confiança garantirá que as informações estejam representadas de modo correto, atingindo um certo grau de confiabilidade. Já a camada de criptografia será responsável por cifrar todas as informações trocadas entre as camadas da Web Semântica, impossibilitando aos humanos e máquinas interpretarem essas informações caso não possuam uma determinada chave de acesso, garantindo assim a confiabilidade.

2.2.6 Camada de interação

O topo da arquitetura da Web Semântica é a utilização da mesma por parte dos usuários finais, onde os mesmos realmente irão interagir com a Web Semântica. As camadas anteriores fornecem por meio de tecnologias recomendadas oficialmente pelo W3C uma arquitetura base para que aplicações possam ser desenvolvidas de modo que humanos e máquinas possam interagir na consulta e manipulação de informações semanticamente (BERNERS-LEE et al, 2001, tradução nossa). Breitman (2005) comenta que existem inúmeras possibilidades de interação com a Web Semântica, mas cita os principais, relacionados ao ambiente Web, como páginas da Web, motores de busca, serviços Web e aplicações em geral.

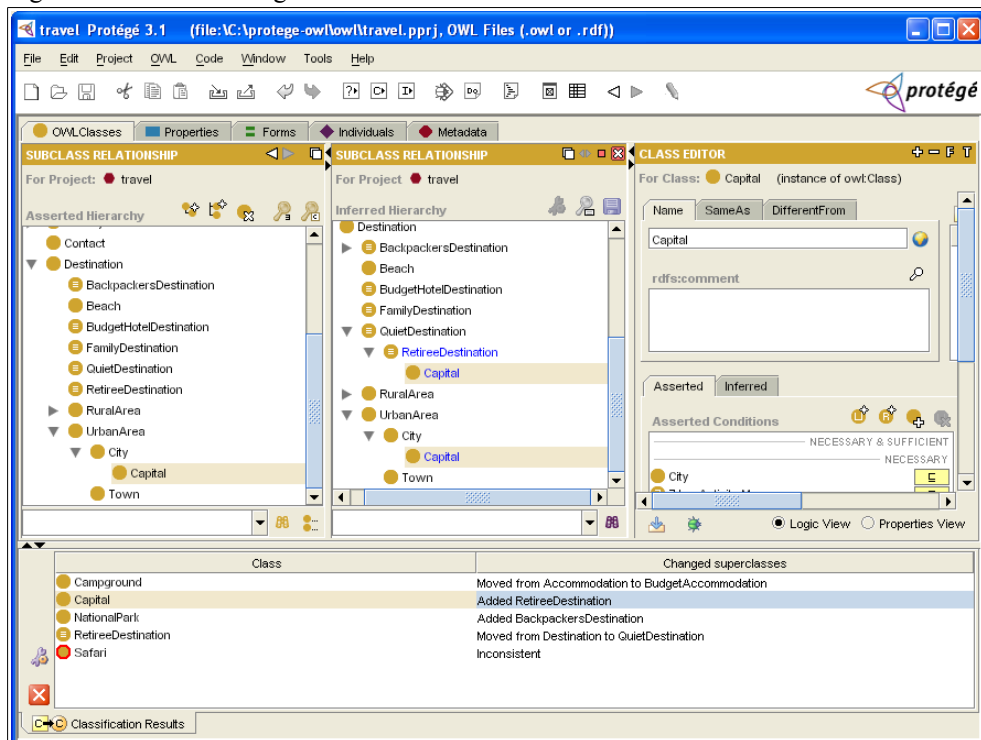
2.3 FERRAMENTAS PARA A WEB SEMÂNTICA

Desde a concepção da Web Semântica várias ferramentas foram e estão sendo desenvolvidas com o intuito de abstrair as tecnologias que compõem a sua arquitetura, assim como no desenvolvimento da mesma. Breitman (2005) classifica as ferramentas da Web Semântica em três categorias:

- a) editores de ontologias;
- b) metadados;
- c) ferramentas de visualização em mecanismos de inferências.

Ainda segundo Breitman (2005) quando se fala de construção e manutenção de ontologias a principal ferramenta visual é o Protégé, um programa livre, multiplataforma, desenvolvido em Java e que permite a construção, edição e visualização de ontologias, assim como também possibilita que desenvolvedores criem *plugins* que aumentam a expressividade da ferramenta. Um exemplo de *plugin* é o SKOSEd, que possibilita criar e editar tesouros e/ou artefatos similares representados por meio de SKOS. A figura 16 apresenta uma tela do Protégé.

Figura 16 - Tela do Protégé.



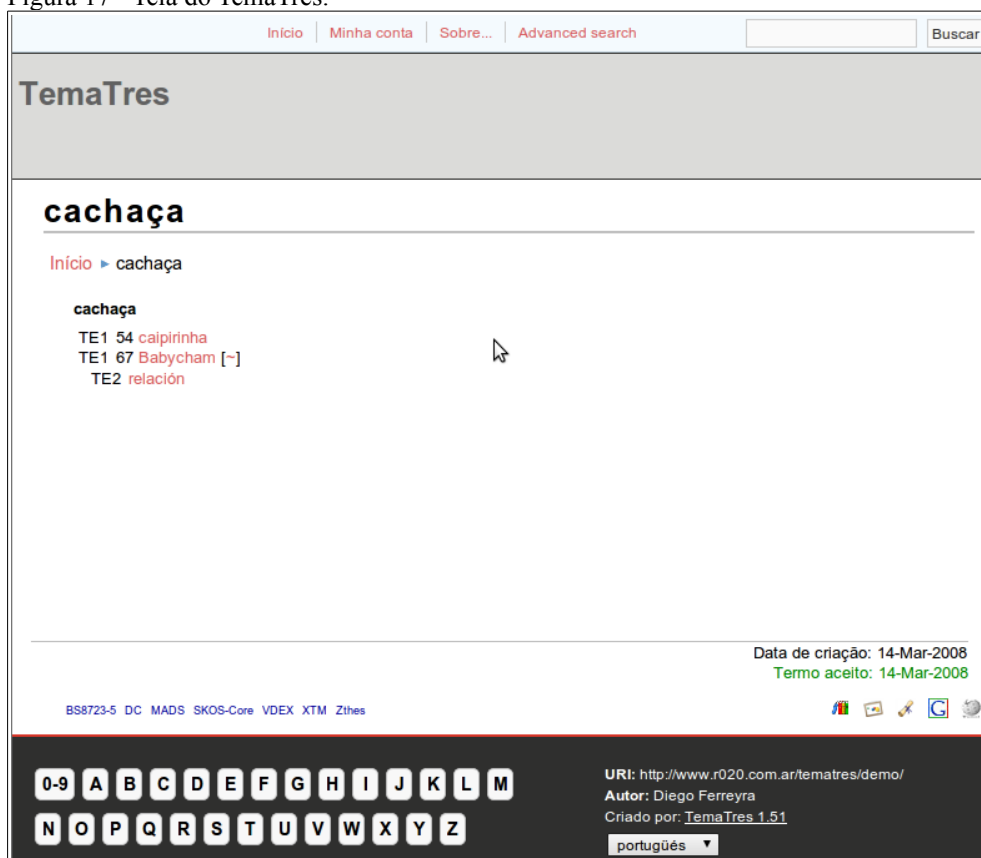
Fonte: *Stanford Center For Biomedical Informatics Research* (2012).

Protégé possui dois editores principais: Protégé-*Frames* e Protégé-OWL, este último designado para a construção e manutenção de ontologias em OWL. Dentre as principais características do Protégé-OWL, destacam-se:

- a) criar, carregar e salvar arquivos em OWL e RDF;
- b) editar e visualizar as classes, propriedades e regras;
- c) definir as características de classes lógicas como expressões OWL;
- d) executar raciocinadores como classificadores de lógica de descrição;
- e) editar indivíduos OWL para marcação da Web Semântica.

Além de editores de ontologias em OWL também existem editores e gerenciadores de estruturas baseadas em SKOS, sendo uma das principais ferramentas visuais o TemaTres. TemaTres é uma aplicação Web desenvolvida em PHP *Hypertext Preprocessor* (PHP) de código aberto que possui utilitários para gerenciamento distribuído de vocabulários, recursos avançados de pesquisa, edição e publicação de serviços baseados em vocabulários controlados (FERREYRA, 2007, tradução nossa). A figura 17 mostra uma tela do TemaTres.

Figura 17 - Tela do TemaTres.



Fonte: Ferreyra (2007).

TemaTres também é um servidor de acesso a tesouros, possuindo algumas funcionalidades específicas:

- a) sugestão de pesquisa (você quis dizer ...?);
- b) exposição de vocabulários como serviços Web;
- c) suporte para tesouro multilíngue;
- d) navegação sistemática ou alfabética;
- e) importação e exportação de dados em SKOS.

A partir da definição do W3C Brasil (2011) que a Web Semântica é um banco de dados global se faz necessário que um banco de dados a ser utilizado semanticamente tenha algumas particularidades, principalmente aceitando consultas no padrão SPARQL. Virtuoso *Universal Server* é um servidor universal de dados multitarefa e multiplataforma, com características de bancos de dados relacional e virtual, disponível em duas edições: comercial e código aberto, esta última denominado OpenLink Virtuoso (*Semantic Web*, 2012, tradução nossa). Segundo a OpenLink *Software* (2012, tradução nossa) as principais características do Virtuoso são:

- a) gerenciamento de dados relacionais;
- b) gerenciamento de dados em estruturas XML, RDF e OWL;
- c) servidor Web de documentos;
- d) servidor de dados ligados;
- e) servidor de aplicações Web;
- f) servidor de serviços Web.

Uma opção poderosa do Virtuoso é o Virtuoso SPARQL *Query Editor*, um editor de consulta visual e via Web que possibilita a execução de pesquisas no formato SPARQL em uma fonte de dados RDF previamente armazenada em uma instância do Virtuoso, possibilitando ainda que o resultado da consulta possa ser convertido em diversos formatos, como por exemplo: JSON, XML, HTML. (SEMANTIC WEB, 2012, tradução nossa). A figura 18 demonstra o Virtuoso SPARQL *Query Editor*.

Figura 18 - Editor de consulta SPARQL.

Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#) | [iSPARQL](#)

Default Data Set Name (Graph IRI)

Query Text

```
select distinct ?Concept where {[] a ?Concept} LIMIT 100
```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format: (The CXML output is disabled, see [details](#))

Execution timeout: milliseconds (values less than 1000 are ignored)

Options: Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#))

Fonte: DBpedia (2012).

TopBraid *Suite* é um conjunto de ferramentas com edições gratuitas e comerciais que abstraem e utilizam as tecnologias da arquitetura da Web Semântica para construir aplicações semânticas flexíveis com base em modelos de dados ligados (TOPQUADRANT, 2012, tradução nossa). A TopBraid *Suite* fornece três ambientes de desenvolvimento que ajudam a projetar, desenvolver e implantar modelo de dados semânticos atendendo as mudanças que tangem a recuperação da informação adaptativa e sustentável, sendo eles: TopBraid *Composer*, TopBraid *Ensemble* e TopBraid *Live*.

TopBraid *Composer* proporciona a modelagem de ontologias, interoperabilidade de dados, consultas e aplicação de regras semânticas. TopBraid *Ensemble* fornece um ambiente Web para criar aplicações orientadas a modelo, com capacidade de utilização de *Rich Internet Applications* (RIA). TopBraid *Live* permite implementar aplicações baseadas em *Service Oriented Architecture* (SOA) e integrar de dados *on-demand* por meio de estruturas baseadas em RDF e/ou OWL.

Ainda segundo a TopQuadrant (2012, tradução) a TopBraid *Suite* suporta em todo

o seu conjunto de ferramentas a utilização de SPARQL, oferecendo inferência flexível e verificação de restrição, assim como ferramentas auxiliares baseadas em SPARQL que complementam a suíte, denominadas: SPARQL *Motion*, SPARQL *Inferencing Notation* (SPIN) e SPARQL *Web Pages* (SWP).

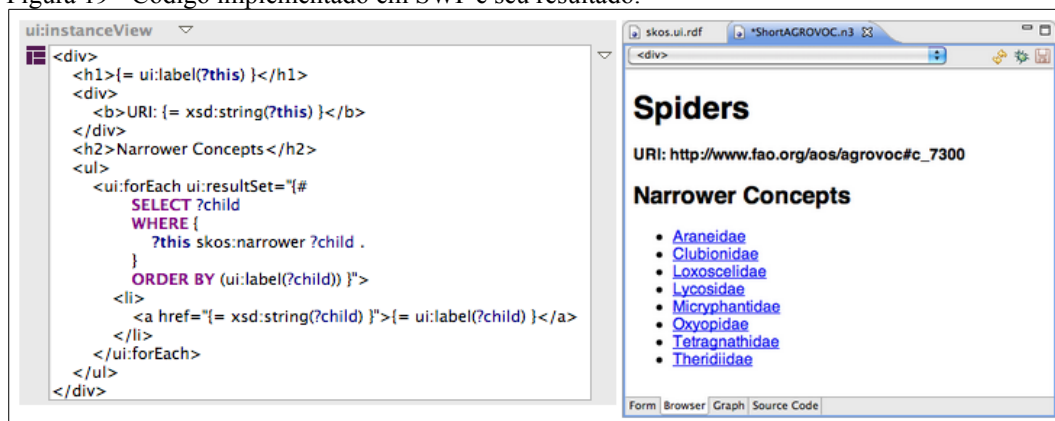
SPARQL *Motion* é uma linguagem de *script* visual com motor de processamento semântico, envolvendo consultas e transformações de dados, onde cada *script* pode ser definido como um serviço Web que retorna dados em XML, RDF, JavaScript *Object Notation* (JSON) e/ou outras interfaces de usuário (TOPQUADRANT, 2012, tradução nossa).

SPIN define regras e restrições de negócio que podem ser expressas em SPARQL e executadas sobre modelos de dados RDF. SPIN é uma especificação que foi submetida ao W3C para se tornar uma recomendação oficial (TOPQUADRANT, 2012, tradução nossa).

SWP é um *framework* para construção de interfaces Web com base em dados ligados, onde as interfaces do usuário estão ligadas a definições de classes e em modelos de ontologias (TOPQUADRANT, 2012, tradução nossa).

A figura 19 exemplifica um código implementado em SWP e seu resultado.

Figura 19 - Código implementado em SWP e seu resultado.



Fonte: TopQuadrant (2012).

O grupo Semantic Web (2012) resume outras ferramentas reconhecidas pelo W3C, que ajudam a implementação da Web Semântica, divididas em editores e gerenciadores de OWL e SKOS, bancos de dados RDF, raciocinadores lógicos, ambientes de desenvolvimento, validação, visualização, extração e edição de dados na Web.

OWL *Application Programming Interface* (API) é um *framework* Java considerado pelo W3C como implementação de referência para desenvolvimento de ontologias em OWL DL e OWL 2 e como base para outras ferramentas que manipulam OWL.

Apache Jena é um *framework* Java que fornece um ambiente de desenvolvimento para RDF, RDF *Schema*, OWL e SPARQL, incluindo uma máquina de inferência baseada em regras, possuindo também uma integração com a ferramenta Protégé.

OWLDiff é uma ferramenta baseada em OWL API que realiza a verificação de diferença estrutural entre duas ontologias. OWLDiff pode ser utilizado de forma independente ou como um *plugin* do Protégé. PoolParty é um sistema de gerenciamento de dicionário de sinônimos e um editor de SKOS para a Web Semântica, incluindo funcionalidades de análise de texto e recursos de dados ligados, extração de entidades em textos não estruturados e uma plataforma de pesquisa semântica.

AllegroGraph é um banco de dados RDF que visa proporcionar uma camada de armazenamento sólido para o raciocínio geo temporal, análise de redes sociais e capacidades de modelagem de ontologias para aplicações semânticas. Sesame é um *framework* Java extensível para o armazenamento, consulta e inferência em RDF. Podendo ser implantado como um servidor Web ou como uma biblioteca Java. Sesame possui armazenamento independente, assim como também é suportado por outros fornecedores de banco de dados RDF. Bigdata é um banco de dados RDF em escala horizontal com suporte a transações e concorrência, aceitando altas taxas agregadas de I/O, sendo projetado em uma arquitetura de banco de dados distribuído, mas também pode ser executado em um único servidor.

RIFLE é uma suíte de ferramentas para RIF. Mais precisamente RIF *Core* e RIF PRD, assim como combinação com outros padrões da Web Semântica, como RDF, OWL e SPARQL. HermiT é raciocinador de ontologias OWL para lógica descritiva. Dado um arquivo OWL, pode-se determinar se a ontologia é consistente, identificar relações de subjunção entre as classes, assim como outras relações. Quest é um raciocinador focado em consultas eficientes de SPARQL, suportando estruturas em RDF *Schema* e OWL 2, com regimes de inferência QL e no contexto de ontologias com mapeamentos para fontes de dados relacionais. Pellet é um raciocinador de lógica descritiva para OWL DL de uso livre, suportando toda a expressividade de OWL DL, incluindo raciocínio sobre nominais, como classes enumeradas.

CubicWeb é um *framework* que capacita desenvolvedores a construir aplicações para a Web Semântica de forma eficiente com a reutilização de componentes (chamados cubos) e seguindo princípios de orientação a objetos de design. SPARQL *Endpoint* interface to Python é uma ferramenta de consulta e conversão de resultados baseada em SPARQL para ser utilizado em implementações Python. jOWL é um *plugin* jQuery para navegação e

visualização de documentos OWL e RDF *Schema*.

Vapour é um serviço de validação para verificar se os dados de uma página supostamente semântica estão corretamente publicados de acordo com as melhores práticas da Web Semântica. Purifyr é uma ferramenta de extração de dados da Web com processamento de linguagem natural, escrito em C e Python, detecta ruídos semânticos, informações de navegação, artigos relacionados ou nuvens de *tags*.

RDF *Validation Service* é uma ferramenta oficial do W3C que valida uma estrutura RDF e em caso de sucesso gera tanto o modelo de dados triplo quanto um grafo RDF (PRUD'HOMMEAUX, 2007, tradução nossa). W3C *Markup Validation Service* é um validador oficial do W3C para verificar e validar a marcação de documentos da Web, como HTML e *eXtensible HyperText Markup Language* (XHTML), desde que os mesmos estejam de acordo com o DTD *Resource Description Framework in Attributes* (RDFa) (W3C, 2012, tradução nossa).

Breitman (2005) destaca uma importante ferramenta para a Web Semântica, visando principalmente a manutenção de páginas da Web, trata-se do editor de metadados DC.dot, que oferece um ambiente Web para edição de conteúdo descrito do padrão *Dublin Core* (DC). Por meio da submissão de uma página da Web a ferramenta DC.dot analisa e sugere um conjunto de metadados DC para a página submetida, facilitando assim a realização *a posteriori* de marcações semânticas em páginas da Web.

Bizer et al (2006, tradução nossa) apresentam o *Information Quality Assessment Framework* (WIQA), um conjunto de componentes para filtragem de informações da Web, visando a busca da qualidade das mesmas, estas contidas no conteúdo, contexto e relevância de uma informação. O WIQA foi desenvolvido para atender requisitos como representação flexível de informações juntamente com metadados relacionados de qualidade, suporte para informações de diferentes políticas de filtragem e explicação de decisões da filtragem.

RDFa é o caminho mais curto para o desenvolvimento de páginas voltadas para a Web Semântica, pois permite aos desenvolvedores Web que por meio de uma especificação estendam os atributos em *tags* da linguagem HTML ou XML com novos atributos, marcando de forma legível por máquinas o conteúdo que está sendo exibido na página. Este tipo de marcação possibilita uma estruturação dos conteúdos e conseqüentemente da página, objetivando na possibilidade de extrair triplos RDF, assim como auxiliar os motores de busca na indexação da mesma. A marcação pode ser simples como a indicação do título da página

ou complexa como a rede de relacionamentos do criador da página (ADIDA et al, 2012, tradução nossa). A figura 20 exemplifica um código em HTML utilizando RDFa.

Figura 20 - Utilização de RDFa em código HTML.

```
1 <html>
2 <head>
3   ...
4 </head>
5 <body>
6   <div>
7     <h2 property="dc:title">The trouble with Bob</h2>
8     ...
9     <h3 property="dc:creator" resource="#me">Alice</h3>
10    ...
11  </div>
12 </body>
13 </html>
```

Fonte: Adaptado de Adida et al (2012).

RDFa é uma recomendação oficial do W3C, assim como uma extensão do HTML5, que incorpora em sua especificação uma evolução de microformatos e microdados como forma de marcação de dados em páginas da Web, podendo ser utilizada de duas formas: RDFa *Core* e RDFa *Lite*. RDFa *Core* compreende todos os atributos básicos e avançados da especificação, possibilitando desenvolvedores expressarem dados complexos e estruturados em páginas da Web, como relações entre pessoas, lugares e eventos. RDFa *Lite* utiliza um subconjunto básico de atributos de RDFa: *vocab*, *typeof*, *property*, *resource*, e *prefix*, indicado para páginas da Web simples e com pouco conteúdo (SPORNY, 2012, tradução nossa).

O projeto schema.org é uma colaboração entre a Google, a Microsoft e o Yahoo! com o objetivo de unificar, simplificar e evoluir as formas de marcação de dados estruturados em páginas da Web denominados por microformatos, microdados e RDFa, gerando um único vocabulário e uma única sintaxe. Esta padronização possibilita que desenvolvedores não necessitem implementar páginas da Web com base no tipo de marcação suportado por cada mecanismo de pesquisa, visto que o schema.org é suportado pelos principais mecanismos de buscas, de forma padrão e concreta. O schema.org é caracterizado pelo uso de microdados, representando um equilíbrio entre a capacidade de extensão de RDFa e a simplicidade do uso de microformatos (GOOGLE, 2012). Sua sintaxe é bastante simples, possuindo três atributos:

- a) *itemscope*: marca o início de um novo item;
- b) *itemtype*: define o tipo do item;
- c) *itemprop*: define a propriedade de um item.

A figura 21 exemplifica a utilização do schema.org em uma página da Web.

Figura 21 - Utilização de schema.org em uma página da Web.

```
<div itemscope itemtype="http://schema.org/Movie">  
  <h1 itemprop="name">Avatar</h1>  
  <span>Director:  
  <span itemprop="director">James Cameron</span>(born August 16, 1954)</span>  
  <span itemprop="genre">Science fiction</span>  
  <a href=" ../movies/avatar-theatrical-trailer.html" itemprop="trailer">Trailer</a>  
</div>
```

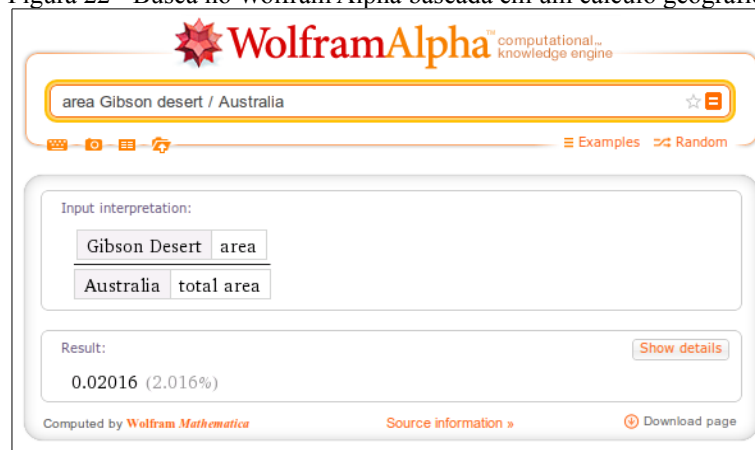
Fonte: Adaptado de Schema.org (2012).

Gleaning Resource Descriptions from Dialects of Languages (GRDDL) é um mecanismo recomendado oficialmente pelo W3C que permite extrair informações organizadas de páginas da Web, independente da estrutura de dados utilizada, resultando em uma estrutura RDF. A estrutura padrão de dados em páginas da Web para a Web Semântica é o RDF, entretanto a maioria das páginas ainda não são estruturadas desta forma, o que leva ao uso do GRDDL para esta transformação (CONNOLLY, 2007, tradução nossa). O *framework* Jena possui uma implementação baseada no GRDDL, denominado *Jena GRDDL Reader*, que possibilita uma maneira mais simples da utilização dos mecanismos de extração.

2.4 EXEMPLOS DE APLICAÇÕES NA WEB SEMÂNTICA

As aplicações na Web Semântica dividem-se principalmente em buscadores semânticos, redes sociais com recursos semânticos, *wikis* semânticas e dicionários semânticos. Wolfram Alpha é um buscador semântico que introduz uma maneira radicalmente nova de obter-se conhecimento e respostas, não por meio de pesquisas na Web, mas realizando cálculos dinâmicos baseados em uma vasta coleção de *built-in* de dados, algoritmos e métodos semânticos, com capacidade de realização de cálculos e respostas a perguntas. Na versão comercial possibilita inserção de arquivos para realização de buscas baseada nos mesmos (WOLFRAM ALPHA, 2012, tradução nossa). A figura 22 demonstra uma busca baseada em um cálculo geográfico.

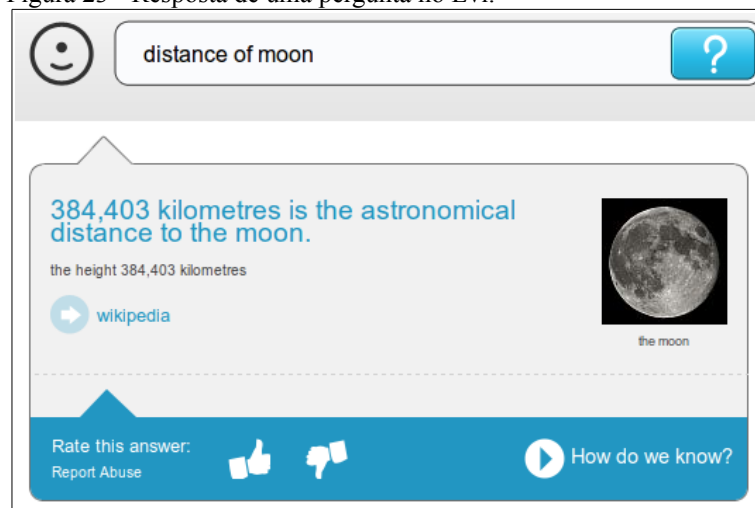
Figura 22 - Busca no Wolfram Alpha baseada em um cálculo geográfico.



Fonte: Wolfram Alpha (2012).

Entre os buscadores semânticos que só respondem perguntas se destacam o Ask e o Evi que alimentam um novo tipo de experiência de pesquisa onde os usuários podem acessar o conhecimento do mundo simplesmente solicitando a informação que precisam de uma maneira natural. O buscador Evi possui o recurso de acesso por meio da Web ou de aplicativos móveis, sendo ainda permitido que o usuário aprove ou não a resposta (EVI, 2012, tradução nossa). A figura 23 exibe a resposta de uma pergunta realizada no buscador Evi.

Figura 23 - Resposta de uma pergunta no Evi.



Fonte: Evi (2012).

O buscador hokia funciona como qualquer buscador sintático, entretanto realiza uma ligação semântica de conceitos nos resultados, sendo que além de ser um site de buscas também oferece uma plataforma de buscas semânticas para as empresas instalarem em suas

aplicações e sites internos e externos (HAKIA, 2012, tradução nossa). Já o buscador semântico Swoogle tem um objetivo um pouco diferente dos anteriores, pois permite que usuários e/ou aplicações realizem buscas de documentos da Web Semântica, especialmente ontologias (SWOOGLE, 2012, tradução nossa).

Os principais buscadores do mercado não estão ignorando a Web Semântica. A Microsoft adquiriu o motor de busca baseado em linguagem natural Powerset e o incorporou no seu buscador Bing (MICROSOFT, 2012, tradução nossa). A Google fez a mesma coisa, disponibilizando na versão em inglês do seu buscador o *Knowledge Graph* (Grafo do Conhecimento). Segundo a Google (2012, tradução nossa), quando o usuário realiza uma busca ele não está apenas procurando uma página da Web e sim respostas, quer entender e explorar conceitos e este é o principal objetivo do *Knowledge Graph*, a compreensão das coisas do mundo real e as relações entre elas.

A figura 24 mostra uma busca efetuada no Google em inglês pelo nome de um pintor, trazendo os resultados padrões a esquerda e a direita os conceitos e relacionamentos sobre o mesmo.

Figura 24 - Busca no Google utilizando Knowledge Graph.

The screenshot shows a Google search for "vincent van gogh". The search bar contains the text "vincent van gogh" and shows "About 19,100,000 results (0.17 seconds)". The search results on the left include:

- Vincent van Gogh - Wikipedia, the free encyclopedia**: en.wikipedia.org/wiki/Vincent_van_Gogh
Vincent Willem **van Gogh** was a Dutch post-Impressionist painter whose work, notable for its rough beauty, emotional honesty, and bold color, had a far-reaching ...
 → List of works - Health - The Starry Night - Sunflowers
- Vincent van Gogh Gallery - Welcome!**: www.vangoghgallery.com/
 The definitive reference for information about **Vincent van Gogh** including his biography and the complete collection of his paintings, drawings, sketches and ...
- Vincent van Gogh Biography - His Life and Times**: www.vangoghgallery.com/misc/bio.html
 Read a biography of Dutch post-Impressionist artist **Vincent van Gogh**. Get quick facts, a timeline, information about his family and artists who influenced him.
- Images for vincent van gogh** - Report images

The Knowledge Graph panel on the right displays:

- Vincent van Gogh**: Vincent Willem van Gogh was a Dutch post-Impressionist painter whose work, notable for its rough beauty, emotional honesty, and bold color, had a far-reaching influence on 20th-century art. Wikipedia
- Born:** March 30, 1853, Zundert
- Died:** July 29, 1890, Auvers-sur-Oise
- Parents:** Anna Carpentus van Gogh
- Siblings:** Theo van Gogh, Wil van Gogh
- Periods:** Post-Impressionism, Impressionism, Divisionism
- Works:**
 - The Starry Night** (1889)
 - Cafe Terrace at Night** (1888)
 - Irises** (1889)
 - The Potato Eaters** (1885)
 - Starry Night Over the...** (1888)

Fonte: Google (2012).

A empresa adquirida pela Google foi a Metaweb, que também possui um repositório de dados composto por informações semânticas, chamado Freebase. Freebase também possui uma API para acesso aos dados por meio de aplicações, disponibilizando inclusive estrutura de dados em RDF (FREEBASE, 2012, tradução nossa). A compra da

Metaweb alavancou o desenvolvimento da Web Semântica na Google, surgindo assim o Google *Refine*, uma ferramenta para trabalhar com dados desorganizados, limpando-os e transformando-os em um formato único, podendo ser estendido como um serviço Web ou vinculando a uma bases de dados, como o Freebase (GOOGLE CODE, 2012, tradução nossa).

Nas redes sociais os destaques semânticos são os sites SciVee e Uptake. SciVee é um site de compartilhamento de arquivos científicos, onde os usuários podem enviar arquivos ao site e relacioná-los a outros arquivos, pessoas e entidades (SCIVEE, 2012, tradução nossa).

Uptake é uma rede social de planejamento de viagens, onde o usuário pode descobrir as pessoas que já foram para um lugar específico e saber a opinião delas sobre o mesmo, assim como o que elas indicam fazer em cada lugar. Uptake também aceita acesso por meio de uma conta da rede social Facebook, aumentando assim um pouco mais a qualidade da informação, já que os amigos do usuário podem fazer parte da base de conhecimento para as respostas do planejamento da viagem (UPTAKE, 2012, tradução nossa).

Entre as principais redes sociais o Facebook saiu na frente e criou o *Open Graph*, uma API que possibilita outros sites e /ou aplicações se conectarem a rede social por meio de uma estrutura padrão: RDFa e do protocolo: *Open Graph Protocol*, de modo que uma mesma informação pode estar vinculada em várias aplicações e/ou sites juntamente com o Facebook (FACEBOOK, 2012, tradução nossa).

No ambiente *wiki* várias aplicações existem para os mais variados fins, desde a construção de *wikis* semânticas até a utilização de conceitos de *wikis* criadas. *Semantic MediaWiki Plus* (SMW+) é um ambiente *wiki* semântico e colaborativo adequado para equipes e empresas que trabalham com fluxos de trabalho complexos e informais. SMW+ oferece um conjunto de características para criar, compartilhar, publicar e reutilizar o conhecimento contido no conteúdo do *wiki* (SEMANTIC WEB, 2011, tradução nossa).

DBpedia é um esforço da comunidade livre que permite extrair as informações da Wikipedia e disponibilizar as mesmas em formato semântico, mais precisamente em estruturas RDF. Possibilitando assim a realização de consultas sofisticadas e vinculação dos dados a outros conjuntos de dados da Web. DBpedia possui uma API para acesso a base de dados rodando em instâncias do Virtuoso, que por sua vez armazena as estruturas RDF e disponibiliza um meio para realização de consultas no padrão SPARQL (DBPEDIA, 2012, tradução nossa).

Faviki é outra aplicação desenvolvida para trabalhar em conjunto com a

Wikipedia, tem um formato de rede social de *tags*, onde é possível adicionar *tags* a assuntos, vinculando as mesmas a conceitos dentro da Wikipedia, assim como relacionar a *tags* de outros usuários (FAVIKI, 2012, tradução nossa).

Dicionários semânticos se concentram basicamente em tesouros para representação de conhecimento e a linguagem SKOS para o desenvolvimento. Segundo Breitman (2005) o tesouro mais utilizado é o WordNet, um banco de dados léxico de língua inglesa desenvolvido por especialistas em cognição. No WordNet cada palavra é associada a uma ou várias definições, representando a memória léxica dos seres humanos. O WordNet pode ser acessado pela Web para realização de consultas de palavras ou pode ser baixado gratuitamente (WORDNET, 2012, tradução nossa).

WordWeb é muito semelhante ao WordNet, diferenciando-se por não ter um único banco de dados, sendo que se uma determinada palavra não for encontrada a aplicação realiza a busca em bancos de dados de outros servidores. WordWeb diferencia-se também por possuir versões Web, *desktop* e para aplicativos móveis (WORDWEB, 2012, tradução nossa).

Visuwords é um dicionário gráfico *on-line* que apresenta os resultados reminiscentes de rede neural representando palavras e conceitos associando-os com outras palavras e/ou conceitos, assim como disponibilizando legendas para explicar as mesmas. Visuwords usa o banco de dados livre WordNet para recuperar as informações (VISUWORDS, 2012, tradução nossa).

O Jornal The New York Times também aderiu a Web Semântica e desde 2009 disponibiliza um dicionário de dados em arquivos no padrão SKOS para download público. O dicionário foi desenvolvido como um vocabulário baseado no *Linked Open Data*, possuindo dados de pessoas, organizações, localidades e descrições diversas (THE NEW YORK TIMES, 2009, tradução nossa).

A Globo, por meio do portal globo.com, foi outra corporação de grande porte que aderiu a Web Semântica, mais precisamente na gestão de conteúdos editoriais, utilizando-se de tecnologias padronizadas desta. O projeto começou em 2009 para resolver o problema dos relacionamentos de um mesmo tema, pessoa ou entidade nos mais diversos editoriais do portal. A solução foi o emprego de uma representação semântica para servir como marcação do conteúdo editorial (CAROLO, 2011). A figura 25 exemplifica a publicação de uma notícia utilizando semântica.

Figura 25 - Ontologia na publicação de uma notícia.

Na lista negra de Mou, juiz de Barça x Real não se incomoda com críticas

Belga é tido pelo técnico do Real como um dos árbitros que favorecem clube catalão. De Bleckere se diz satisfeito por ter sido escolhido

Url (seo): na-lista-negra-de-mou-juiz-de-barca-x-real-nao-se-incomoda-com-criticas

<http://globoesporte.globo.com/futebol/liga-dos-campeoes/noticia/2011/05/na-lista-negra-de-mou-juiz-de-barca-x-real-nao-se-incomoda-com-criticas.html>



Criticado por Mourinho, o árbitro belga, Frank de Bleckere, que vai apitar Barcelona e Real Madrid, pela semifinal da [Liga dos Campeões](#), preferiu não comentar as declarações do técnico merengue, que reclamou da escolha da Uefa.

Junto com outros árbitros, [De Bleckere está](#)

Anotação semântica

Esportes (Esporte, Categoria e Modalidade)

Pessoas (Atleta, Árbitro e Técnico)

Organizações (Clube, Agremiação, Time e Seleção)

Eventos (Edição de Campeonato, Jogo e Prova)

Cita o Canal ou Programa de TV

Trata do Assunto

Fonte: Carolo (2011).

Carolo (2011) destaca que o próximo passo é adicionar nas páginas do portal atributos pertencentes a especificação RDFa.

3 REPRESENTAÇÃO DE CONHECIMENTO NA WEB

A representação de conhecimento na Web está embutida dentro do conceito de Sistemas de Organização do Conhecimento (SOC) tradução do original inglês *Knowledge Organization System* (KOS), onde o conhecimento está associado a algum tipo de sistema, responsável pela organização e recuperação da informação de forma ágil e precisa (MEDEIROS, 2011).

Segundo Carlan (2010), KOS estruturam semanticamente os conceitos, contemplando termos, definições, propriedades e relacionamentos, facilitando assim a indexação da informação por parte dos usuários. Medeiros (2011) complementa que KOS possuem vertentes tanto no campo da Ciência da Informação quanto na Ciência da Computação.

Na Ciência da Informação KOS são geralmente empregados em vocabulários controlados e/ou linguagens documentárias para a área de biblioteconomia, que por meio de instrumentos informacionais são realizadas as traduções dos conteúdos originais e completos para um esquema estruturado e sistemático. Carlan (2010) destaca que os principais instrumentos para representação de conhecimento no campo da Ciência da Informação são tesouros, taxonomias, ontologias e sistemas de classificação.

Na Ciência da Computação se teve primeiramente uma estruturação de informações, com o uso de banco de dados relacionais e orientados a objeto. Banco de dados relacionais ou *Relational Database Management System* (RDBMS) possuem em sua estrutura um modelo relacional, composto por tabelas e seus relacionamentos. Bancos de dados orientados a objetos ou *Object-Oriented Database Management Systems* (OODBMS) também possuem tabelas, entretanto seus relacionamentos são modelados por meio de herança e/ou uma hierarquia de tipos e sub-tipos (ABITEBOUL et al, 2000). Segundo Carvalho e Oliveira (2008), houve então uma evolução na estruturação de informações, chegando-se na capacidade de representação de conhecimento em banco de dados por meio da aplicação de técnicas de Inteligência Artificial (IA), gerando assim as bases de conhecimento, constituídas de sistemas computacionais inteligentes e/ou especialistas, onde os registros e seus relacionamentos estáticos são substituídos por regras, fatos e heurísticas correspondentes ao conhecimento de um ou mais especialistas.

O advento da Web Semântica trouxe um pouco da Ciência da Informação para a

Ciência da Computação, utilizando-se de várias técnicas de sucesso na recuperação de informações, objetivando uma interoperabilidade entre sistemas, padronização e marcação semântica em páginas da Web (MEDEIROS, 2011). Berners-Lee (2009, tradução nossa) destaca que a representação de conhecimento em páginas da Web dentro do contexto da Web Semântica envolve características básicas para seu funcionamento, divididos em:

- a) estrutura: URI (descritores de recursos);
- b) sintaxe: XML e/ou HTML (linguagens de marcação);
- c) semântica: OWL, SKOS, metadados (representação semântica);
- d) interoperabilidade: RDF (ligação dos dados dos dados representados).

Estas características delegam a capacidade ou não de uma forma de representação ser utilizada em páginas da Web, envolvendo a descrição do que vai ser representado, ser passível de marcação semântica em linguagens como XML e/ou HTML e possuir um modelo conceitual semântico, resultando em uma estrutura RDF, assim como a ligação de dados na Web.

3.1 FORMAS DE REPRESENTAÇÃO DE CONHECIMENTO NA WEB

Existem várias formas de representação de conhecimento que podem ser utilizadas em páginas da Web, dentre as mais utilizadas destacam-se: metadados, mapas conceituais, ontologias e tesouros. Estas formas de representação de conhecimento são conceitualizadas e caracterizadas a seguir.

3.1.1 Metadados

Segundo Breitman (2005) e Ferreira (2006) a Web proporciona uma significativa quantidade de informações sobre os mais variados tipos de assuntos, entretanto a tecnologia deixa a desejar no quesito representacional, falta informação sobre a informação, dificultando assim a recuperação de resultados mais precisos em buscas realizadas em páginas da Web. Segundo Ferreira (2006) o uso de metadados é a base para a representação de conhecimento na Web Semântica, sendo conceituado por Breitman (2005) como dados sobre dados, informações sobre um determinado recurso, como: autor, descrição, data de criação. W3C Brasil (2011) complementa que metadados são informações sobre dados em páginas da Web

que podem ser compreendidos por máquinas. Alves (2005) classifica os metadados nos seguintes níveis:

- a) simples: descrições simples e sem padronização;
- b) estruturados: descrições simples, genéricas e padronizadas;
- c) ricos: descrições complexas, completas e padronizadas.

Alves (2005) e Ferreira (2006) finalizam que para a Web Semântica o formato de metadados apropriado são os metadados estruturados, enfatizando o uso do padrão DC. O padrão DC é um vocabulário básico para descrever documentos e/ou recursos em páginas da Web por meio da linguagem XML, utilizando-se de descritores como: título, criador, descrição (SEMANTIC WEB, 2006, tradução nossa). Juntamente com o DC outro padrão que se destaca em páginas da Web é o *Friend of a Friend* (FOAF).

FOAF é um padrão de metadados para descrever pessoas e sua rede social de amigos de forma semântica. Geralmente utilizado em páginas pessoais e em publicações de artigos, textos e documentos diversos, descrevendo informações de uma pessoa, como nome, telefone, site, e-mail (SEMANTIC WEB, 2005, tradução nossa).

O shema.org, utilizado como um vocabulário de metadados, objetiva um vocabulário rico e único para tudo que possa ser representado em uma página da Web, sua representação tem como tipo genérico o *Thing* (pensamento), extensível a outros tipos específicos, como: *Event*, *Place*, *Person*, *Product*, entre outros, assim como a possibilidade de criação de novos tipos a partir dos já existentes.

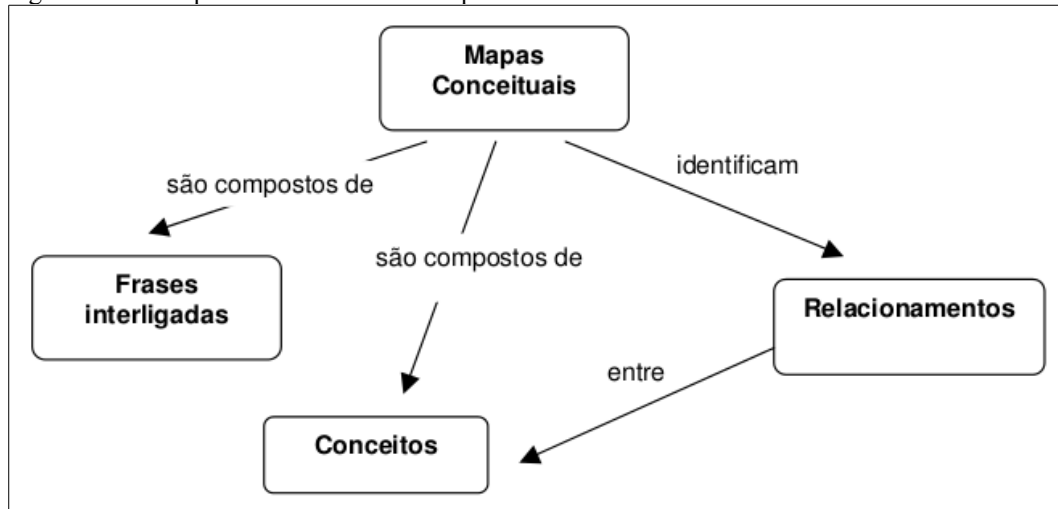
Iannella e Waugh (1997, tradução nossa) relatam alguns problemas da representação de conhecimento por metadados encontrados na maioria dos padrões e que ainda não foram solucionados:

- a) acessibilidade e documentação dos padrões;
- b) interoperabilidade entres os padrões;
- c) ampliação de um padrão existente;
- d) internacionalização de um padrão;
- e) relacionamento entre o dado e o metadado;
- f) metadado do metadado.

3.1.2 Mapas conceituais

Lima (2004) define que um mapa conceitual é a representação de um conceito e sua relação com outros conceitos e/ou ideias de pensamento, constituindo assim uma rede semântica baseada em uma estrutura cognitiva. Machion (2007) complementa que um mapa conceitual possui conceitos principais e secundários, sendo que setas podem dar um sentido de direção nos relacionamentos, assim como definir a hierarquia dos conceitos, ou seja, exibindo os conceitos em uma ordem do mais geral e/ou principal ao mais específico e/ou secundário. A figura 26 demonstra um exemplo de uma mapa conceitual e ao mesmo tempo explica a sua estrutura.

Figura 26 - Exemplo e estrutura de um Mapa Conceitual.



Fonte: Lima (2004).

Segundo Lima (2004) o conhecimento estruturado por meio de um mapa conceitual necessita que a partir de um conceito, a relação deste com os demais seja analisada, interpretada e compreendida, assim construindo novos conhecimentos. A construção de um mapa conceitual envolve as seguintes etapas:

- a) seleção: escolha dos conceitos;
- b) ordenação: do mais geral ao mais específico;
- c) agrupamento: união dos conceitos com mesmo nível de abstração;
- d) arranjo: organização dos conceitos em forma de diagrama;
- e) preposição: frases que indicam a relação.

Pizzato (2003) identifica que o maior problema na representação de conhecimento

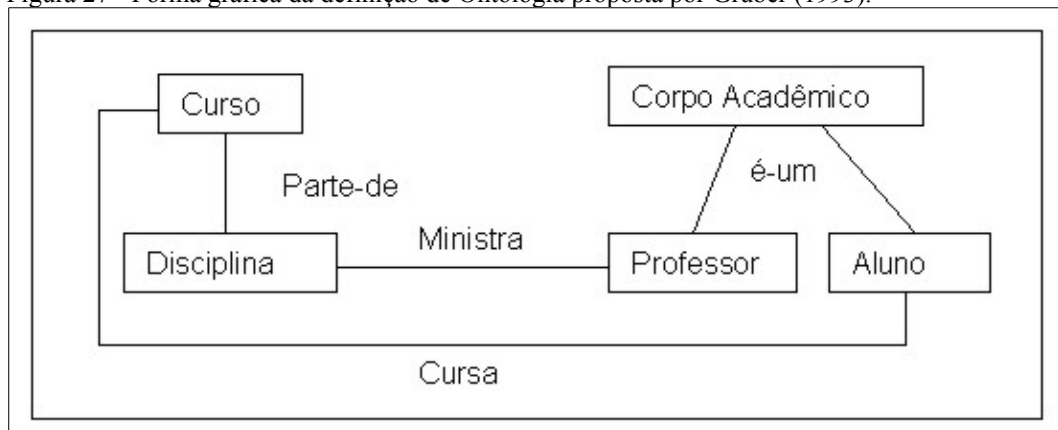
por meio de mapas conceituais é a ambiguidade, mais precisamente em relação ao uso da linguagem natural.

3.1.3 Ontologias

Ontologia é um artefato tecnológico que descreve um modelo conceitual de um determinado domínio em uma linguagem lógica e formal, a partir da descrição dos aspectos semânticos de conteúdos informacionais, possibilitando a realização de inferências automáticas por máquinas (RAMALHO, 2006). Carlan (2006) complementa que uma ontologia é composta por conceitos, propriedades, relações, funções, restrições e axiomas explicitamente definidos que representam um modelo abstrato de um fenômeno ou objeto do mundo real, sendo que o objetivo da construção de uma ontologia é a necessidade de um vocabulário compartilhado, onde as informações podem ser trocadas e reusadas por humanos e/ou máquinas.

A partir da definição de Gruber (1993, tradução nossa) que ontologia é uma especificação formal e explícita de uma conceitualização compartilhada, Moreira (2003) realiza uma exemplificação de forma gráfica da definição na figura 27.

Figura 27 - Forma gráfica da definição de Ontologia proposta por Gruber (1993).



Fonte: Moreira (2003).

Medeiros (2011) complementa que o formalismo de uma ontologia permite que a representação do conhecimento seja clara e concisa assim como suas inferências, permitindo também o processamento por máquinas. A conceitualização é a visão e/ou abstração das entidades e suas relações, o compartilhamento de uma ontologia significa o aceite e

entendimento da mesma perante uma comunidade.

Guarino (1998, tradução nossa) conceitua que uma ontologia define regras que regulam a combinação entre conceitos e seus relacionamentos, de modo que os relacionamentos entre os conceitos são criados por especialistas e os usuários realizam consultas de acordo com os conceitos especificados, resultando assim em uma linguagem de consulta. Segundo Ramalho (2010) uma ontologia deve possuir os seguintes componentes:

- a) classes e subclasses: conceitos;
- b) propriedades descritivas: atributos que o conceito possui;
- c) propriedades relacionais: relação entre conceitos;
- d) regras e axiomas: lógicas de restrição de valores e inferências;
- e) instâncias: conteúdo das classes e subclasses;
- f) valores: conteúdo das propriedades descritivas.

Machion (2007) destaca que uma ontologia dentro da Web Semântica deve possibilitar a realização de inferências entre conceitos, sendo caracterizadas por:

- a) satisfatibilidade: coerência entre conceitos;
- b) inclusão: união de conceitos;
- c) equivalência: igualdade de conceitos;
- d) disjunção: distinção de conceitos.

Breitman (2005) explica que existem classificações para ontologias, caracterizando-se por sua expressividade, sendo que uma ontologia de pouca expressividade pode ser considerada como um vocabulário controlado simples, média expressividade um tesouro e de alta expressividade uma ontologia que possui restrições de valores e/ou de lógica, como disjunção, parte de, inverso, entre outros. Guarino (1998, tradução nossa) apresenta um outro sistema de classificação para ontologias, utilizando a generalidade como critério de classificação, definindo o que cada tipo de ontologia deve descrever:

- a) ontologias de nível superior: conceitos genéricos;
- b) ontologias de domínio: vocabulário de um domínio específico;
- c) ontologias de tarefas: vocabulário de uma tarefa ontológica de nível superior;
- d) ontologias de aplicação: conceitos específicos de uma tarefa.

Segundo Uschold e Gruninger (1996, tradução nossa) uma ontologia ainda pode ser classificada pelo seu grau de formalismo:

- a) altamente informal: expressas em linguagem natural, sem controle;

- b) semi-informal: expressa em linguagem natural, restrita e estruturada;
- c) semi-formal: expressa em linguagem artificial, vocabulário controlado;
- d) rigorosamente formal: definição de termos com semântica formal, teoremas, provas e propriedades.

Um exemplo de ontologia é a *Toronto Virtual Enterprise* (TOVE), uma ontologia voltada ao desenvolvimento de estruturas ontológicas para empresas públicas e privadas, visando demonstrar as características de senso comum de uma empresa, aplicadas a área de IA e sistemas especialistas. A ontologia TOVE divide-se em ontologia de nível superior e de domínio (FOX, 2011, tradução nossa).

O'Leary (1997 apud CARLAN, 2006) cita alguns problemas para a construção e uso de ontologias:

- a) a escolha de uma ontologia é um processo político, já que nenhuma ontologia pode ser totalmente adequada a todos os indivíduos ou grupos;
- b) as ontologias necessitam evoluir, não são estáticas;
- c) estender ontologias não é um processo direto. Ontologias são, geralmente, estruturadas de maneira precisa e, como resultado, são particularmente vulneráveis a questões de extensão, dado o forte relacionamento entre complexidade e precisão das definições;
- d) a noção de bibliotecas de ontologias sugere uma relativa independência entre diferentes ontologias. A interface entre elas constitui, portanto, um impedimento, especialmente porque cada uma delas é desenvolvida no contexto de um processo político.

3.1.4 Tesouros

O termo tesouros se origina do grego *thesaurus* e do latim tesouro, este, que por muito tempo foi utilizado para designar léxico, possuindo duas vertentes: a europeia e a norte americana. A vertente europeia se baseia em um vocabulário organizado com termos que se relacionam de acordo com o seu significado. Já a vertente norte americana se caracteriza pela ordem alfabética dos termos do vocabulário. Um dos primeiros tesouros é o *Thesaurus of English Words and Phrases*, criado por Peter Mark Roget em Londres, no ano de 1852, uma espécie de dicionário que se diferenciava por ser organizado de acordo com o significado dos

termos e não somente com a ordem alfabética (CAMPOS, 2001). Com uma utilização mais frequente, principalmente pela Ciência da Informação, o tesauro ganhou mais características e formas de utilização, se tornando um padrão internacional controlado pela *International Organization for Standardization* (ISO), sob número 2788 de 1986, que culmina na definição de Currás (1995, p. 53):

Tesauro é uma linguagem especializada, normalizada, pós-coordenada, usada com fins documentários, onde os elementos linguísticos que a compõem – termos, simples ou compostos – encontram-se relacionados entre si sintática e semanticamente.

Moreira (2003) esclarece que um tesauro atua em um domínio restrito, possuindo uma linguagem controlada, com termos que se combinam no momento do uso, podendo ser utilizado dentro de um ambiente organizacional, na representação de assuntos e/ou em buscas de informações, sendo classificados no quesito idiomas como monolíngues ou multilíngues e no quesito quantidade de descritores como macrotesauros ou microtesauros, onde nos macrotesauros a quantidade de descritores é menor que nos microtesauros.

Segundo a documentação da ISO 2788 (1986, tradução nossa) um tesauro deve possuir além da definição dos termos do vocabulário, informações extras, como: notas e descrições, assim como os seguintes tipos de relacionamentos entre termos, baseados em relações hierárquicas, partitivas, associativas e de equivalência:

- a) *Broader Term* (BT) ou Termo Genérico (TG): hierárquica;
- b) *Narrower Term* (NT) ou Termo Específico (TE): hierárquica;
- c) *Broader Term Partitive* (BTP) ou Termo Genérico Partitivo (TGP): partitiva;
- d) *Narrower Term Partitive* (NTP) ou Termo Específico Partitivo (TEP): partitiva;
- e) *Related Term* (RT) ou Termo Associado (TA): associativa;
- f) *Used For* (UF) ou Usado Para (UP): equivalência;
- g) *USE* ou Uso (U): equivalência.

Um exemplo de Tesauro de um domínio específico é o *Transportation Research Thesaurus* (TRT), um tesauro que tem a finalidade de melhorar a indexação e recuperação de informações sobre transportes, mantido pela *Transportation Research Board* (TRB). O dicionário de sinônimos cobre todos os modos e aspectos referentes a transportes. O TRT proporciona uma linguagem comum e consistente entre produtores e usuários de informações de transporte. A página da Web do TRT permite pesquisar e navegar o dicionário de sinônimos em vários formatos, sendo possível recuperar a "família" (relações mais gerais, mais restritas e termos relacionados) (TRB, 2012, tradução nossa). A figura 28 exemplifica uma parte do

tesauro TRT.

Figura 28 – Parte do tesauro TRT.

The screenshot displays the 'Transportation Research Thesaurus' interface. At the top, the title 'Transportation Research Thesaurus' is prominently displayed. Below the title, there is a 'Term Details' section. This section includes a search bar labeled 'TRT Keywords:' with a 'Search' button. Below the search bar, there are navigation options: 'Display Hierarchical | Alphabetical | KWOC | KWIC'. A breadcrumb trail shows the path: 'Top Terms > Transportation operations > Freight and passenger services > Freight service > Air cargo'. The main content area for 'Air cargo (Bbbc)' includes a 'Definition' (Freight, mail or other commercial articles carried on an aircraft. (Source: Team)), 'Use For' (Air freight), 'Broader Term' (Freight service (Bbb)), 'Related Terms (Hierarchical)' (Intermodal services (Bbba), Unitized cargo (Bbbb)), and 'Related Terms (Associative)' (Air freight forwarders (Nbcca)). On the right side, a vertical menu is partially visible with options like 'Fir', 'Sea', 'TR', 'Re', 'TR', 'Exp', 'nar', 'TR', 'Re', 'TR'.

Fonte: TRB (2012).

Soergel et al (2004 apud MURAKAMI, 2005) resumem algumas limitações de tesouros como forma de representação de conhecimento:

- a) falta de uma abstração conceitual: tesouros são coleções de termos, organizados em uma ordem hierárquica e interligada com alguns relacionamentos muito gerais e básicos. A distinção entre um conceito (significado) e sua lexicalização (palavras) não cria consistência, não refletindo o modo humano de entender o mundo em termos de significado e linguagem;
- b) cobertura semântica limitada: a maioria dos tesouros não diferencia conceitos em tipos e têm um conjunto muito limitado de relacionamentos entre conceitos, distinguidos somente entre relacionamentos hierárquicos e associativos, não suportando inferências;
- c) falta de consistência: devido à falta de precisão semântica dos relacionamentos nos tesouros, eles são aplicados inconsistentemente, criando ambiguidade na interpretação dos relacionamentos e resultando em uma estrutura

- semântica interna total que é irregular e não prognosticável;
- d) limitado processamento automático: tradicionalmente tesouros são projetados para indexar e formular consultas por pessoas e não para processamento automatizado. A semântica ambígua que caracteriza muitos tesouros os torna não adequados para processamento automático.

3.1.5 Comparativo das formas de representação de conhecimento

Hodge (2000, tradução nossa) comenta que apesar da diversidade de sistemas de organização e representação de conhecimento, existem algumas características críticas para a utilização das mesmas, destacando-se:

- a) visão particular do mundo em uma coleção de conceitos e seus itens;
- b) uma mesma entidade pode ser caracterizada de diferentes maneiras;
- c) semelhança entre o conceito expresso em um KOS e o objeto do mundo na qual o conceito de refere, assim como o inverso.

Os estudos de Berners-Lee (2009), Hodge (2000) e Zeng (2008) elencaram características e pontos as formas de representação de conhecimento descritas nos tópicos anteriores, de acordo com a sua estrutura (nível sintático) e sua representação (nível semântico) respectivamente. As características estruturais foram divididas em: descritores de recursos (URI), linguagem de marcação semântica, representação semântica e interoperabilidade, as características semânticas foram divididas em: eliminação de ambiguidade, controle de sinônimo, relação hierárquica, relação associativa e propriedades. Estas características culminaram no comparativo demonstrado no quadro 3, avaliadas pelos autores de forma quantitativa quanto ao nível de presença das características em cada forma de representação de conhecimento, onde, no quadro, a não presença do caractere “X” representa que a forma de representação não apresenta determinada característica, um “X” corresponde pouca presença e assim sucessivamente até os cinco “X”, denotando que a forma de representação apresenta muita presença de determinada característica.

Quadro 3 - Comparativo das características das formas de representação de conhecimento.

Característica	Forma de representação de conhecimento			
	Metadados	Mapas conceituais	Ontologias	Tesouros
Descritores de recursos (URI)			XXX	XX
Linguagem de marcação semântica	XX		XXXX	XXX
Representação semântica	XX		XXXX	XXX
Interoperabilidade	X		XXXX	XXX
Eliminação de ambiguidade	XX	X	XX	XXXX
Controle de sinônimo		X	XX	XXXX
Relação hierárquica		XXX	XXX	XXX
Relação associativa		XX	XXXXX	XXXX
Propriedades			XXXXX	

Fonte: Adaptado de Berners-Lee (2009); ZENG (2008).

Em virtude da expressiva presença de importantes características no que tange a representação de conhecimento em páginas da Web, este trabalho passa a focar-se exclusivamente em ontologias e tesouros como formas de representação de conhecimento, mais precisamente em suas evoluções: ontologias de fundamentação e tesouros conceituais.

3.2 ONTOLOGIAS DE FUNDAMENTAÇÃO

Ontologias de fundamentação são uma evolução de ontologia formal, que no grau de formalismo está entre uma ontologia semi-formal e uma rigorosamente formal, assim como também está classificada como uma ontologia de nível superior. Ontologias de fundamentação identificam categorias gerais de certos aspectos da realidade que não são específicos a um campo científico, descrevendo conhecimento independentemente de linguagem, de um estado particular das coisas ou ainda do estado de agentes (GUIZZARDI, 2005, tradução nossa). Segundo Schneider (2003, tradução nossa) ontologias de fundamentação são teorias axiomáticas de nível superior independentes de domínio, como: atributo, objeto, evento, partição, dependência e relação espaço-temporal. Borgo e Masolo (2009) caracterizam ontologias de fundamentação da seguinte maneira:

- a) têm vasto alcance;
- b) podem ser reutilizáveis em cenários de modelagem diferentes;
- c) são filosófica e conceitualmente bem fundamentadas;

d) são semanticamente transparentes e, portanto, axiomatizadas.

Medeiros (2011) comenta que ontologias de fundamentação são altamente reutilizáveis, caracterizando-se pelo comprometimento ontológico, determinando regras de restrição, conceitos, categorias e meta propriedades, ainda como apresentando os seguintes elementos:

- a) relação entre identidade e classificação;
- b) noções de tipos e suas instâncias;
- c) objetos, e suas propriedades intrínsecas;
- d) distinções entre tipos e suas relações;
- e) relações parte-todo;
- f) classes com suas propriedades, relacionamentos, valores e regras.

3.2.1 Exemplos de ontologias de fundamentação

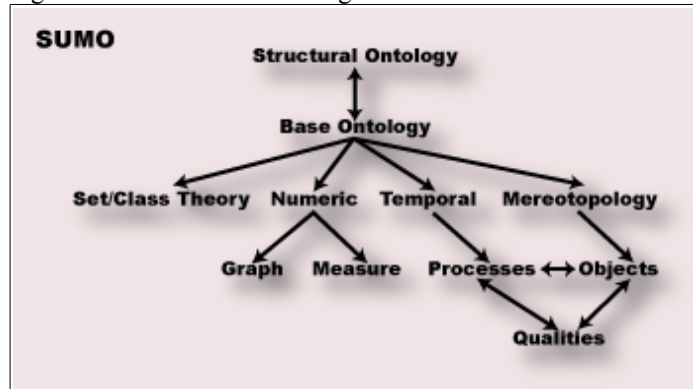
Entre as ontologias de fundamentação mais reconhecidas encontram-se a *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE), *General Formal Ontology* (GFO), *Suggested Upper Merged Ontology* (SUMO) e *Unified Foundational Ontology* (UFO). DOLCE foi uma das primeiras Ontologias de Fundamentação criadas, sendo que seu objetivo nunca foi ser um padrão universal ou uma ontologia oficial e sim ser base para a construção de novas ontologias, um ponto de partida, assim como também para esclarecer os pressupostos ocultos subjacentes em ontologias existentes ou recursos linguísticos, tais como o WordNet (LABORATORY FOR APPLIED ONTOLOGY, 2006, tradução nossa).

GFO é uma ontologia para a modelagem conceitual focada na integração de processos e objetos, que inclui categorias como elaborações de objetos, processos, tempo e espaço, propriedades, relações, papéis, funções, fatos e situações (ONTO-MED, 2010, tradução nossa).

SUMO foi desenvolvida por Adam Pease dentro do grupo de trabalho Ontologia do *Institute of Electrical and Electronics Engineers* (IEEE), sendo uma ontologia de nível superior que objetiva a interoperabilidade de dados, busca e recuperação de informação, inferência automatizada e processamento de linguagem natural. Limita-se a conceitos genéricos suficientes para resolver uma ampla gama de áreas de domínios, fornecendo uma

estrutura e um conjunto de conceitos sobre os quais poderão ser criadas ontologias de domínio, como: medicina, finanças, engenharia (ONTOLOGY PORTAL, 2012, tradução nossa). A figura 29 exemplifica a estrutura da ontologia SUMO.

Figura 29 - Estrutura da ontologia SUMO.



Fonte: Ontology Portal, 2012).

Segundo Martins (2009) a ontologia UFO foi concebida com base em outras duas ontologias de nível superior, a DOLCE e a GFO, aproveitando-se as características positivas e corrigindo as limitações das mesmas. Guizzardi (2005, tradução nossa) refere-se a UFO como uma ontologia de apoio ao modelo conceitual e esclarece os fragmentos da mesma:

- a) UFO-A: ontologia UFO base, modela objetos e propriedades;
- b) UFO-B: extensão da UFO-A, modela eventos;
- c) UFO-C: extensão da UFO-B, modela entidades sociais e intencionais.

Medeiros (2011) complementa que a ontologia UFO-A aumenta o nível de semântica na modelagem conceitual de um determinado domínio de conhecimento, apresentando todos os elementos de uma ontologia de fundamentação.

3.2.2 Ferramentas para ontologias de fundamentação

Existem várias ferramentas para construção de ontologias, sendo uma das principais a ferramenta Protégé, anteriormente já mencionada, que possibilita a construção de ontologias de forma completa. Entretanto essas ferramentas possuem um alto grau de complexidade de desenvolvimento, o que não vai de encontro a construção de ontologias de fundamentação, que têm como forte característica um nível mais geral de abstração. Em consequência disso tem-se a necessidade de ferramentas que auxiliam no desenvolvimento deste tipo de ontologia, objetivando a simplicidade na definição de suas características, neste

propósito, dois editores de destacam: OntoUML e DOE.

OntoUML Editor é um editor gráfico, gratuito e de código fonte aberto que suporta a criação de ontologias de fundamentação, baseando-se na ideia de construção de modelos conceituais e ontologias de domínio. OntoUML Editor protege o usuário da complexidade de princípios ontológicos, porém também reforça estes princípios, verificando de forma automática as restrições formais, excluindo erros de sintaxe nos modelos desenvolvidos. OntoUML Editor pode ser utilizado em qualquer sistema operacional e não possui limitações diferenciadas de processamento e armazenamento, podendo ser utilizado em conjunto com a ferramenta de desenvolvimento Eclipse (BENEVIDES, 2012, tradução nossa).

DOE é outro editor de ontologias simples e gratuito, visa a construção de ontologias de forma hierárquica, com relacionamentos e definição de conceitos e sinônimos, apoiando-se bastante na Ciência da Informação no que diz respeito a construção de taxonomias. Por meio das taxonomias são definidas as restrições, inserindo semântica na ontologia e assim sendo traduzida em uma linguagem de representação de conhecimento. O objetivo de editor DOE não é ser uma ferramenta completa para construção de ontologias e sim ser um apoio no início da construção de uma ontologia, oferecendo um ponto de vista humano e compreensível. DOE é um aplicativo feito em Java e portanto pode ser utilizado em qualquer sistema operacional, bastando que o mesmo tenha instalado uma Máquina Virtual Java (ISAAC; TRONCY, 2012, tradução nossa).

3.2.3 Representação de conhecimento em ontologias de fundamentação

A evolução de ontologias para ontologias de fundamentação culminou na presença de novas características, gerando uma representação de conhecimento mais rica em comparação com as ontologias até então, principalmente no que tange a redução de ambiguidade, reutilização e definição de conceitos. Como já mencionado anteriormente ontologias de fundamentação são oriundas de características elencadas por Guarino (1998), quanto a generalidade, e Uschold e Gruninger (1996), quanto ao formalismo, resultando assim como característica fundamental de Ontologias de Fundamentação, uma ontologia de nível superior e formal. A partir destas características e das levantadas por Borgo e Masolo (2009), Guizzardi (2005) e Ramalho (2010) chega-se a comparação apresentada no quadro 4, onde

compara-se ontologias e ontologias de fundamentação, esta última como uma evolução das primeiras.

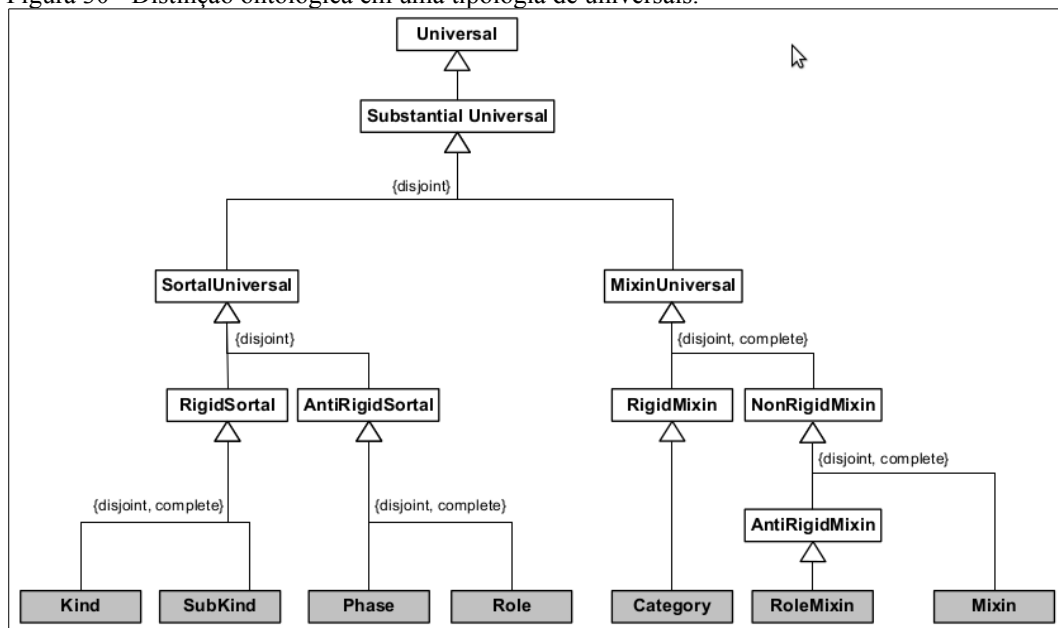
Quadro 4 - Comparativo entre as características das ontologias a partir de sua evolução.

Característica	Evolução	
	Ontologias	Ontologias de fundamentação
Classes	X	X
Propriedades	X	X
Regras e axiomas	X	X
Instâncias	X	X
Valores	X	X
Relacionamentos	X	X
Tipos e sub-tipos		X
Fases		X
Funções		X
Categorias		X
Funções genéricas		X
Genéricas		X

Fonte: Do Autor.

As características encontradas somente em Ontologias de Fundamentação estão baseadas na modelagem conceitual e podem ser distinguidas dentro do conceito de ontologias de fundamentação por meio da figura 30.

Figura 30 - Distinção ontológica em uma tipologia de universais.



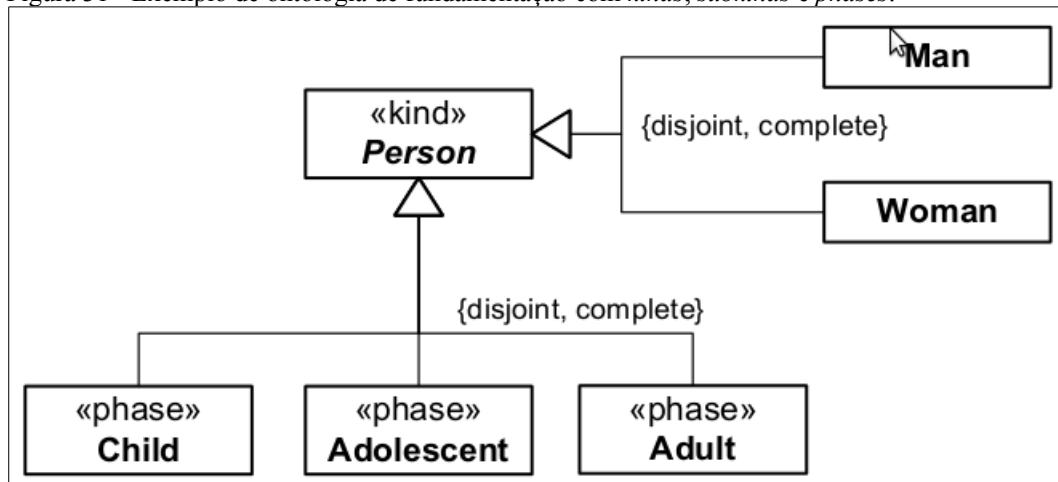
Fonte: Guizzardi (2005).

Em contraste com o quadro 4, na figura 30, os elementos: Kind, SubKind, Phase, Role, Category, RoleMixin e Mixin são respectivamente: Tipos, Sub-tipos, Fases, Funções, Categorias, Funções genéricas e Genéricas. Os demais elementos são contextualizados por Medeiros (2011) a seguir:

- a) *Universal*: são entidades que comportam um conjunto de características;
- b) *Substantial Universal*: são entidades que mantêm sua identidade no tempo;
- c) *Sortal Universal*: são entidades que permitem diferenciar suas instâncias a partir de suas características;
- d) *Mixin Universal*: são classes que agregam entidades, ainda que de essência diferenciada;
- e) *Rigid Sortal*, *Rigid Mixin*: quando a entidade é uma instância durante todo tempo em que o *substantial universal* existir, independente do mundo a que é aplicado;
- f) *Anti Rigid Sortal* e *Anti Rigid Mixin*: é quando a entidade instanciada a ele puder deixar de existir ao longo de sua existência, assim, não é aplicável a totalidade de instâncias do *Universal*;
- g) *Non Rigid Mixin*: são entidades aplicáveis a pelo menos uma de suas instâncias.

A figura 31 exemplifica uma parte de uma possível ontologia de fundamentação, apresentando o uso de *kinds* (tipos), *subkinds* (sub-tipos) e *phases* (fases).

Figura 31 - Exemplo de ontologia de fundamentação com *kinds*, *subkinds* e *phases*.



Fonte: Guizzardi (2005).

Medeiros (2011) explica que ontologias de fundamentação representam o conhecimento por meio de uma visão Aristotélica do mundo, ou seja, a origem das primeiras ontologias, que classificam os objetos estabelecendo categorias gerais para os mesmos, estes sendo os *kinds*, organizando os conceitos em instâncias, universais, propriedades e classes, isso faz com que ontologias de fundamentação representem o conhecimento com superordenação e/ou subordinação, possuindo identidade, rigidez e dependência, devendo-se principalmente ao fato de que as mesmas explicitam os conceitos por meio de sua tipologia, estabelecendo a posição dos mesmos dentro da cadeia de elementos, evitando-se assim por exemplo, proposições errôneas.

3.3 TESAUROS CONCEITUAIS

Segundo Campos e Gomes (2006), a origem dos tesouros conceituais se deu com a insuficiência da descrição de uma palavra ou termo em um vocabulário controlado para designar o conteúdo de uma informação, como por exemplo: quando o vocabulário é convertido para outras línguas, onde dependendo da cultura o mesmo termo passa a ter outro conceito. Durante vários anos várias soluções foram discutidas para resolver o problema, como a utilização de uma expressão verbal na unidade do conceito (CAMPOS; GOMES, 2006); a Teoria da Classificação Facetada de Ranganathan, que se baseava na junção de conceitos similares em categorias (MEDEIROS, 2011); o *Thesaurofacet*, que utiliza a Teoria da Classificação Facetada, na parte da organização de uma sintaxe, adicionando uma

organização dos conceitos de um domínio específico e permitindo a construção de tesouros tanto na forma sistemática quanto na alfabética, objetivando a compreensão das relações entre conceitos (CAMPOS, 2001) e a Teoria do Conceito de Ingetraut Dahlberg, 1970, que uniu-se as teorias anteriores, gerando uma nova forma de construir tesouros resolvendo os problemas encontrados até então. Esta teoria se baseia na utilização de categorias como forma de organização de conceitos, dando origem aos tesouros conceituais (MEDEIROS, 2011).

Medeiros (2011) conceitua que tesouros conceituais são formados por uma parte alfabética perante os termos, assim como a especificação das relações entre eles, e a parte sistemática, onde os termos são definidos por um conteúdo conceitual, ganhando destaque em suas definições e recebendo uma categorização de conceitos, o autor então sintetiza tesouros conceituais com base na Teoria da Classificação Facetada e na Teoria dos Conceitos nos seguintes elementos:

- a) conceitos: representam os termos ou palavras;
- b) categorias e classes: estabelecem a ordenação lógica e hierárquica dos conceitos;
- c) relações entre conceitos: categorial, hierárquica, partitiva e funcional-sintagmática;
- d) definições: permitem posicionar um conceito dentro de um sistema de conceitos.

Maniez (1993 apud GONZÁLEZ, 2011) determina que um Tesouro Conceitual é um modelo de relações associativas que categoriza os termos e conceitos no contexto do usuário, uma rede semântica em que cada nó contém um único conceito semântico com uma possível série de conceitos associados, possuindo relações preferenciais, hierárquicas e associativas.

3.3.1 Exemplos de tesouros conceituais

O Tesouro do Folclore e Cultura Popular Brasileira, mantido pelo Centro Nacional de Folclore e Cultura Popular (CNFCP), criado em 2004 por especialistas em folclore brasileiro, contemplando as seguintes categorias: alimento, artefato, associação, atividade produtiva, atividade ritual, indivíduo, matéria-prima, literatura oral, medicamento, construção artesanal, sistema de crença, lugar, tempo, atividade musical, atividade narrativa, prática

religiosa e farmacopeia popular. Este tesouro possui todas as relações de equivalência, hierarquia, associação e partitativa, assim como descrições, notas explicativas e o acréscimo de um qualificador semântico colocado entre parênteses, exemplo: balaio (artefato), balaio (dança) (CNFCP, 2012). A figura 32 exemplifica uma parte do Tesouro de Folclore e Cultura Popular Brasileira.

Figura 32 - Parte do Tesouro de Folclore e Cultura Popular Brasileira.

Jogo

Atividade ritual, submetida a regras, realizada com ou sem o uso de objetos, individual ou coletiva, cuja finalidade fundamental vise, prioritariamente, à disputa entre seus participantes. A ênfase está no caráter competitivo do evento.

Termo Genérico ▶

- ▶ **Atividade ritual**

Termos Específicos

- ▶ Bate-coxa
- ▶ Capoeira
- ▶ Finca
- ▶ Jogo com animal
- ▶ Jogo de azar
- ▶ Jogo de bola
- ▶ Malha
- ▶ Pica-pau (jogo)

Termos Associados

- ▶ Bola
- ▶ Brincadeira

Fonte: Do Autor.

O Tesouro de Cultura Material dos Índios no Brasil representa a forma tesourográfica pela qual é realizado o controle terminológico das obras indígenas disponíveis no Museu do Índio. Visando permitir a comunicação com acervos diferenciados da instituição, sendo um modelo de representação do domínio de cultura material indígena brasileira, também fazendo parte de um movimento contra a falta de registros da cultura indígena (MEDEIROS, 2011).

3.3.2 Ferramentas para tesouros conceituais

Conforme Campos et al (2006) não existem muitos estudos relativos a ferramentas e tesouros, no que tange a elaboração e manutenção destes, assim como não existe um software de referência para tal caso. Sendo que em tesouros conceituais o problema aumenta, pois segundo as autoras os softwares preocupam-se muito com a questão operacional, deixando de lado o fato da metodologia para construção de tesouros, e nos tesouros conceituais essa lacuna é exemplificada pela falta do recurso de categorização com base no conceito, este, um dos pontos fortes dos tesouros conceituais sobre tesouros comuns. Entretanto algumas ferramentas para construção de tesouros se destacam e conseguem ajudar no processo de construção de tesouros conceituais, fundamentando-se nas características destes, incluindo categorização simples, estas ferramentas são o MultiTes e o Synaptica.

MultiTes é uma ferramenta comercial baseada em tecnologia Web com finalidade de criar e gerenciar tesouros, taxonomias e vocabulários controlados, objetivando também a publicação e gerenciamento dos mesmos na internet ou na própria intranet. O MultiTes é dividido em 3 partes:

- a) MultiTes Pro: ambiente de desenvolvimento;
- b) MultiTes WDK: publicação de tesouros na Web;
- c) MultiTes EDK: instalação de servidores para tesouros na Web.

MultiTes pode ser utilizado nas plataformas Windows 2000, XP ou superior (MULTISYSTEMS, 2012, tradução nossa).

Synaptica é uma ferramenta profissional e comercial para desenvolvimento de tesouros, taxonomias e afins, tendo sua comercialização baseada em duas categorias: *Enterprise* e *Express*, a primeira voltada empresas de grande porte e a segunda para empresas de pequeno porte. Entre os principais objetivos do Synaptica estão:

- a) redução de ambiguidade em consultas;
- b) vocabulário com indexação de sinônimos;
- c) hierarquias navegáveis;
- d) links de associação;
- e) navegação facetada;
- f) visualização nos mais variados formatos textuais e gráficos.

O Synaptica não pode ser baixado na internet, precisa ser solicitada uma análise

comercial, assim como o agendamento de demonstração e verificação de requisitos funcionais e estruturais (SYNAPTICA, 2012, tradução nossa).

3.3.3 Representação de conhecimento em tesauros conceituais

A construção de tesauros sempre se baseou em aspectos metodológicos, diferenciando-se por vertentes, teorias e características, motivada principalmente pelo uso dos mesmos dentro da Ciência da Informação, mais precisamente a recuperação de informações. Segundo Campos (2001), as teorias do conceito e da classificação originaram uma nova tendência para tesauros, os tesauros conceituais, um tesouro baseado no estabelecimento do termo/conceito e das relações entre eles, assim como uma mudança de sentido, onde o conteúdo conceitual dos termos passa a ser determinante em relação a base linguística.

Para Campos et al (2006) essa evolução vêm de encontro com a vertente europeia dos tesauros, que já possuíam a categorização dos conceitos em sua construção, entretanto somente nos tesauros conceituais é que a definição dos conceitos começou a estar presente em sua completude, categorizando-o e definindo-o. A partir das características elencadas por Campos (2001), ISO 2788 (1986) e Medeiros (2011) chega-se ao comparativo entre as evoluções de tesauros, demonstrado no quadro 5.

Quadro 5 - Comparativo entre as características dos tesauros a partir de sua evolução.

Característica	Evoluções		
	Tesauros (vertente norte americana)	Tesauros (vertente europeia)	Tesauros conceituais
Termos	X	X	
Notas	X	X	X
Descrições	X	X	X
Relações de termos	X	X	
Organização alfabética	X		X
Organização sistemática		X	X
Conceitos			X
Definições para conceitos			X
Relações de conceitos			X
Categorias		X	X

Fonte: Do Autor.

Medeiros (2011) explica que tesouros conceituais ao somarem as características das vertentes norte americana e europeia conseguiram humanizar a representação de conhecimento, principalmente pelo uso de categorias e conceitos, estes últimos considerados os objetos de representação de conhecimento, pertencentes a categorias como: objetos ou entidades, eventos ou ações, entre outras categorias. O autor esclarece que os conceitos não chegam a substituir um termo na representação de conhecimento e sim agregam informações para alinhar o modelo conceitual, esta agregação é composta pelos referentes e suas características, assim representando algo do mundo real, classificados a partir do mais geral ao mais individual, por exemplo: Ponte Rio-Niterói, é um objeto do mundo real considerado como individual e Ponte é um objeto geral. O autor ainda reitera que o objeto que se conceitua nos tesouros conceituais é o objeto geral, entretanto o conhecimento é obtido por meio da análise dos objetos individuais e suas relações.

3.4 AMBIGUIDADE

Segundo Luft (2001) ambiguidade é a duplicidade de significados, falta de clareza, indecisão, onde uma palavra tem dois ou mais sentidos. Fuchs (1996, tradução nossa) complementa que a ambiguidade faz com que uma palavra ou frase seja compreendida de diferentes maneiras por um receptor. O autor classifica a ambiguidade nos seguintes tipos:

- a) morfológica: ocorre quando não é possível determinar a categoria gramatical, provocada pela policategorização, onde uma palavra pode ser um substantivo, adjetivo ou verbo;
- b) lexical: ocorre quando há vários significados para uma unidade lexical, provocado pela homografia, onde há colisão de signos linguísticos, exemplo: cobre (metal), cobre (verbo cobrir) ou pela polissemia, que ocorre quando uma só e mesma expressão envolve significados distintos, sendo um único signo linguístico, exemplo: arquivo (móvel, instituição, conjunto de documentos);
- c) sintática: ocorre na estruturação da frase em constituintes hierarquizados, quando se definem as ligações que se estabelecem entre os sintagmas, provocada principalmente por frases preposicionais, exemplo: a professora de dança espanhola (A professora é espanhola ou é a dança que é

espanhola?);

- d) predicativa: ocorre na interpretação das relações temáticas que articulam predicados, argumentos e participantes, exemplo: A crítica deste autor (autor é o objeto da crítica ou é o agente da crítica?);
- e) semântica: ocorre quando há mais de uma interpretação possível para o relacionamento dos termos em uma frase, exemplo: Ela não chora mais porque ele partiu (ela chorava porque ele havia partido ou ela parou de chorar uma vez que ele já foi embora?);
- f) pragmática: ocorre no cálculo dos valores enunciativos e/ou na reconstrução destes valores, exemplo: Paulo vai à escola. (ele é estudante ou ele está indo à escola neste momento?).

Neste trabalho será dada ênfase a ambiguidade semântica, haja vista que o objetivo do mesmo é a qualidade semântica na representação de conhecimento, sendo um fator de aumento de qualidade a redução de ambiguidade semântica. Como já destacado por Fuchs (1996, tradução nossa) a ambiguidade semântica se encontra no relacionamento dos termos em uma frase, entretanto é válido ressaltar que a ambiguidade semântica também é gerada pelo próprio termo, causando uma multiplicidade de conceitos relacionados com a sua aplicabilidade, e até mesmo por polissemia (SILVA, 2006). Pagani (2009) reforça constatando que a ambiguidade semântica se deve ao fato de os pronomes poderem ter diversos antecedentes e enunciados, denominando-se como correferencialidade, ou ainda como uma indeterminação referencial dos pronomes, com isso, o autor divide a ambiguidade semântica em:

- a) anafórico: quando o pronome (demonstrativo, relativo ou adverbial) tem um antecedente linguístico;
- b) dêitico: quando o pronome (demonstrativo) tem por objetivo localizar o fato no tempo e espaço sem defini-lo.

Silva (2006) realiza uma classificação dos fenômenos linguísticos que causam a ambiguidade, inclusive e principalmente a ambiguidade semântica:

- a) policategorização: ocorre quando uma palavra pode ser categorizada como sendo um substantivo, adjetivo ou verbo;
- b) ambivalência: é a presença simultânea de dois componentes de sentidos contrários: seu domínio de eleição e o dos sentimentos e atitudes, um

- dualismo de relações;
- c) metáfora: é o emprego de palavra fora do seu contexto, por efeito de analogia ou comparação. Exemplo: “A Amazônia é o pulmão do mundo”. A metáfora está contida nos termos “Amazônia” e “pulmão”;
 - d) duplo sentido: diferencia-se de ambiguidade pois este fenômeno não existe sem a presença do sujeito falante;
 - e) indeterminação: ocorre na ausência de determinação ou presença de relativismo;
 - f) vaguidade: é a falta de informação e/ou especificação. Exemplo: “José chutou Carlos”. A vaguidade está na resposta para a pergunta: Porque e como José chutou Carlos?;
 - g) polissemia: ocorre nas palavras que têm a capacidade de assumir significados diferentes, mas mantendo uma relação de sentido entre elas;
 - h) anáfora: é quando se tem uma repetição sistemática da mesma palavra no início de diferentes frases ou de membros da mesma frase. Exemplo: “Ao depositar a panela em cima da mesa, quebrei-a ”, a anáfora está contida em determinar o que foi quebrado;
 - i) sinonímia: ocorre quando uma palavra é substituída em uma determinada frase, sem alterar o contexto de verdade;
 - j) homonímia: são palavras que se escrevem e pronunciam da mesma maneira, mas que têm significado e origem diferentes. Exemplo: “banco”, móvel e instituição financeira;
 - k) homografia: é quando uma palavra aceita vários significados. Exemplo: “cobrir”, verbo cobrir e tipo de metal.

A ambiguidade faz parte tanto da Ciência da Informação quanto da Ciência da Computação, oriunda primeiramente dos sistemas de recuperação de informação e ultimamente com a utilização da Web Semântica no que tange a representação de conhecimento. Silva (2006) revela que a ambiguidade está contida tanto no emissor de uma mensagem (*a priori*) quanto no receptor da mesma (*a posteriori*), o primeiro gera a ambiguidade em um texto e/ou enunciado mal formulado, o segundo por conter uma diversidade de condições, como cultural, histórica e/ou de contexto, resultando em uma ambiguidade na interpretação, relativa a linguagem natural utilizada. Silva (2006)

complementa que mesmo a partir de um texto gerado de forma ambígua o receptor precisa ter informações necessárias para realizar a interpretação correta, sendo assim a ambiguidade *a priori* pode inicialmente não ser um problema, desde que se tenha opções de interpretação no momento da recepção da mensagem, ou seja, a ambiguidade é reconhecida e pode ser resolvida *a posteriori*.

Bräscher (2002) afirma que a ambiguidade causa ruído na recuperação da informação, onde a partir de um mesmo termo pode-se encontrar informação relevante e irrelevante, exemplo: na frase “Neutralização de contaminação com leite” é encontrada uma ambiguidade sintática, não permitindo que um sistema de recuperação de informação decida se está buscando-se “neutralização de contaminação” ou “neutralização com leite”. Segundo Ferneda (2006) um aspecto que também gera ambiguidade na recuperação de informação é o conteúdo utilizado pelos usuários para execução de uma busca, haja vista, que geralmente este conteúdo são palavras isoladas, sem a utilização de expressões e/ou ligações, dificultando que um sistema de recuperação de informação interprete o significado correto do termo pesquisado, pois o termo apresentará ambiguidade pela falta de termos relacionados por parte do usuário.

Para Bräscher (2002) um sistema de recuperação de informações necessita de uma distinção entre os significados segundo um contexto, conseguindo assim maior precisão no resultado da busca não gerando ambiguidade de conceitos, entretanto, a maioria dos sistemas de recuperação de informações voltados à Web utilizam métodos estatísticos e análises sintáticas para extração de termos, requerendo pouco esforço ao sistemas, mas não solucionando a ambiguidade, diferentemente de um sistema que utiliza linguagens documentárias.

Fuchs (1996, tradução nossa) esclarece que toda forma de representação de conhecimento que suporta associação de vários conceitos a um termo é virtualmente ambígua, necessitando ser analisada dentro de um contexto específico, afim de se tornar unívoca ou ainda ser considerada efetivamente ambígua. Bräscher (2002) complementa que nem todo tipo de informação contextual pode ser representado formalmente, o que na maioria das vezes impede a resolução de problemas de ambiguidade durante a representação do conhecimento e recuperação da informação.

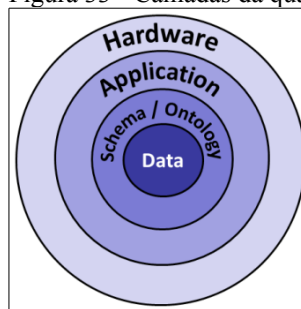
Brandt (2009) conclui que a ambiguidade na representação do conhecimento faz com que a informação contida em um documento tenha pouca relevância, ou seja, falta de

correspondência entre o documento e a consulta. Berners-Lee et al (2001, tradução nossa) finalizam que uma das principais preocupações da Web Semântica é a ambiguidade na representação de conhecimento, pois a precisão na recuperação da informação requer diferenciação de representação nos variados conceitos de uma palavra em uma página da Web, caso contrário, os motores de busca recuperarão todas as páginas que a palavra é utilizada e não o conceito em si. A diferenciação dos conceitos, segundo os autores pode ser obtida com a utilização de URIs, onde para cada conceito de uma palavra existe uma URI diferente.

3.5 QUALIDADE SEMÂNTICA

Antes de definir o que é qualidade semântica é preciso definir o que é qualidade de um dado e/ou informação. Wang e Strong (1996, tradução nossa) definem que um dado tem qualidade quando está adequado para a utilização por parte dos consumidores de dados. A ISO 8000-102 (2009, tradução nossa) define que a qualidade dos dados é o grau que preenche os seus requisitos. Com base nas duas definições pode-se concluir que um dado tem qualidade quando o mesmo está formalmente definido para atender as expectativas do consumidor de dados. Fürber (2011) ilustra por meio da figura 33 as camadas que podem gerar a qualidade de um dado e/ou informação em um sistema de informação.

Figura 33 - Camadas da qualidade da informação.



Fonte: Fürber (2011).

Segundo Fürber (2011, tradução nossa), qualquer uma das camadas podem gerar problemas na qualidade da informação, sendo que cada camada tem sua responsabilidade e agente ativo, por exemplo:

- a) *hardware*: computador, performance;
- b) *application*: motores de busca, algoritmos;

c) *schema/ontology*: representação, semântica;

d) *data*: os dados em si.

Wang e Strong (1996, tradução nossa) identificam no quadro 6 as 15 dimensões mais importantes de requisitos de qualidade de dados aos consumidores de dados.

Quadro 6 - Requisitos de qualidade de dados importantes aos consumidores de dados.

Categoria	Dimensão	Definição
Intrínseco	Credibilidade	É aceito ou considerado como verdadeiro.
	Precisão	Está correto, livre de erros.
	Objetividade	É imparcial.
	Reputação	É confiável, tanto origem como conteúdo.
Contexto	Valor agregado	Proporciona vantagens em sua utilização.
	Relevância	É aplicável.
	Oportunidade	Idade do dado é apropriada para uso.
	Plenitude	Profundidade do dado é suficiente.
	Quantidade	Quantidade e o volume do dado são apropriados.
Representação	Interpretabilidade	Idioma do dado está apropriado e unidades e definições estão claras.
	Compreensibilidade	Não possui ambiguidade, sendo facilmente compreendido.
	Consistência	É apresentado sempre com o mesmo formato, sendo compatível com os dados anteriores.
	Concisa	Tem apresentação compacta, mas é completo quanto seu objetivo.
Acesso	Acessibilidade	Facilmente e rapidamente recuperável.
	Segurança	Pode ter seu acesso restringido, mantido em segurança.

Fonte: Wang e Strong (1996, tradução nossa).

Ainda segundo Wang e Strong (1996, tradução nossa) as características que fazem uma informação ter qualidade semântica estão categorizadas na representação dos dados, envolvendo as dimensões: interpretabilidade e compreensibilidade na significação dos dados, assim como consistência e concisão no seu formato, detalhadas nas seguintes características:

- a) completa: define completamente um conceito;
- b) sem ambiguidade: não há duas representações para o mesmo conceito;
- c) significativa: conceito está contido dentro de um vocabulário;
- d) correta: o conceito é recuperado de acordo com seu significado.

O quadro 7 apresenta os critérios para a qualidade dos dados quanto a sua representação voltados para a Web Semântica.

Quadro 7 - Critérios de qualidade dos dados quanto a sua representação para a Web Semântica.

Critério	Indicadores
Consistência	<p>Para evitar contradições de dados, deve-se:</p> <ul style="list-style-type: none"> a) não definir uma entidade como membro de classes disjuntos; b) usar corretamente as propriedades funcionais inversas, exemplo: setar valor nula em uma dessas propriedades; c) estender um vocabulário ao invés de alterar as propriedades para aumentar a expressividade do mesmo; d) usar tipos de dados homogêneos; e) não informar valores inconsistentes para propriedades.
Uniformidade	<p>Para uma uniformidades dos dados, deve-se:</p> <ul style="list-style-type: none"> a) usar um formato de dados definido, exemplo: RDF, RDFa; b) usar vocabulários definidos e reconhecidos; c) realizar referências de URIs.
Versatilidade	<p>Para representações alternativas de dados, deve-se:</p> <ul style="list-style-type: none"> a) dispor os dados em vários formatos reconhecidos; b) dispor os dados em vários idiomas; c) dispor opções de preferência de dados, exemplo: idioma; d) dispor métodos alternativos de acesso aos dados, exemplo: terminal de consulta SPARQL.
Compreensibilidade	<p>Para facilitar o consumo dos dados, deve-se:</p> <ul style="list-style-type: none"> a) rotular de forma legível as classes, propriedades e entidades; b) descrever de forma clara as classes, propriedades e entidades; c) indicar os metadados disponíveis para o conjunto de dados; d) indicar expressão regular das URIs, caso exista; e) disponibilizar um exemplo de consulta SPARQL; f) indicar alguns vocabulários utilizados; g) dispor um HTML claro; h) disponibilizar listas de discussão.

Fonte: Flemming (2011, tradução nossa).

Flemming (2011, tradução nossa) comenta que uma representação dos dados que engloba formatos, vocabulários e identificadores bem definidos gera vantagens ao consumidor de dados. Destacando-se o uso de vocabulários que aumentam a performance do processamento da lógica por um programa e ajudam humanos na compreensão semântica (se os mesmos possuírem afinidade com o vocabulário). Um HTML bem escrito e com uso de metadados internacionalmente reconhecidos ajudam tanto máquinas quanto humanos na recuperação da informação. A referenciação de URIs ajuda no aumento da quantidade de

conjunto de dados, enriquecendo o vocabulário, assim como a rotulação legível de classes, propriedades e entidades por meio de rótulos e comentários ajudam no significado do dado.

Para Berners-Lee (2009, tradução nossa) um dado tem qualidade na Web Semântica, se:

- a) possuir licença para dados abertos;
- b) for legível por máquinas;
- c) estar em um padrão aberto do W3C;
- d) conter ligação com outros dados.

3.6 AMBIGUIDADE E QUALIDADE SEMÂNTICA EM ONTOLOGIAS DE FUNDAMENTAÇÃO E TESAUROS CONCEITUAIS

Segundo Medeiros (2011) ontologias de fundamentação melhoram a comunicação entre agentes durante a exportação, recuperação e extração de informações, na Web Semântica e em um todo, facilitando a interoperabilidade entre sistemas e agentes de *software*, possuindo uma semântica baseada no mundo real por meio de um vocabulário próprio, restringindo interpretações sobre seus conceitos e diminuindo erros de representação, reduzindo a ambiguidade, testando e validando o modelo conceitual gerado por meio de um formalismo bem definido.

Guizzardi (2005, tradução nossa) comenta que ontologias de fundamentação contribuem para uma caracterização fiel da conceitualização, revelando seu compromisso ontológico, permitindo a representação do domínio empregado, possibilitando a reutilização de ontologias.

Bräscher e Schiessi (2012) ressalta que a conceitualização serve para idealizar algo inacessível na significação do conteúdo, sendo que para tanto é necessário descrever detalhadamente as características de entidades e relacionamentos, tornando o conceito implícito em explícito, permitindo o acesso da mesma informação para pessoas e máquinas. As autoras ainda complementam que para um conceito estar bem definido em uma estrutura de representação de conhecimento o mesmo precisa permitir a generalização.

Medeiros (2011) explica que ontologias de fundamentação têm em sua estrutura um formato que permite de forma automática a generalização, assim como é capaz de deixar o conceito explícito, possuindo superordenação/subordinação das entidades, onde por exemplo,

um *Phase* será sempre um *Anti Rigid Sortal*, não podendo um *Kind*, que é um *Rigid Sortal*, estar subordinado a um *Phase*.

Guizzardi (2005, tradução nossa) destaca que uma das razões da qualidade obtida com a utilização de Ontologias de Fundamentação se deve ao fato de as mesmas ativarem o uso de *labels* (rótulos), que identificam a natureza conceitual das entidades de acordo com o estado do mundo em que se encontram.

Schneider (2003, tradução nossa) conclui que ontologias de fundamentação por meio da explicitação de categorias gerais independentes de domínio, como objetos, atributos, eventos, partição, dependência e conexões espaço-temporais, permitem que além da própria generalização dos conceitos exista também uma forma mais clara de representá-lo, delimitando o seu significado.

Para Medeiros (2011), a principal vantagem dos tesouros conceituais são as definições dos conceitos, que se bem constituídas permitem um bom posicionamento do mesmo dentro da categorização dos conceitos, ressaltando ainda que esta definição manifesta as características do objeto e sua função em um domínio, evidenciando o que o objeto realmente é, ou seja, sua natureza. Dalberg (1983 apud MEDEIROS, 2011) cita que a definição de um conceito envolve:

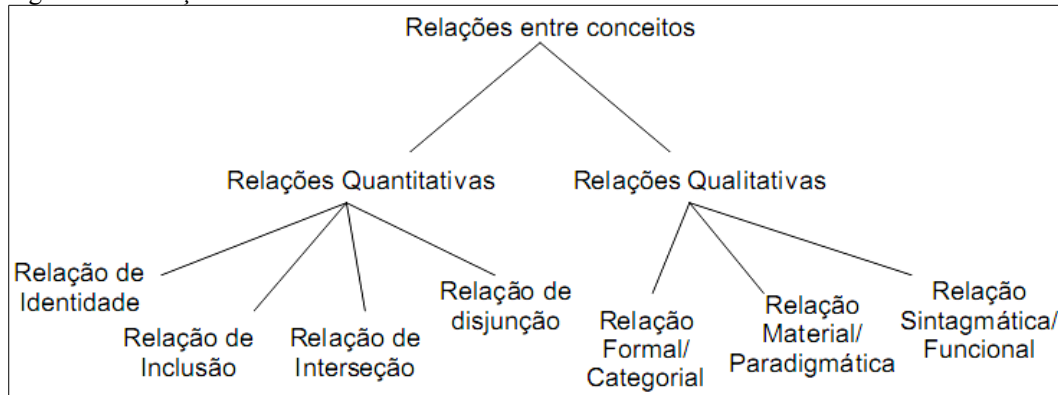
- a) referentes dos conceitos;
- b) a(s) categorias a(s) qual(is) pertence(m) um conceito;
- c) a expressão verbal adequada aos usuários.

González (2011) complementa que a utilização de verbos ativam a significação dos descritores de conceitos, realizando uma relação explícita, objetivando a não ambiguidade conceitual, assim como auxiliam na interoperabilidade entre sistemas, mostrando os agentes que interagem com o sistema e como fazem. As categorias segundo o autor permitem que conceitos possam ser relacionados usando inúmeras possibilidades da linguagem natural, simplificando o modelo de representação de conhecimento e enriquecendo a recuperação da informação.

Campos (2001) comenta que a agregação de conceitos aos termos nos tesouros conceituais deixa a representação de conhecimento mais rica, pois consegue abstrair a informação para um nível menos complexo e mais claro, assim como seus relacionamentos ganham a possibilidade de novas relações, adquirindo verbalização e permitindo o posicionamento do conceito dentro de um sistema de conceitos. Na figura 34 é possível

perceber as novas relações que podem ser obtidas entre conceitos.

Figura 34 - Relações entre conceitos.



Fonte: Campos (2001).

Campos et al (2006) concluem que tesouros carecem de especialistas em biblioteconomia, dentro da área de Ciência da Informação, mas que as evoluções da metodologia na linhas de construção de tesouros estão simplificando o processo, permitindo que softwares sejam capazes de ajudar neste processo de construção de tesouros, aproximando-os da área da Ciência da Computação. Isto pode ser percebido com a mudança no formato de representação de conhecimento estabelecido pelos tesouros conceituais, onde os tesouros agregam categorias e conceitos, e ao mesmo tempo, sistematizam o conhecimento e aproximam o vocabulário do entendimento humano.

4 TRABALHOS CORRELATOS

Do ano 2000 até os dias atuais vários trabalhos voltados para a Web Semântica foram e estão sendo desenvolvidos, variando-se entre dissertações, teses e principalmente artigos, envolvendo estudos sobre a arquitetura, ferramentas, linguagens e formas de representação de conhecimento em páginas da Web, estas formas de representação norteiam-se basicamente em ontologias e tesouros. Relacionamos neste capítulo alguns trabalhos importantes que vão de encontro com o propósito deste trabalho, particularmente nas comparações entre ontologias e tesouros e qualidade semântica.

4.1 TESAUROS CONCEITUAIS E ONTOLOGIAS DE FUNDAMENTAÇÃO: ANÁLISE COMPARATIVA ENTRE AS BASES TEÓRICO- METODOLÓGICAS UTILIZADAS EM SEUS MODELOS DE REPRESENTAÇÃO DE DOMÍNIOS

A dissertação apresentada por Medeiros (2011) em Niterói para obtenção do título de Mestre em Ciência da Informação na Universidade Federal Fluminense teve como objetivo identificar de forma comparativa, os elementos que constituem os formalismos de representação de modelos conceituais para a construção de Tesouros Conceituais e Ontologias de Fundamentação. Com base no objetivo proposto o autor buscou metodologias dentro da área de Ciências Sociais Aplicadas, mais precisamente os métodos qualitativos e comparativos, resultando na verificação das qualidades de cada modelo conceitual assim como a comparação por meio de suas características levantadas, esta, dividida em elementos comparáveis e não comparáveis dentro de Ontologias de Fundamentação e Tesouros Conceituais.

Para efeito de validação dos modelos conceituais comparados o autor construiu um modelo conceitual baseado em Ontologias de Fundamentação (UFO-A), modelada por meio do software livre brModelo. A Ontologia de Fundamentação desenvolvida baseou-se um Tesouro Conceitual já modelado, o Tesouro de Cultura Material de Índios do Brasil.

Medeiros (2011) chegou na conclusão que as áreas de Ciência da Informação e Ciência da Computação (áreas de estudos) deveriam manter um maior relacionamento entre si, sendo esta uma forte razão para as diferenças entre ontologias e tesouros, ontologias são mais voltadas a área da Ciência da Computação, representando um artefato tecnológico sobre

um vocabulário, e tesouros voltados a área da Ciência da Informação, visando um controle terminológico de um vocabulário. O autor ainda comenta que ontologias conseguem representar um conhecimento como os tesouros, entretanto a Ciência da Computação negligencia a modelagem de domínios, gerando muitas ontologias e não utilizando-se de uma classificação consistente. Medeiros (2011) finaliza sua conclusão ressaltando que há mais diferenças do que semelhanças entre Ontologias de Fundamentação e Tesouros Conceituais e destaca que a Ciência da Informação deve se esforçar mais em busca de ferramentas que auxiliam na organização, recuperação e visualização de informações de um determinado domínio.

4.2 TESAUROS E ONTOLOGIAS SOB A LUZ DA TEORIA COMUNICATIVA DA TERMINOLOGIA

Na dissertação de Sales (2008) para obtenção do título de Mestre na Universidade Federal de Santa Catarina, Florianópolis, o objetivo foi identificar por meio de uma pesquisa documental os aspectos que aproximam e distanciam os fundamentos do tesouro e o da ontologia. A metodologia aplicada na pesquisa consiste no Método de Análise de Conteúdo de Laurence Bardin e por meio da Teoria Comunicativa da Terminologia de Maria Teresa Cabré, plano metodológico e teórico respectivamente. Para a constituição do corpus de análise foram utilizadas as bases de dados da *Library and Information Science Abstracts*, da *Wilson Library Literature and Information Science Full Text* e da Biblioteca Digital de Teses e Dissertações do IBICT.

Segundo Sales (2008) a análise dos resultados e interpretações proporcionaram significativas contribuições, haja vista que várias semelhanças e diferenças foram encontradas no que se diz respeito a tesouros e ontologias no campo da representação de conhecimento. Constatando-se que ambos os instrumentos se aproximam no que tange o esclarecimento dos aspectos relativos aos termos e suas estruturas conceituais. Suas diferenças se acentuam na esfera das aplicações, pois os recursos informáticos que suportam as ontologias concedem objetivos que vão além daqueles atribuídos aos tesouros.

4.3 REPRESENTAÇÃO DE INFORMAÇÃO E CONHECIMENTO: ESTUDO DAS DIFERENTES ABORDAGENS ENTRE A CIÊNCIA DA INFORMAÇÃO E A CIÊNCIA DA COMPUTAÇÃO

Furgeri (2006) realiza uma pesquisa para obtenção do título de Mestre em Ciência da Informação na Pontifícia Universidade Católica de Campinas, Campinas, onde o foco foi a representação do conhecimento e da informação, investigando quais são os pontos convergentes e divergentes entre as linguagens documentárias da Ciência da Informação e as linguagens de marcação desenvolvidas e utilizadas na Ciência da Computação, visando identificar ações, teorias e processos necessários para uma maior integração entre as duas áreas.

Segundo o autor, a pesquisa baseou-se em identificar os elementos fundamentais e necessários para a representação da informação e do conhecimento no âmbito da Ciência da Informação, por meio de metadados, tesauros e ontologias e no âmbito da Ciência da Computação se buscou formas e linguagens de representação de conhecimento que ajudam na construção da Web Semântica. Furgeri (2006) comenta que a validação da pesquisa culminou na construção de uma ontologia para a representação de recursos informacionais utilizando o *software* Protégé.

Como resultados o autor constatou que as formas de representação de conhecimento oriundas da Ciência da Informação e da Ciência da Computação buscam representar o conhecimento de maneiras similares, estabelecendo conceitos (gerais e específicos) e seus relacionamentos, entretanto o público-alvo nas duas áreas são distintos, enquanto na área da Ciência da Informação com os tesauros, por exemplo, busca-se auxiliar a representação e recuperação de informação para seres humanos, a área da Ciência da Computação com as ontologias, por exemplo, buscam auxiliar agentes de *software* objetivando numa interpretação de informações de forma automática. Furgeri (2006) ainda conclui que a ontologia proposta na pesquisa pode auxiliar não apenas seres humanos a recuperar informações de maneira mais efetiva, mas também ser um ponto de partida para que mecanismos de software realizem inferências sobre artigos na Web.

4.4 TESAUROS E A WORLD WIDE WEB

O trabalho de conclusão de curso de Murakami (2005) na Universidade de São Paulo, São Paulo, para obtenção do título de Bacharel em Biblioteconomia Documentação, objetivou-se em identificar como os tesauros estão sendo utilizados e/ou incorporados à nova dinâmica de gestão de informações na Web.

A metodologia aplicada para solucionar a pesquisa foi a utilização de uma pesquisa exploratória nas áreas da Ciência da Informação e Ciência da Computação em relação aos termos de tesauros e organização e representação do conhecimento, culminando na separação sistemática dos resultados em dois grupos:

- a) representação tecnológica dos tesauros documentários na Web;
- b) emprego dos tesauros documentários em várias etapas da gestão da informação em ambientes informacionais da Web.

Murakami (2005) conclui que ferramentas de vocabulário controlado, como os tesauros, podem contribuir com a organização de ambientes informacionais, mas ainda possuem muitas barreiras tecnológicas, principalmente no que tange a linguagem de representação e gerenciamento de conteúdo do tesauros, destacando também algumas tecnologias e ferramentas que estão sendo construídas para sanar estas limitações. Outro ponto que o autor destaca é o nível de automatização de tarefas que um tesouro pode apresentar na Web e indica algumas funções que o mesmo pode exercer:

- a) ferramenta de linguagem para recuperação da informação ;
- b) mediação para o aprendizado do usuário durante a navegação;
- c) padronização dos sistemas de comunicação disponíveis na Web.

4.5 DETECÇÃO E CORREÇÃO DE PROBLEMAS DE QUALIDADE DOS DADOS:

MODELO, SINTAXE E SEMÂNTICA

Oliveira (2008) apresenta em sua tese de doutorado, *Detecção e Correção De Problemas de Qualidade dos Dados: Modelo, Sintaxe e Semântica*, na Universidade do Minho, Braga, Portugal, uma exploração dos problemas e limitações encontrados nas atuais soluções de melhoria na qualidade de dados e a concepção de um novo modelo para detecção e correção de problemas na qualidade dos dados. A metodologia aplicada na pesquisa se

baseou em conceituar as áreas da pesquisa, realizar um estudo de caso e criar um modelo de detecção e correção, detalhados a seguir:

- a) definição de qualidade dos dados;
- b) definição de limpeza de dados;
- c) definição de problemas oriundos da qualidade dos dados;
- d) formalização das operações de detecção de problemas;
- e) formalização das operações de correção de problemas;
- f) proposta de um modelo para limpeza de dados;
- g) realização de um estudo de caso para limpeza de dados;
- h) interoperabilidade das operações de limpeza de dados.

Segundo o autor os estudos metodológicos resultaram no desenvolvimento de um protótipo para detecção e correção de dados, denominado SmartClean, que aplicado ao estudo de caso alcançou seus objetivos, confirmando sua aplicabilidade, validade e utilidade na detecção e correção de dados em problemas de qualidade de dados. Oliveira (2008) ainda destaca o estudo realizado na pesquisa sobre interoperabilidade das operações de detecção e correção em bases de dados diferentes, alcançado por meio da especificação das operações em um nível conceitual que as isola do nível do esquema.

4.6 CARACTERÍSTICAS DA QUALIDADE DE DADOS RELACIONADOS PUBLICADOS EM FONTES DE DADOS

Na tese de doutorado, *Qualitätsmerkmale Von Linked Data-Veröffentlichenden Datenquellen*, realizada por Flemming (2011) na Universidade Humboldt, Berlin, Alemanha, o objetivo foi elaborar um conjunto de critérios que representam a qualidade de uma fonte de dados estruturada e vinculada na Web.

Segundo a autora foram pesquisados critérios, considerando as melhores práticas atuais e as necessidades dos consumidores, sejam eles humanos ou máquinas. Cada critério foi examinado e determinado a sua relevância para a qualidade da fonte de dados. A partir de 30 critérios houveram exclusões e combinações, resultando em 11 critérios qualitativos agrupados em 4 categorias, denominadas conteúdo, representação, uso e sistema.

Flemming (2011, tradução nossa) comenta que além dos critérios serem divididos em categorias eles também receberam indicadores que constituem aspectos mensuráveis de

um critério e, assim, permitir a avaliação final da qualidade de uma fonte de dados de forma quantitativa por meio de cálculos que representam a pontuação de cada indicador.

Para efeito de validação a autora destaca que foi utilizado o método de apresentação dos indicadores, realizada em listas de discussões, a desenvolvedores e usuários em geral, demonstrando a pontuação de uma fonte de dados a partir de exemplos, haja vista que seria complicado utilizar e testar os critérios em uma fonte de dados estruturados na Web de terceiros. Como resultado Flemming (2011, tradução nossa) vê com sucesso o aceite dos critérios durante as apresentações e ressalva alguns critérios que não estavam presentes na pesquisa, como segurança, reputação e relevância da fonte de dados, que ou são incomuns ou subjetivos quanto a avaliação da qualidade.

5 VALIDAÇÃO COMPARATIVA DE QUALIDADE SEMÂNTICA NA REPRESENTAÇÃO DE CONHECIMENTO EM PÁGINAS DA WEB: UM ESTUDO DE CASO

Neste capítulo é apresentada a metodologia de desenvolvimento do presente trabalho de conclusão de curso e a aplicação prática dos conceitos e tecnologias abordados na fundamentação teórica. O objetivo do desenvolvimento prático foi validar comparativamente ontologias de fundamentação e tesouros conceituais, focando nas características de qualidade semântica que reduzem a ambiguidade nas referidas formas de representação de conhecimento.

5.1 METODOLOGIA

Dentro das linhas de pesquisa do curso de Ciência da Computação, este projeto se insere na área de Web Semântica e Sistemas de Informação.

As etapas de desenvolvimento deste trabalho foram: levantamento bibliográfico, estudo de caso e critérios de qualidade semântica para avaliação de uso de recursos em ontologias de fundamentação e tesouros conceituais.

5.1.1 Levantamento bibliográfico

A pesquisa bibliográfica deste trabalho foi extensa, haja vista que a Web Semântica é uma área de conhecimento em expansão e tem-se produzido muitos trabalhos e pesquisas com este tema. Porém, conforme já mencionado por Flemming (2011, tradução nossa), estas pesquisas e trabalhos focam tecnologias e linguagens e muito pouco na qualidade semântica que as mesmas apresentam. Os objetos de estudo (Web Semântica, Representação de Conhecimento, Ambiguidade, Qualidade Semântica, Ontologias de Fundamentação e Tesouros Conceituais) foram pesquisados a partir das abordagens realizadas em teses de doutorado, dissertações de mestrado, trabalhos de conclusão de curso, artigos científicos, livros e sites oficiais das tecnologias, em especial o site do W3C.

A partir do levantamento bibliográfico foi possível atender os seguintes objetivos específicos:

- a) contextualização da Web Semântica e sua contribuição na organização e representação do conhecimento em páginas da Web a partir da definição de conceitos, compressão de sua arquitetura, identificação de ferramentas para sua elaboração e apresentação de exemplos de sua aplicação – Capítulo 2;
- b) identificação de diferentes formas de representação de conhecimento em páginas da Web, em especial as ontologias de fundamentação e os tesouros conceituais – Capítulo 3;

5.1.2 Estudo de caso

O estudo de caso deste trabalho realizou uma validação comparativa nas formas de representação de conhecimento compreendidas por ontologias de fundamentação e tesouros conceituais, focando-se na validação de suas características de qualidade semântica que reduzem a ambiguidade na representação do conhecimento.

Para realização da validação foi necessário delimitar um domínio de conhecimento específico que atendesse completamente o estudo de caso, sendo que uma das principais dificuldades dentro do levantamento bibliográfico foi encontrar uma ontologia de fundamentação e/ou tesouro conceitual estruturado e populado que apresentasse as características de qualidade semântica em seus conceitos, assim como explicitasse as ambiguidades que são resolvidas pelas mesmas.

A partir da definição de um domínio de conhecimento, foram levantados os tipos de ambiguidade semântica a serem testados, assim como a identificação das características de qualidade semântica que tratam as ocorrências de ambiguidade em ontologias de fundamentação e tesouros conceituais para a realização de uma validação comparativa entre as duas formas de representação do conhecimento. No resultado da validação comparativa são discutidos os tipos de ambiguidade semântica e a solução que as formas de representação do conhecimento, ontologias de fundamentação e tesouros conceituais, propõem para resolvê-los a partir das características de qualidade semântica apresentada em seus conceitos.

Sendo assim, as etapas do estudo de caso consistiram em: delimitação do domínio de conhecimento, levantamento dos tipos de ambiguidade semântica a serem testados, identificação das características de qualidade semântica que tratam as ocorrências de ambiguidade em ontologias de fundamentação e tesouros conceituais, validação comparativa e

discussão dos resultados da validação.

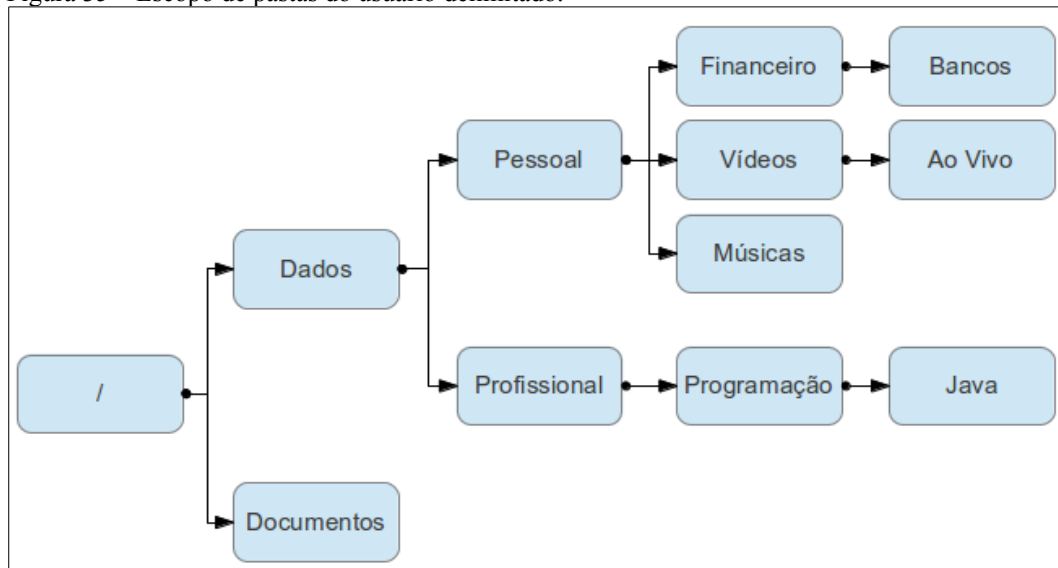
5.1.2.1 Domínio de conhecimento

Usuários de computadores sempre utilizaram pastas em seus computadores para guardar seus documentos, seja uma hierarquia simples ou complexa, e estes detêm o conhecimento sobre a estrutura da mesmas. Entretanto com o advento da Web Semântica e com o próprio conceito de Computação em Nuvem os dados estão sendo colocados na nuvem, ou seja, na Web. Uma vez que estes dados estão na Web é preciso que haja uma representação formal, principalmente se os mesmos forem compartilhados, não só entre usuários diferentes, mas principalmente entre aplicações diferentes. Baseado nisso o grupo KDE, criou um sub-projeto chamado Nepomuk, uma espécie de indexador de uma área de trabalho KDE do usuário perante seus dados, que visa fazer com que o computador funcione mais como a mente humana no que se refere a organização do conhecimento (KDE, 2012, tradução nossa). Entre os objetivos principais do Nepomuk, destacam-se:

- a) armazenar todas as informações de uma forma consistente e compatível permitindo que diferentes aplicativos possam acessá-lo e entendê-los;
- b) organizar essas informações em grafos contendo relações entre as entidades que fazem sentido.

O projeto Nepomuk reúne vários vocabulários para ontologias de nível superior que podem ser utilizados para criação de novas ontologias e/ou estruturas de representação de conhecimento. Entretanto essas ontologias são complexas quanto a sua estrutura, inviabilizando a utilização das mesmas neste estudo de caso, com isso foi necessário delimitar uma destas ontologias e um escopo dentro desta, neste caso, a ontologia delimitada foi a *Nepomuk File Ontology* (NFO) e escopo sendo como as pastas dos usuários, responsáveis por armazenar os documentos que ficam nos computadores dos usuários. A figura 35 demonstra o escopo de pastas de usuários delimitado.

Figura 35 – Escopo de pastas do usuário delimitado.



Fonte: Do autor.

5.1.2.2 Tipos de ambiguidade semântica

Para o estudo de caso os tipos de ambiguidade foram delimitados e agrupados perante o local onde os mesmos se apresentam, como identificados a seguir:

- a) conceitos: homografia, polissemia e vaguidade;
- b) relações conceituais: ambivalência e indeterminação.

Os conceitos e as relações conceituais foram escolhidos por possuírem uma maior gama de possibilidade de ocorrência de ambiguidade na representação do conhecimento, assim como a delimitação dos tipos de ambiguidade foi definida perante o grau de problemática semântica que os mesmos apresentam.

5.1.2.3 Características de qualidade semântica

Ontologias de fundamentação e tesouros conceituais apresentam diversas características e recursos para a representação do conhecimento, entre estas características, algumas representam qualidades semânticas que ajudam a reduzir a ambiguidade nos conceitos e relações conceituais contidos na representação do conhecimento, sendo que neste estudo de caso serão testadas e validadas as seguintes características:

- a) ontologias de fundamentação: categorias, rótulos, tipos e sub-tipos, relação disjuntiva e completa;

b) tesouros conceituais: agregação de conceitos, categorização, definição de conceitos, rótulos, relação conceitual quantitativa e qualitativa.

5.1.2.4 Validação comparativa

A validação comparativa apresentada no quadro 8 é realizada por meio das ambiguidades semânticas delimitadas para este estudo de caso e a solução utilizada para a redução de ambiguidade na representação de conhecimento proporcionada por Ontologias de Fundamentação e Tesouros Conceituais.

As possíveis soluções, previamente delimitadas, foram validadas por meio da estruturação de informações baseado em um escopo de pastas de usuários, sendo que tanto para ontologias de fundamentação quanto para tesouros conceituais foi utilizado o mesmo domínio de conhecimento e os mesmos dados, possibilitando assim a representação de conhecimento por ambas as formas representacionais.

Para construção da estrutura informacional e o próprio código fonte, foi utilizada as linguagens OWL para ontologias de fundamentação e SKOS para tesouros conceituais, o ambiente de desenvolvimento foi o gedit, versão 3.4.1 para Linux. Não foi utilizado a ferramenta Protégé, pois a mesma necessita que seja criada uma estrutura informacional completa, sendo que neste estudo de caso e para esta validação comparativa não se tinha esta necessidade, visto que a estrutura e os dados foram validados de forma isolada no escopo de pastas do usuário previamente definido.

Quadro 8 – Solução para ambiguidade em ontologias de fundamentação e tesouros conceituais.

Ambiguidade	Solução			
	Ontologias de Fundamentação		Tesouros Conceituais	
	Característica	Código fonte	Característica	Código fonte
Homografia	Categorias	Apêndice A	Categorização, Rótulo alternativo	Apêndice B
Polissemia	Tipos e sub-tipos	Apêndice A	Agregação de conceitos	Apêndice B
Vaguidade	Rótulos	Apêndice A	Definição de conceitos, Rótulo preferencial	Apêndice B
Ambivalência	Relação disjuntiva	Apêndice A	Relação quantitativa	Apêndice B
Indeterminação	Relação completa	Apêndice A	Relação qualitativa	Apêndice B

Fonte: Do Autor.

5.1.2.5 Resultados da validação comparativa

A definição do domínio de conhecimento inicialmente se tornou um problema, visto que existem poucas ontologias de fundamentação e/ou tesouros conceituais desenvolvidos, e este número diminui ainda mais quando necessita-se que estes estejam populados, sendo que a validação comparativa deste trabalho baseia-se principalmente na estruturação de informações que objetivam uma representação do conhecimento, ou seja, necessidade de estudar a estrutura e os dados em conjunto, de forma isolada e específica. Para tanto, foi utilizado um domínio de conhecimento limitado a uma estrutura de pastas de um usuário de computador comum, estruturado e populado sem necessidade de especialistas, se tornando conciso para a validação comparativa, pois os dados puderam ser delimitados de acordo com o escopo do estudo de caso estabelecido, objetivando uma comparação livre de preferencialismo, ou seja, possibilitando assim que as referidas formas de representação fossem capazes de solucionar os mesmos problemas nos mesmos dados.

Para cada tipo de ambiguidade encontrada na representação de conhecimento do domínio estabelecido foi identificado uma ou mais características de qualidade semântica dentro de ontologias de fundamentação e tesouros conceituais que reduzem e/ou solucionam estas ambiguidades, conforme detalhado a seguir.

O tipo de ambiguidade por Homografia, onde uma forma gramatical semelhante possui duas ou mais significações, foi encontrada dentro do domínio de conhecimento, na pasta conceituada por “Bancos”, sendo que a mesma poderia ser uma pasta contendo informações de instituições financeiras ou de arquivos de banco de dados. Ontologias de fundamentação apresentaram como solução a utilização da característica categoria, separando assim o respectivo conceito em uma categoria geral, que neste caso foi proporcionada pela categoria “Financeiro”, que também é uma pasta. Tesouros conceituais apresentou duas soluções para esta ambiguidade, categorização e rótulo alternativo, a categorização foi possível da mesma forma que em ontologias de fundamentação, ou seja, utilizando-se da categoria definida por “Financeiro”, o rótulo alternativo possibilitou que além do conceito “Bancos”, fosse adicionado um novo conceito para a pasta, que neste caso foi conceituado em “Instituições Financeiras”.

A ambiguidade por Polissemia, em que uma mesma forma gramatical possui diversas significações, encontra-se dentro do domínio, na pasta conceituada em “Java”, onde

esta deixa ambíguo se é uma linguagem de programação ou o nome de uma ilha. Ontologias de fundamentação tiveram como solução a utilização de tipos e sub-tipos, onde foi possível especificar qual é o tipo do conceito, ou seja, um conceito pai, que neste caso foi definido por “Linguagem de Programação”. Tesouros conceituais apresentaram como solução a agregação de conceitos, ou seja, um outro conceito dentro do domínio de conhecimento é agregado ao conceito ambíguo para delimitá-lo, neste caso o conceito agregado foi o de “Programação”.

O tipo de ambiguidade por Vaguidade, onde há uma falta de informação ao conceito, está contido na pasta conceituada em “Financeiro”, em que não é possível definir do que se trata o mesmo. Ontologias de fundamentação solucionam esta ambiguidade por meio de rótulos, que substituem o conceito na representação, neste caso, o conceito muda para “Contas a Pagar/Receber”. Tesouros conceituais utilizam-se de duas características que resolvem este tipo de ambiguidade, rótulo preferido e definição de conceitos, no rótulo preferido é adicionado ao conceito um rótulo preferencial que revela com mais exatidão o conceito, assim sendo, foi utilizado o rótulo preferencial “Contas a Pagar/Receber”, na definição do conceito, onde o mesmo pode ser posicionado dentro de um sistema de conceitos, foi utilizado a definição “Financeiro contém dados sobre finanças pessoais respectivos a contas a pagar e a receber”.

A ambiguidade por Ambivalência, em que há presença de sentidos contrários para o mesmo conceito perante um usuário e seu contexto, é verificada na pasta “Ao Vivo”, que para usuários diferentes podem se tratar de músicas, shows, vídeos, entre outros. Ontologias de fundamentação dispõem da característica denominada por “Relação disjuntiva” para resolver esta ambiguidade, eliminando relações conceituais indevidas, que neste caso, seria uma relação com “Músicas”, já que no domínio estabelecido a pasta “Ao Vivo” se refere a “Vídeos”. Tesouros conceituais basicamente fazem a mesma coisa que ontologias de fundamentação, mas de maneira inversa, ou seja, por meio de uma relação quantitativa é delimitado uma relação entre conceitos que são específicos a outro, baseado nisso, é realizado a relação da pasta “Ao Vivo” com a pasta “Vídeos”.

O tipo de ambiguidade por Indeterminação, onde há ausência de determinação ou presença de relativismo, é encontrado na pasta conceituada por “Pessoal”, referindo-se dentro do domínio de conhecimento a documentos pessoais do usuário, que não fica determinado se é isso, haja vista que existe uma pasta chamada “Documentos”. Ontologias de fundamentação realizam uma relação completa para resolver isso, fazendo que com que “Pessoal” e

“Documentos” sejam pastas equivalentes dentro de um relacionamento de conceitos. Tesouros conceituais resolvem este tipo de ambiguidade por meio da característica denominada por relação qualitativa, relacionando de forma associativa a pasta conceituada em “Pessoal” com a pasta conceituada em “Documentos”.

Ontologias de fundamentação e tesouros conceituais por meio de suas características de qualidade semântica conseguiram reduzir e/ou solucionar as ambiguidades encontradas dentro da estruturação de informações com foco na representação do conhecimento, sendo que em alguns casos, ambas as formas de representação proporcionaram que uma mesma característica resolvesse mais de uma ambiguidade. Entretanto, em dois dos cinco tipos de ambiguidade delimitados tesouros conceituais dispuseram mais de uma característica para solucionar a mesma ambiguidade, isso se deve principalmente as características de rótulos preferenciais e alternativos e das definições de conceitos, que não são características estruturais e sim conceituais, diferentemente das características de qualidade obtidas em ontologias de fundamentação, onde são todas características estruturais, importantes para a interpretação por máquinas, mas não tão alinhadas ao entendimento humano perante a uma conceituação representada em linguagem de máquina.

A utilização de ontologias de fundamentação ou tesouros conceituais deve ser estudada caso a caso, partindo-se principalmente da definição de quem terá maior contato com a representação do conhecimento, máquinas ou humanos. A formalização obtida com ontologias de fundamentação ajuda na interpretação dos conceitos por máquinas e as características conceituais de Tesouros Conceituais realizam um maior alinhamento entre o domínio de conhecimento do ser humano com a referida representação do conhecimento.

Na representação de conhecimento em páginas da Web necessita-se de que haja uma cooperação entre máquinas e humanos, as primeiras sendo possibilitadas de interpretarem e processarem o conhecimento, os últimos de compreenderem a informação recuperada e relacionada. Com isso, mesmo delimitando uma área para representação, neste caso, páginas da Web, ainda se tem outras variáveis envolvidas, como: o usuário precisará em algum momento interpretar uma informação, entender um relacionamento, se sim, tesouros conceituais são os mais indicados, caso contrário, ontologias de fundamentação resolvem melhor o problema, pois serão as máquinas que interpretarão e processarão toda a informação, assim como tomarão as decisões.

5.1.3 Critérios de qualidade semântica para uso em ontologias de fundamentação e tesouros conceituais

Ontologias de fundamentação e tesouros conceituais são evoluções de ontologias e tesouros respectivamente, que agregaram aos seus antecessores um alinhamento entre o modelo estrutural e o modelo conceitual, permitindo que a representação de conhecimento seja clara tanto para pessoas quanto para máquinas. Esta evolução como já mencionado, baseou-se principalmente na utilização de categorias, definições de conceitos e generalizações, que aumentam a qualidade na representação de conhecimento, principalmente no que tange a redução da ambiguidade. No quadro 9 é apresentado os tipos de ambiguidades e os locais onde as mesmas se apresentam em uma representação de conhecimento, assim como os critérios de qualidade semântica para uso em ontologias de fundamentação e tesouros conceituais que solucionam as respectivas ambiguidades. Estes critérios foram elencados a partir da realização do estudo de caso, onde por meio da validação comparativa pode ser validado quais critérios apresentam qualidade semântica para avaliação de uso de recursos em ontologias de fundamentação e tesouros conceituais, estes critérios se tornam soluções para a redução de ambiguidade na estruturação de informações com finalidade de representação de conhecimento em páginas da Web.

Quadro 9 – Critérios de qualidade semântica para uso em ontologias de fundamentação e tesouros conceituais.

Local	Ambiguidade	Critérios de qualidade semântica	
		Ontologias de Fundamentação	Tesouros Conceituais
Conceitos	Duplo sentido, Homonímia, Homografia, Policategorização, Polissemia, Vaguidade	Categorias, Fases, Funções, Rótulos, Tipos e Sub-tipos	Agregação de conceitos, Categorização, Definição de conceitos, Rótulos preferenciais e alternativos
Relações conceituais	Ambivalência, Anáfora, Indeterminação, Metáfora, Sinonímia	Relação disjuntiva e completa, Restrições, Superordenação e/ou Subordinação	Organização alfabética, Organização sistemática, Relação conceitual quantitativa e qualitativa, Verbalização.

Fonte: Do Autor.

6 CONCLUSÃO

O estudo da Web Semântica é algo relativamente novo, iniciando-se no final da década de 90 e atingindo um alto grau de aceitação cerca de 10 anos depois, estando em constante evolução, tendo como principal objetivo a significação de conteúdos perante máquinas e humanos de forma cooperativa. Muito antes disso, tanto a área da Ciência da Computação quanto a área da Ciência da Informação estudam formas de representação de conhecimento e recuperação de informações. Sendo que a Ciência da Informação é um pouco mais antiga, vindo antes do próprio computador, possuindo assim uma maior experiência no campo da significação da informação, auxiliando os seres humanos na organização do conhecimento, principalmente com a utilização de tesouros. A Ciência da Computação por sua vez investiu na questão de bancos de dados e IA e com o advento da Web Semântica introduziu os estudos com ontologias.

A constante comparação entre ontologias e tesouros dentro da área de Ciência da Computação e da Ciência da Informação objetivou o desenvolvimento deste trabalho, que iniciou-se com a conceitualização da Web Semântica, sua arquitetura, tecnologias, ferramentas e aplicações. Esta conceitualização possibilitou toda a fundamentação para mais uma comparação entre ontologias e tesouros, que neste caso tratou-se de uma comparação entre suas evoluções: ontologias de fundamentação e tesouros conceituais. Antes disso porém foi necessário validar se ontologias e tesouros realmente são melhores para a representação de conhecimento em páginas da Web, para tanto, a fundamentação provou ser verdadeiro por meio de uma comparação com as formas de representação de conhecimento denominadas por Mapas Conceituais e Metadados.

A comparação entre ontologias de fundamentação e tesouros conceituais teve como objetivo uma validação de qualidade semântica na estruturação de informações com finalidade de representação de conhecimento em páginas da Web, sendo o fator qualidade semântica as características encontradas nas referidas formas de representação de conhecimento que reduzem a ambiguidade semântica e/ou conceitual de um determinado domínio de conhecimento. A ambiguidade e a qualidade semântica então foram estudadas de maneira separada, objetivando uma fundamentação do que seria problema e o que seria solução. Este objetivo foi alcançado por meio do levantamento dos tipos de ambiguidade semântica, características de qualidade semântica em ontologias de fundamentação e tesouros

conceituais, assim como uma definição de um domínio de conhecimento, originando assim um estudo de caso com finalidade de uma validação comparativa entre os objetos de estudo.

O resultado da validação comparativa pode esclarecer os motivos que fazem com que ontologias e tesauros possuam várias pesquisas comparativas tanto na Ciência da Computação quanto na Ciência da Informação, haja vista que são formas de representação de conhecimento relativamente diferentes e que resolvem quase os mesmos problemas. Entretanto as soluções propostas por cada forma de representação são distintas, ontologias melhoram a interoperabilidade e interpretação por máquinas e tesauros melhoram a representação e recuperação da informação para humanos. No que se diz respeito a ontologias de fundamentação e tesauros conceituais quanto a qualidade semântica na estruturação de informações e representação de conhecimento em páginas da Web os resultados finais são semelhantes, sendo que em comparação com as suas respectivas origens, ontologias e tesauros, os mesmos apresentam maior qualidade semântica, porém ontologias de fundamentação se especializam em máquinas e tesauros conceituais em humanos.

Para trabalhos futuros sugere-se:

- a) desenvolver uma nova ferramenta ou um *plugin* para a ferramenta Protege, objetivando a construção de ontologias de fundamentação e/ou tesauros conceituais;
- b) desenvolver um comparador de qualidade semântica para ontologias de fundamentação e/ou tesauros conceituais;
- c) criar uma ontologia de fundamentação e/ou tesouro conceitual completo para estruturação de pastas e documentos de usuários de computadores;
- d) desenvolver um buscador semântico a partir da representação da estrutura de pastas e documentos de usuários de computadores.

REFERÊNCIAS

- ABITEBOUL, Serge et al. Gerenciamento de dados na Web. Rio de Janeiro, 2000. 251 p.
- ADIDA, Ben et al. RDFa 1.1 Primer. 2012. Disponível em: <<http://www.w3.org/TR/rdfa-primer/>>. Acesso em: 25 jun. 2012.
- ALVES, Rachel C. V. WEB SEMÂNTICA: uma análise focada no uso de metadados. 2005. 180 f. Dissertação, UNESP, Marília, 2006.
- APACHE. Apache Jena. 2012. Disponível em: <<http://jena.apache.org/>>. Acesso em: 3 jun. 2012.
- ASK. **About**. 2012. Disponível em: <<http://www.ask.com/>>. Acesso em: 13 jun. 2012.
- BECHHOFFER, Sean; MILES, Alistair. SKOS Simple Knowledge Organization System Reference. 2009. Disponível em: <<http://www.w3.org/TR/skos-reference/>>. Acesso em: 13 mai. 2012.
- BENEVIDES, Alessandro B. OntoUML. 2012. Disponível em: <<http://code.google.com/p/ontouml/>>. Acesso em: 25 jun. 2012.
- BERNERS-LEE, Tim. Semantic Web on XML. 2000. Disponível em: <<http://www.w3.org/2000/Talks/1206-xml2k-tbl/Overview.html>>. Acesso em: 7 abr. 2012.
- _____. Semantic Web Concepts. 2005. Disponível em: <<http://www.w3.org/2005/Talks/0517-boit-tbl>>. Acesso em: 6 abr. 2012.
- _____. Semantic Web and Linked Data. 2009. Disponível em: <<http://www.w3.org/2009/Talks/0120-campus-party-tbl/>>. Acesso em: 29 abr. 2012.
- _____. Linked Data. 2009. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 25 jun. 2012.
- BERNERS-LEE, Tim et al. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American, New York, p.29-37, 17 maio 2001.
- BIZER, Chris et al. WIKI. 2006. Disponível em: <<http://www4.wiwiss.fu-berlin.de/bizer/wiqa/>>. Acesso em: 25 jun. 2012.
- BOLEY, Harold; KIFER, Michael. RIF Overview. 2005. Disponível em: <<http://www.w3.org/2005/rules/wiki/Overview>>. Acesso em: 27 mai. 2012.
- BOLEY, Harold et al. RIF Core Dialect. 2010. Disponível em: <<http://www.w3.org/TR/rif-core/>>. Acesso em: 27 mai. 2012.

BORGO, S.; MASOLO, C. Foundational choices in DOLCE. 2009. Disponível em: <<http://www.loa-cnr.it/Papers/IOShandDOLCEv16.pdf>>. Acesso em: 21 jun. 2012.

BRANDT, Mariana Baptista. Etiquetagem e Folksonomia: uma análise sob a óptica dos processos de organização e recuperação da informação na web. 2009. Dissertação (Pós-Graduação) - Departamento de Ciência da Informação e Documentação. Universidade de Brasília, Brasília, 2000.

BRANSKI, Regina M. Recuperação de informações na Web. *Perspectivas em Ciência da Informação*, Belo Horizonte, p.70-87, 02 jan. 2004.

BRAY, Tim et al. Namespaces in XML 1.0 (Third Edition). 2009. Disponível em: <<http://www.w3.org/TR/REC-xml-names/>>. Acesso em 7 de abr. 2012.

BRÄSCHER, Marisa. A Ambiguidade na Recuperação da Informação. *Revista de Ciência da Informação*, Rio de Janeiro, v. 3, n. 1, fevereiro de 2002. Disponível em <http://www.dgz.org.br/fev02/Art_05.htm>. Acesso em: 18 abr. 2012.

BRÄSCHER, Marisa.; SCHIESSI, Marcelo. Ontologia: ambiguidade e precisão. *Revista eletrônica de biblioteconomia e ciência da informação*, Florianópolis, v. 17, n. 1, p. 125-141, 2012. Disponível em <<http://www.periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17nesp1p125/22729>>. Acesso em: 29 set. 2012.

BREITMAM, Karin K. *Web Semântica: a internet do futuro*. Rio de Janeiro, 2005. 190 p.

BRICKLEY, Dan; GUHA, R.V. RDF vocabulary description language 1.0. 2004. Disponível em: <<http://www.w3.org/TR/rdf-schema/>>. Acesso em: 7 abr. 2012.

_____. Resource Description Framework (RDF). 2004. Disponível em: <<http://www.w3.org/RDF/>>. Acesso em 7 de abr. 2012.

BRICKLEY, Dan; MILES, Alistair. SKOS Core Guide. 2005. Disponível em: <<http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>>. Acesso em: 19 mai. 2012.

CAMPOS, Maria L. A. *Linguagem documentária: teorias que fundamentam sua elaboração*. Niterói: UFF, 2001. 133 p.

CAMPOS, Maria L. A. et al. Estudo comparativo de softwares de construção de tesouros. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 11, n. 1, p. 68-81, jan./abr. de 2006. Disponível em <www.scielo.br/pdf/pci/v11n1/v11n1a06.pdf>. Acesso em: 29 set. 2012.

CAMPOS, Maria L. A.; GOMES, Hagar E. Metodologia de Elaboração de Tesouro Conceitual: A Categorização como Princípio Norteador. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 11, n. 3, p. 346-359, setembro de 2006. Disponível em <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/273/66>>. Acesso em: 22 jun. 2012.

CARLAN, Eliana. Ontologia e Web Semântica. 2006. 60 f. Monografia (Mestrado) – Universidade de Brasília, UnB, Brasília, 2006.

_____. Sistemas de Organização do Conhecimento: uma reflexão no contexto da Ciência da Informação. 2010. 195 f. Dissertação – Programa de Pós Graduação em Ciência da Informação, UB, Brasília, 2010.

CAROLO, Fernando. Gestão de Conteúdo Web Com Ferramentas Semânticas. 2011. Disponível em: <<http://globo.com>>. Acesso em: 25 jun. 2012.

CARVALHO, Cedric L.; OLIVEIRA, Hellen C. Gestão e Representação do Conhecimento. 2008. 19 f. Relatório técnico, UFG, Goiânia, 2008. Disponível em: <www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_003-08.pdf>. Acesso em: 14 jun. 2012.

CNFCP. Tesouro de Folclore e Cultura Popular Brasileira. 2012. Disponível em: <<http://www.cnfcp.gov.br/tesouro/>>. Acesso em: 17 jun. 2012.

COATES, Tony et al. URIs, URLs, and URNs: Clarifications and Recommendations 1.0. 2001. Disponível em: <<http://www.w3.org/TR/uri-clarification/>>. Acesso em 7 de abr. 2012.

CONNOLLY, Dan. Gleaning Resource Descriptions from Dialects of Languages (GRDDL). 2007. Disponível em: <<http://www.w3.org/TR/grddl/>>. Acesso em: 25 jun. 2012.

CURRÁS, Emilia. Tesouros: linguagens terminológicas. Brasília : IBICT, 1995. 286 p.

DBPEDIA. About. 2012. Disponível em: <<http://dbpedia.org/About>>. Acesso em: 13 jun. 2012.

DEAN, Mike; SCHREIBER, Guus. OWL Web Ontology Language Reference. 2004. Disponível em: <<http://www.w3.org/TR/owl-ref/>>. Acesso em: 30 abr. 2012.

DEVEDZIC, V. Semantic Web and Education. Springer, 2006.

EVI. About. 2012. Disponível em: <<http://evi.com/>>. Acesso em: 13 jun. 2012.

FACEBOOK. Open Graph. 2012. Disponível em: <<https://developers.facebook.com/docs/opengraph/>>. Acesso em: 13 jun. 2012.

FAVIKI. About. 2012. Disponível em: <<http://www.faviki.com/pages/about/>>. Acesso em: 13 jun. 2012.

FERNEDA, Edberto. Recuperação de Informação: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 2003. Tese (Doutorado) - Escola de Comunicação e Artes da Universidade de São Paulo, São Paulo, 2003.

FERREIRA, EVELINE C. H. GOMES. Geração Automática de Metadados: uma

Contribuição para a Web Semântica . 2006. 229 f. Tese, USP, São Paulo, 2006.

FERREYRA, Diego. TemaTres. 2007. Disponível em: <<http://www.r020.com.ar/tematres/>>. Acesso em: 3 jun. 2012.

FLEMMING, Annika. Qualitätsmerkmale von Linked Data-veröffentlichenden Datenquellen. 2011. 173f. Tese (Doutorado em Ciência da Computação) – Universidade Humboldt, Berlin, Alemanha, 2011.

FOKOUÉ, Achille et al. OWL 2 Web Ontology Language Profiles. 2009. Disponível em: <<http://www.w3.org/TR/owl2-profiles/>>. Acesso em: 13 mai. 2012.

FOX, Mark S. TOVE. 2011. Disponível em: <<http://www.eil.utoronto.ca/enterprise-modelling/tove/index.html>>. Acesso em: 22 jun. 2012.

FREEBASE. Wiki. 2012. Disponível em: <http://wiki.freebase.com/wiki/Main_Page>. Acesso em: 13 jun. 2012.

FUCHS, C. Les ambiguïtés du français. Paris: Orphys, 1996. 183p.

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

FÜRBER, Christian. Semantic Web Quality. 2011. Disponível em: <<http://semwebquality.org/mediawiki/index.php?title=SemWebQuality.org>>. Acesso em: 20 jun. 2012.

GINSBERG, Allen et al. RIF Use Cases and Requirements. 2005. Disponível em: <<http://www.w3.org/2005/rules/wiki/UCR>>. Acesso em: 27 mai. 2012.

GOOGLE. Knowledge Graph. 2012. Disponível em: <<http://www.google.com/insidesearch/features/search/knowledge.html>>. Acesso em: 13 jun. 2012.

GOOGLE. Schema.org. <<http://support.google.com/webmasters/bin/answer.py?hl=pt-BR&answer=1211158&topic=1088472&ctx=topic>>. Acesso em: 19 ago. 2012.

GOOGLE CODE. About. 2012. Disponível em: <<http://code.google.com/p/google-refine/>>. Acesso em: 13 jun. 2012.

GOLBREICH, Christine et al. OWL 2 Web Ontology Language New Features and Rationale. 2009. Disponível em: <<http://www.w3.org/TR/owl2-new-features/>>. Acesso em: 13 mai. 2012.

GONZÁLEZ, José A. M. Evolução ontológica das linguagens documentárias: relato de uma experiência de curso organizado conjuntamente para o DT/SIBI-USP e o PPGCI/ECA.

Revista de Ciência da Informação e Documentação, Ribeirão Preto, n. 1, p.143-164, janeiro 2011.

GUARINO, Nicola. Formal Ontology and Information Systems. Trento, 1998.

GUIZZARDI, G. Ontological foundations for structural conceptual models. 416 f. Tese (PhD em Computer Science) – Twente University of Technology, Twente, Holanda, 2005.

GRUBER, Thomas R. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, v. 5, n. 2, p. 199-220, abr. 1993.

HAKIA. About. 2012. Disponível em: <<http://hakia.com/>>. Acesso em: 13 jun. 2012.

HARRIS, Steve; SEABORNE, Andy. SPARQL 1.1 Query Language. 2012. Disponível em: <<http://www.w3.org/TR/sparql11-query/>>. Acesso em: 20 mai. 2012.

HITZLER, Pascal et al. OWL 2 Web Ontology Language Primer. 2009. Disponível em: <<http://www.w3.org/TR/owl2-primer/>>. Acesso em: 13 mai. 2012.

HODGE, G. Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. Abril 2000. Disponível em: <<http://www.clir.org/pubs/reports/pub91/1knowledge.html/contents.html> >. Acesso em: 29 set. 2012.

IANNELLA, Renato; WAUGH, Andrew. Metadata: Enabling the Internet. Distributed Systems Technology Centre. 1997. Disponível em: <<http://ifla.queenslibrary.org/documents/libraries/cataloging/metadata/ianr1.pdf>>. Acesso em: 18 jun. 2012.

ISAAC, Antoine; SUMMERS, Ed. SKOS Simple Knowledge Organization System Primer. 2009. Disponível em: <<http://www.w3.org/TR/skos-primer/>>. Acesso em: 13 mai. 2012.

ISAAC, Antonie; TRONCY, Raphael. DOE. Disponível em: <<http://www.eurecom.fr/~troncy/DOE/>> Acesso em 9 set. 2012.

ISO 2788. Guidelines for the establishment and development of monolingual thesauri. 1986. Disponível em: <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?cnumber=7776>. Acesso em: 17 jun. 2012.

ISO 8000-102. Master data: Exchange of characteristic data: Vocabulary. 2009. Disponível em: <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?cnumber=50799>. Acesso em: 19 jun. 2012.

ISOTANI, Seiji et al. Web 3.0: Os Rumos da Web Semântica e da Web 2.0 nos Ambientes Educacionais. In: XIX SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 2008, Fortaleza. Anais do Simpósio Brasileiro de Informática na Educação. Itajaí: Sbie, 2008. p. 785 – 795.

KDE. Nepomuk. 2012. Disponível em: <<http://nepomuk.kde.org/>>. Acesso em: 10 nov. 2012.

LABORATORY FOR APPLIED ONTOLOGY. DOLCE. 2006. Disponível em: <<http://www.loa.istc.cnr.it/DOLCE.html>>. Acesso em: 21 jun. 2012.

LIMA, Gercina Ângela B. O. Mapa Hipertextual (MHTX): um modelo para organização hipertextual de documentos. 2004. 199 f. Tese, UFMG, Belo Horizonte, 2004.

LUFT, Celso P. Minidicionário Luft. 20ª ed. São Paulo: Ática, 2001.

MACHION, Andréia C. G. Uso de ontologias e mapas conceituais na descoberta e análise de objetos de aprendizagem: um estudo de caso em eletrostática. 2007. 120 f. Tese, USP, São Paulo, 2007.

MANIEZ, J. Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales. Madrid, Fund. Germán Sánchez Rupérez, 1993.

MANOLA, Frank; MILLER, Eric. RDF Primer. 2004. Disponível em: <<http://www.w3.org/TR/rdf-primer/>>. Acesso em 29 de abr. 2012.

MARTINS, Aline F. Construção de Ontologias de Tarefa e sua Reutilização na Engenharia de Requisitos. 160 f. Dissertação (Mestrado) - Universidade Federal do Espírito Santo, UFES, Vitória, 2009.

MCGUINNESS, Deborah L. et al. OWL Web Ontology Language Guide. 2004. Disponível em: <<http://www.w3.org/TR/owl-guide/>>. Acesso em: 30 abr. 2012.

MCGUINNESS, Deborah L.; HARMELEN, Frank V. OWL Web Ontology Language Overview. 2004. Disponível em: <<http://www.w3.org/TR/owl-features/>>. Acesso em: 30 abr. 2012.

MEALLING, M.; DENENBERG, R. RFC 3305: URIs, URLs, and URNs. 2002. Disponível em: <<http://www.rfc-editor.org/rfc/rfc3305.txt>>. Acesso em 30 de abr. 2012.

MEDEIROS, Jackson S.. Tesouros conceituais e ontologias de fundamentação: análise comparativa entre as bases teórico-metodológicas utilizadas em seus modelos de representação de domínios. 2011. 145 f. Monografia (Mestrado) – Programa de Pós Graduação em Ciência da Informação, UFF, Niterói, 2011.

MICROSOFT. COMMUNITY. 2012. Disponível em: <<http://www.bing.com/community/>>. Acesso em: 13 jun. 2012.

MOREIRA, Alexandra. Tesouros e Ontologias: estudo de definições presentes na literatura das áreas das Ciências da Computação e da Informação, utilizando-se o método analítico sintético. 2003. 150 f. Dissertação (Mestrado) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2003.

MULTISYSTEMS. Multites. Disponível em: <<http://www.multites.com>> Acesso em 6 set.

2012.

MURAKAMI, Tiago Rodrigo Marçal. Tesouros e a World Wide Web. 2005. 92 f. Monografia (Graduação) - Curso de Biblioteconomia e Documentação, USP, São Paulo, 2005.

OLIVEIRA, Paulo J. M. Detecção e Correção de Problemas de Qualidade dos Dados: Modelo, Sintaxe e Semântica . 2008. 383f. Tese (Doutorado em Engenharia da Computação) - Universidade do Minho, Braga, Portugal, 2008.

ONTO-MED. GFO. 2010. Disponível em: <<http://www.onto-med.de/ontologies/gfo/>>. Acesso em: 21 jun. 2012.

ONTOLOGY PORTAL. SUMO. 2012. Disponível em: <<http://www.ontologyportal.org/>>. Acesso em: 17 jun. 2012.

OPENLINK SOFTWARE. Virtuoso. 2012. Disponível em: <<http://virtuoso.openlinksw.com/>>. Acesso em: 11 jun. 2012.

PAGANI, Luiz Arthur. Avaliação epistemológica de um exemplo de análise de ambigüidade num manual de introdução à semântica. 2009. Disponível em <<http://people.ufpr.br/~arthur/textos/ambig.pdf>>. Acesso em: 25 out. 2012.

PIZZATO, L. A. S. Estrutura Multitesouro para Recuperação de Informações. 2003. Dissertação (Mestrado) - Faculdade de Informática, PUC, Porto Alegre. 2003.

PRUD'HOMMEAUX, Eric. W3C RDF Validation Service. 2007. Disponível em: <<http://www.w3.org/RDF/Validator/>>. Acesso em: 13 jun. 2012.

PRUD'HOMMEAUX, Eric; SEABORNE, Andy. SPARQL Query Language for RDF. 2008. Disponível em: <<http://www.w3.org/TR/rdf-sparql-query/>>. Acesso em: 20 mai. 2012.

QUIN, Liam. Namespaces in XML 1.0 (Third Edition). 2012. Disponível em: <<http://www.w3.org/XML/>>. Acesso em 7 de abr. 2012.

RAMALHO, Rogério A. S. Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação. 2006. 120 f. Dissertação (Mestrado) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2006.

RAMALHO, Rogério A. S. Desenvolvimento e utilização de ontologias em bibliotecas digitais: uma proposta de aplicação. 145 f. Tese (Doutorado em Ciência da Informação) - Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Marília, 2010.

RAMALHO, Rogério A. S. et al. Web semântica: uma investigação sob o olhar da Ciência da Informação. Revista de Ciência da Informação - v. 8, n. 6, dezembro de 2007. Disponível em <http://www.dgz.org.br/dez07/Art_04.htm>. Acesso em: 6 abr. 2012.

SALES, Rodrigo de. Tesouros e ontologias sob a luz da Teoria Comunicativa da

Terminologia. 2008. 164f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação. Universidade Federal de Santa Catarina, UFSC, Florianópolis, 2008.

SCHNEIDER, L. How to build a foundational ontology: the object-centered high-level reference ontology. 2003. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.1.3440>>. Acesso em: 21 jun. 2012.

SEMANTIC WEB. Virtuoso. 2012. Disponível em: <<http://semanticweb.org/wiki/Virtuoso>>. Acesso em: 11 jun. 2012.

_____. Tools. 2012. Disponível em: <<http://semanticweb.org/wiki/Tools>>. Acesso em: 12 jun. 2012.

_____. SMW+. 2011. Disponível em: <<http://semanticweb.org/wiki/SMW%2B>>. Acesso em: 13 jun. 2012.

_____. Dublin Core. 2006. Disponível em: <http://semanticweb.org/wiki/Dublin_Core>. Acesso em: 15 jun. 2012.

_____. FOAF. 2005. Disponível em: <<http://semanticweb.org/wiki/FOAF>>. Acesso em: 15 jun. 2012.

_____. Openlink Virtuoso. 2012. Disponível em: <http://www.w3.org/2001/sw/wiki/OpenLink_Virtuoso>. Acesso em: 11 jun. 2012.

SILVA, Lúcio B. Ambiguidades da língua portuguesa: recorte classificatório para a elaboração de um modelo ontológico. 2006. 135 f. Dissertação (Mestrado) - Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2006.

SOUZA, R. R.; ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. *Ciência da Informação*, Brasília, v. 33, n. 1, p. 132-141, jan./abr. 2004.

SPORNY, Manu. RDFa Lite 1.1. 2012. Disponível em: <<http://www.w3.org/TR/rdfa-lite/>>. Acesso em: 25 jun. 2012.

STANFORD CENTER FOR BIOMEDICAL INFORMATICS RESEARCH. The Protégé Ontology Editor and Knowledge Acquisition System. 2012. Disponível em: <<http://protege.stanford.edu/overview/po-screenshots.html>>. Acesso em: 3 jun. 2012.

SWOOGLE. **About**. 2012. Disponível em: <<http://swoogle.umbc.edu>>. Acesso em: 13 jun. 2012.

SYNAPTICA. Synaptica. Disponível em: <<http://www.synaptica.com>> Acesso em 30 set. 2012.

THE NEW YORK TIMES. Linked Open Data. 2009. Disponível em: <<http://data.nytimes.com/>>. Acesso em: 13 jun. 2012.

THE UNICODE CONSORTIUM. **UNICODE**. 2011. Disponível em: <<http://www.unicode.org/standard/principles.html>>. Acesso em: 19 de abr. 2012.

TOPQUADRANT. TopBraid Suite. 2012. Disponível em: <http://topquadrant.com/products/TB_Suite.html>. Acesso em: 11 jun. 2012.

_____. SPARQLMotion. 2012. Disponível em: <<http://topquadrant.com/products/SPARQLMotion.html>>. Acesso em: 11 jun. 2012.

_____. SPIN. 2012. Disponível em: <<http://topquadrant.com/products/SPIN.html>>. Acesso em: 11 jun. 2012.

_____. SPARQL Web Pages. 2012. Disponível em: <<http://topquadrant.com/swp/>>. Acesso em: 11 jun. 2012.

TRB. TRT. 2012. Disponível em: <<http://trt.trb.org/>>. Acesso em: 22 jun. 2012.

USCHOLD, M.; GRUNINGER, M. Ontologies: principles, methods and applications. *The Knowledge Engineering Review*, v. 11, n. 12, 1996.

VISUWORDS. About. 2012. Disponível em: <<http://www.visuwords.com/>>. Acesso em: 13 jun. 2012.

W3C. W3C Markup Validation Service. 2012. Disponível em: <<http://validator.w3.org/>>. Acesso em: 25 jun. 2012.

W3C BRASIL. Web Semântica. 2011. Disponível em: <<http://www.w3c.br/Padroes/WebSemantica>>. Acesso em: 11 jun. 2012.

W3C OWL WORKING GROUP. OWL 2 Web Ontology Language Document Overview. 2009. Disponível em: <<http://www.w3.org/TR/owl2-overview/>>. Acesso em: 13 mai. 2012.

WANG, R. Y.; STRONG, D. M. Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems* - v. 12, n. 4, março de 1996. Disponível em: <<http://dl.acm.org/citation.cfm?id=1189572>>. Acesso em: 20 jun. 2012.

WOLFRAM ALPHA. About. 2012. Disponível em: <<http://www.wolframalpha.com/>>. Acesso em: 13 jun. 2012.

WORDNET. About. 2012. Disponível em: <<http://wordnet.princeton.edu/wordnet/>>. Acesso em: 13 jun. 2012.

WORDWEB. About. 2012. Disponível em: <<http://www.wordwebonline.com/>>. Acesso em: 13 jun. 2012.

ZENG, M. L. Knowledge Organization Systems (KOS). *Knowledge Organization*, v. 35, n. 2-3, p. 160-182, 2008.

APÊNDICE A – Código fonte Ontologias de Fundamentação.

```

<?xml version="1.0"?>

<!DOCTYPE rdf:RDF [
  <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
  <!ENTITY folder "http://www.semanticweb.org/ontologies/2012/10/folder.owl#" >
]>

<rdf:RDF xmlns="http://www.semanticweb.org/ontologies/2012/10/folder.owl#"
  xml:base="http://www.semanticweb.org/ontologies/2012/10/folder.owl"
  xmlns:folder="http://www.semanticweb.org/ontologies/2012/10/folder.owl#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <owl:Ontology rdf:about="http://www.semanticweb.org/ontologies/2012/10/folder.owl"/>

  <!-- Tipos e sub-tipos -->
  <owl:Class rdf:about="&folder;Linguagem_de_Programação"/>

  <!-- Categorias -->
  <owl:Class rdf:about="&folder;Dados"/>
  <owl:Class rdf:about="&folder;Documentos"/>

  <owl:Class rdf:about="&folder;Pessoal">
    <rdfs:subClassOf rdf:resource="&folder;Dados"/>
    <!-- Solução para Indeterminação -->
    <owl:equivalentClass rdf:resource="&folder;Documentos"/>
  </owl:Class>

  <owl:Class rdf:about="&folder;Profissional">
    <rdfs:subClassOf rdf:resource="&folder;Dados"/>
  </owl:Class>

  <owl:Class rdf:about="&folder;Programação">
    <rdfs:subClassOf rdf:resource="&folder;Profissional"/>
  </owl:Class>

  <owl:Class rdf:about="&folder;Músicas">
    <rdfs:subClassOf rdf:resource="&folder;Pessoal"/>
  </owl:Class>

```

```

<owl:Class rdf:about="&folder;Vídeos">
  <rdfs:subClassOf rdf:resource="&folder;Pessoal"/>
</owl:Class>

<owl:Class rdf:about="&folder;Ao_Vivo">
  <rdfs:subClassOf rdf:resource="&folder;Vídeos"/>
  <!-- Solução para Ambivalência -->
  <owl:disjointWith rdf:resource="&folder;Músicas"/>
</owl:Class>

<owl:Class rdf:about="&folder;Financeiro">
  <!-- Solução para Vaguidade -->
  <rdfs:label>Contas a Pagar/Receber</rdfs:label>
  <rdfs:subClassOf rdf:resource="&folder;Pessoal"/>
</owl:Class>

<owl:Class rdf:about="&folder;Bancos">
  <!-- Solução para Homografia -->
  <rdfs:subClassOf rdf:resource="&folder;Financeiro"/>
</owl:Class>

<owl:Class rdf:about="&folder;Java">
  <rdfs:subClassOf rdf:resource="&folder;Programação"/>
  <!-- Solução para Polissemia -->
  <rdfs:subClassOf rdf:resource="&folder;Linguagem_de_Programação"/>
</owl:Class>

</rdf:RDF>

```

APÊNDICE B – Código fonte Tesouros Conceituais.

```

<?xml version="1.0"?>

<!DOCTYPE rdf:RDF [
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
  <!ENTITY skos "http://www.w3.org/2004/02/skos/core#" >
]>

<rdf:RDF
  xml:base="http://www.w3.org/2004/02/skos/core"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">

  <skos:Concept rdf:about="Documentos" />

  <skos:Concept rdf:about="Dados">
    <skos:hasTopConcept rdf:resource="Pessoal" />
    <skos:hasTopConcept rdf:resource="Profissional" />
  </skos:Concept>

  <skos:Concept rdf:about="Pessoal">
    <skos:topConceptOf rdf:resource="Dados"/>
    <skos:hasTopConcept rdf:resource="Financeiro" />
    <skos:hasTopConcept rdf:resource="Músicas" />
    <skos:hasTopConcept rdf:resource="Vídeos" />
    <!-- Solução para Indeterminação -->
    <skos:related rdf:resource="Documentos" />
  </skos:Concept>

  <skos:Concept rdf:about="Profissional">
    <skos:topConceptOf rdf:resource="Dados" />
    <skos:hasTopConcept rdf:resource="Programação" />
  </skos:Concept>

  <skos:Concept rdf:about="Músicas" />
    <skos:topConceptOf rdf:resource="Pessoal" />
  </skos:Concept>

  <skos:Concept rdf:about="Vídeos" />
    <skos:topConceptOf rdf:resource="Pessoal" />
  </skos:Concept>

  <skos:Concept rdf:about="Ao_Vivo" />
    <!-- Solução para Ambivalência -->
    <skos:narrower rdf:resource="Vídeos" />
    <skos:topConceptOf rdf:resource="Vídeos" />

```

```
</skos:Concept>
```

```
<skos:Concept rdf:about="Programação">
  <skos:topConceptOf rdf:resource="Profissional" />
  <skos:hasTopConcept rdf:resource="Java" />
</skos:Concept>
```

```
<skos:Concept rdf:about="Java">
  <!-- Solução para Polissemia -->
  <skos:topConceptOf rdf:resource="Programação" />
</skos:Concept>
```

```
<skos:Concept rdf:about="Bancos" />
  <!-- Solução para Homografia -->
  <skos:altLabel>Instituições financeiras</skos:altLabel>
  <skos:topConceptOf rdf:resource="Financeiro" />
</skos:Concept>
```

```
<skos:Concept rdf:about="Financeiro">
  <!-- Solução para Vaguidade -->
  <skos:prefLabel>Contas a Pagar/Receber</skos:prefLabel>
  <skos:definition>
    Financeiro contém dados sobre finanças pessoais respectivos a contas a pagar e a receber
  </skos:definition>
  <skos:topConceptOf rdf:resource="Pessoal" />
  <skos:hasTopConcept rdf:resource="Bancos" />
</skos:Concept>
```

```
</rdf:RDF>
```

APÊNDICE C – Artigo.

CRITÉRIOS DE QUALIDADE SEMÂNTICA PARA AVALIAÇÃO DE USO DE RECURSOS EM ONTOLOGIAS DE FUNDAMENTAÇÃO E TESAuros CONCEITUAIS

Joel da Rosa¹

¹Universidade do Extremo Sul Catarinense – (UNESC)
Caixa Postal 3167 – 88806-000 – Criciúma – SC – Brazil

webjoel@hotmail.com

***Abstract.** The Semantic Web is a source of research on technology and its possibilities, however, few refer to quality in representing knowledge in Web pages. Besides the significance of content to provide interpretation by machines and humans is necessary to have a particular quality, especially in relation to natural language and the ambiguity generated by it. Foundational ontologies and conceptual thesauri are forms of knowledge representation that have extra features for semantic quality, these, can be defined as quality criteria for structuring information with a focus on knowledge representation. The quality criteria semantics contribute to the use evaluation of foundational ontologies and / or conceptual thesauri..*

***Resumo.** A Web Semântica é fonte de pesquisas sobre suas tecnologias e possibilidades, entretanto, pouco se referem à qualidade na representação do conhecimento em páginas da Web. Além da significação de conteúdos para prover interpretação por máquinas e humanos é necessário que exista uma qualidade específica, principalmente no que tange a linguagem natural e a ambiguidade gerada pela mesma. Ontologias de fundamentação e tesauros conceituais são formas de representação de conhecimento que apresentam características extras para qualidade semântica, estas, podendo ser definidas como critérios de qualidade para estruturar informações com foco na representação de conhecimento. Os critérios de qualidade semântica contribuem para a avaliação de uso de ontologias de fundamentação e/ou tesauros conceituais.*

1. Introdução

Um dos desafios da Web Semântica é criar uma estrutura que consiga expressar o significado de seu conteúdo e ao mesmo tempo estabelecer regras para processar esse significado de forma a inferir novos dados e regras com uma qualidade semântica provida para máquinas e humanos, diminuindo a flexibilidade da linguagem natural assim como a ambiguidade gerada pela mesma. Algumas soluções têm sido desenvolvidas para representar conhecimento e anotar a informação contida na Web envolvendo linguagens documentárias e/ou modelos conceituais, utilizadas nos sistemas informacionais para indexação, armazenamento e recuperação da informação (GUIZZARDI, 2005, tradução nossa).

Dentre as formas de representação de conhecimento com maior reconhecimento, destacam-se: tesauros e ontologias. O uso de ontologias em conteúdo Web, permite expressar

regras e compartilhar representações conceituais (ISOTANI et al, 2008). Os tesouros têm aplicação na Web no contexto de estruturação e recuperação da informação (SALES, 2008).

Ontologias de fundamentação e tesouros conceituais, evoluções de ontologias e tesouros respectivamente, tem em sua natureza uma representação de conhecimento mais alinhada entre o modelo conceitual e o domínio de conhecimento empregado, apresentando características extras para a qualidade semântica na estruturação de informações com foco na representação de conhecimento em páginas da Web (MEDEIROS, 2011).

As características de qualidade semântica encontradas em ontologias de fundamentação e tesouros conceituais proveem uma base para o levantamento de critérios de qualidade semântica que podem ser utilizadas na avaliação de uso dos recursos nas respectivas formas de representação de conhecimento. O levantamento destes critérios é realizado a partir da estruturação de informações com foco na representação de conhecimento em páginas da Web. Cada característica de qualidade semântica objetiva resolver uma ou mais ambiguidades encontradas nas referidas formas de representação de conhecimento, assim, estas características resultam nos critérios de qualidade semântica.

A validação das características de qualidade semântica em ontologias de fundamentação e tesouros conceituais gerou o levantamento dos critérios de qualidade semântica, possibilitando uma base para a criação de estruturas representacionais de conhecimento com alinhamento entre o domínio de conhecimento e o modelo conceitual, de forma que seja possível ter qualidade em uma representação de conhecimento utilizada por máquinas e humanos ao mesmo tempo, facilitando a interpretação e cooperação por ambas as partes.

2. Web Semântica

A Web Semântica é o projeto que visa concretizar a Web 3.0, a “Web dos Dados Linkados”, a Web que está sucedendo a Web 2.0, a “Web dos Documentos” (W3C Brasil, 2011), que por sua vez sucedeu a Web 1.0, estas sendo as primeiras páginas da Web. Breitman (2005) comenta que os conteúdos semânticos das páginas sintáticas são codificados de maneira acessível apenas as pessoas e cita alguns problemas que motores de buscas encontram dentro da Web Sintática (Web 1.0 e Web 2.0):

- a) muitas páginas encontradas, porém pouco precisas;
- b) resultados são sensíveis ao termo procurado;
- c) páginas encontradas não se relacionam com outras páginas.

Segundo Berners-Lee et al (2001, tradução nossa) a Web Semântica é a extensão da Web Sintática, onde as páginas da Web são complementadas por metadados que organizam e concedem significado, ou seja, semântica as informações publicadas. Essas informações podem ser transformadas em uma rede de dados que potencialmente conecta conhecimento, este, podendo ser compartilhado e interpretado por máquinas e humanos (DEVEDZIC, 2006, tradução nossa).

Segundo Ramalho et al (2007), o funcionamento da Web Semântica depende de um conjunto de tecnologias que se relacionam entre si, porém de maneira escalonável, estruturadas de modo que uma ou mais tecnologias contemplem uma determinada camada. Entre as camadas da Web Semântica, a camada semântica é a responsável pela representação do conhecimento, onde necessita-se de um vocabulário que permita descrever os aspectos semânticos dos recursos, criando relações de conceitos entre os mesmos, possibilitando a geração de significado para cada recurso em uma página da Web (RAMALHO et al, 2007). Como tecnologias para representação de conhecimento que podem ser utilizadas nesta camada

destacam-se: *Ontology Language (OWL)* e *Simple Knowledge Organization System (SKOS)*. OWL adiciona um vocabulário completo para descrever classes e propriedades, assim como suas relações de disjunção, cardinalidade, igualdade, entre outras respectivos ao um domínio de conhecimento (MCGUINNESS; HARMELEN, 2004, tradução nossa) . SKOS é modelado em uma estrutura RDF que fornece uma linguagem leve e intuitiva, podendo ser utilizada por si só ou em conjunto com a OWL, apresentando o conceito como principal elemento do vocabulário (BECHHOFFER; MILES, 2009, tradução nossa).

3. Representação de conhecimento na Web

A representação de conhecimento na Web está embutida dentro do conceito de Sistemas de Organização do Conhecimento (SOC) tradução do original em inglês: *Knowledge Organization System (KOS)*, onde o conhecimento está associado a algum tipo de sistema, responsável pela organização e recuperação da informação de forma ágil e precisa (MEDEIROS, 2011).

Segundo Carlan (2010), KOS estruturam semanticamente os conceitos, contemplando termos, definições, propriedades e relacionamentos, facilitando assim a indexação da informação por parte dos usuários. Medeiros (2011) complementa que KOS possuem vertentes tanto no campo da Ciência da Computação quanto na Ciência da Informação.

3.1. Ontologias de fundamentação

Ontologia é um artefato tecnológico que descreve um modelo conceitual de um determinado domínio em uma linguagem lógica e formal, a partir da descrição dos aspectos semânticos de conteúdos informacionais, possibilitando a realização de inferências automáticas por máquinas (RAMALHO, 2006) .

Ontologias de fundamentação são uma evolução de ontologia formal, que no grau de formalismo está entre uma ontologia semi-formal e uma rigorosamente formal, assim como também está classificada como uma ontologia de nível superior. Ontologias de fundamentação identificam categorias gerais de certos aspectos da realidade que não são específicos a um campo científico, descrevendo conhecimento independentemente de linguagem, de um estado particular das coisas ou ainda do estado de agentes (GUIZZARDI, 2005, tradução nossa). Segundo Schneider (2003, tradução nossa) ontologias de fundamentação são teorias axiomáticas de nível superior independentes de domínio, como: atributo, objeto, evento, partição, dependência e relação espaço-temporal .

A evolução de ontologias para ontologias de fundamentação culminou na presença de novas características, gerando uma representação de conhecimento mais rica em comparação com as ontologias até então, principalmente no que tange a redução de ambiguidade, reutilização e definição de conceitos e a categorização dos mesmos.

3.2. Tesouros conceituais

O termo tesouros se origina do grego *thesaurus* e do latim tesouro, este, que por muito tempo foi utilizado para designar léxico, possuindo duas vertentes: a europeia e a norte americana. A vertente europeia se baseia em um vocabulário organizado com termos que se relacionam de acordo com o seu significado. Já a vertente norte americana se caracteriza pela ordem alfabética dos termos do vocabulário.

Segundo Campos e Gomes (2006), a origem dos tesouros conceituais se deu com a insuficiência da descrição de uma palavra ou termo em um vocabulário controlado para designar o conteúdo de uma informação, como por exemplo: quando o vocabulário é convertido para outras línguas, onde dependendo da cultura o mesmo termo passa a ter outro conceito.

Medeiros (2011) conceitua que tesouros conceituais são formados por uma parte alfabética perante os termos, assim como a especificação das relações entre eles, e a parte sistemática, onde os termos são definidos por um conteúdo conceitual, ganhando destaque em suas definições e recebendo uma categorização de conceitos

Segundo Campos (2001) as teorias do conceito e da classificação originaram uma nova tendência para tesouros, os tesouros conceituais, um tesouro baseado no estabelecimento do termo/conceito e das relações entre eles, assim como uma mudança de sentido, onde o conteúdo conceitual dos termos passa a ser determinante em relação a base linguística.

3.3. Ambiguidade

Segundo Luft (2001) ambiguidade é a duplicidade de significados, falta de clareza, indecisão, onde uma palavra tem dois ou mais sentidos. Fuchs (1996, tradução nossa) complementa que a ambiguidade faz com que uma palavra ou frase seja compreendida de diferentes maneiras por um receptor. O autor classifica a ambiguidade nos seguintes tipos: morfológica, lexical, sintática, predicativa, semântica e pragmática.

A ambiguidade semântica se encontra no relacionamento dos termos em uma frase, entretanto é válido ressaltar que a ambiguidade semântica também é gerada pelo próprio termo, causando uma multiplicidade de conceitos relacionados com a sua aplicabilidade, e até mesmo por polissemia (SILVA, 2006). Pagani (2009) reforça constatando que a ambiguidade semântica se deve ao fato de os pronomes poderem ter diversos antecedentes e enunciados, denominando-se como correferencialidade, ou ainda como uma indeterminação referencial dos pronomes.

Silva (2006) realiza uma classificação dos fenômenos linguísticos que causam a ambiguidade, inclusive e principalmente a ambiguidade semântica, dividindo as mesmas em: policategorização, ambivalência, metáfora, duplo sentido, indeterminação, vaguidade, polissemia, anáfora, sinonímia, homonímia, homografia. O autor ainda revela que a ambiguidade está contida tanto no emissor de uma mensagem (*a priori*) quanto no receptor da mesma (*a posteriori*), o primeiro gera a ambiguidade em um texto e/ou enunciado mal formulado, o segundo por conter uma diversidade de condições, como cultural, histórica e/ou de contexto, resultando em uma ambiguidade na interpretação, relativa a linguagem natural utilizada, sendo que mesmo a partir de um texto gerado de forma ambígua o receptor precisa ter informações necessárias para realizar a interpretação correta, sendo assim a ambiguidade *a priori* pode inicialmente não ser um problema, desde que se tenha opções de interpretação no momento da recepção da mensagem, ou seja, a ambiguidade é reconhecida e pode ser resolvida *a posteriori*.

3.4. Qualidade semântica

Wang e Strong (1996, tradução nossa) definem que um dado tem qualidade quando está adequado para a utilização por parte dos consumidores de dados. A ISO 8000-102 (2009, tradução nossa) define que a qualidade dos dados é o grau que preenche os seus requisitos. Com base nas duas definições pode-se concluir que um dado tem qualidade quando o mesmo

está formalmente definido para atender as expectativas do consumidor de dados .

Ainda segundo Wang e Strong (1996, tradução nossa) as características que fazem uma informação ter qualidade semântica estão categorizadas na representação dos dados, envolvendo as dimensões: interpretabilidade e compreensibilidade na significação dos dados, assim como consistência e concisão no seu formato, detalhadas nas seguintes características:

- a) completa: define completamente um conceito;
- b) sem ambiguidade: não há duas representações para o mesmo conceito;
- c) significativa: conceito está contido dentro de um vocabulário;
- d) correta: o conceito é recuperado de acordo com seu significado.

Flemming (2011, tradução nossa) comenta que uma representação dos dados que engloba formatos, vocabulários e identificadores bem definidos gera vantagens ao consumidor de dados. Destacando-se o uso de vocabulários que aumentam a performance do processamento da lógica por um programa e ajudam humanos na compreensão semântica (se os mesmos possuírem afinidade com o vocabulário). Um HTML bem escrito e com uso de metadados internacionalmente reconhecidos ajudam tanto máquinas quanto humanos na recuperação da informação. A referência de URIs ajuda no aumento da quantidade de conjunto de dados, enriquecendo o vocabulário, assim como a rotulação legível de classes, propriedades e entidades por meio de rótulos e comentários ajudam no significado do dado.

4. Validação comparativa

A validação comparativa teve como objetivo validar comparativamente ontologias de fundamentação e tesouros conceituais, focando nas características de qualidade semântica que reduzem a ambiguidade nas referidas formas de representação de conhecimento, gerando desse modo critérios de avaliação de uso de recursos em ontologias de fundamentação e tesouros conceituais.

Para realização da validação comparativa foi delimitado um domínio de conhecimento, os tipos de ambiguidade semânticas e as características de qualidade semântica encontradas em ontologias de fundamentação e tesouros conceituais, culminando na própria validação comparativa.

4.1. Domínio de conhecimento

Nepomuk é uma espécie de indexador de uma área de trabalho KDE do usuário perante seus dados, que visa fazer com que o computador funcione mais como a mente humana no que se refere a organização do conhecimento (KDE, 2012, tradução nossa). Entre os objetivos principais do Nepomuk são: armazenar todas as informações de uma forma consistente e compatível, permitindo que diferentes aplicativos possam acessá-lo e entendê-los e organizar essas informações em grafos contendo relações entre as entidades que fazem sentido.

O projeto Nepomuk reúne vários vocabulários para ontologias de nível superior que podem ser utilizados para criação de novas ontologias e/ou estruturas de representação de conhecimento, sendo delimitada para esta validação comparativa a ontologia *Nepomuk File Ontology* (NFO) e como escopo específico as pastas dos usuários.

4.2. Tipos de ambiguidade semântica

Os tipos de ambiguidade foram delimitados e agrupados perante o local onde os mesmos se apresentam, identificados nos conceitos por: homografia, polissemia e vaguidade e nas

relações conceituais por: ambivalência e indeterminação.

Os conceitos e as relações conceituais foram escolhidos por possuírem uma maior gama de possibilidade de ocorrência de ambiguidade na representação do conhecimento, assim como a delimitação dos tipos de ambiguidade foi definida perante o grau de problemática semântica que os mesmos apresentam.

4.3. Características de qualidade semântica

Ontologias de fundamentação e tesouros conceituais apresentam diversas características e recursos para a representação do conhecimento, entre estas características, algumas representam qualidades semânticas que ajudam a reduzir a ambiguidade nos conceitos e relações conceituais contidos na representação do conhecimento, sendo que nesta validação para ontologias de fundamentação foi delimitadas as características: categorias, rótulos, tipos e sub-tipos, relação disjuntiva e completa. Para tesouros conceituais foram delimitadas: agregação de conceitos, categorização, definição de conceitos, rótulos, relação conceitual quantitativa e qualitativa.

4.4. Validação e resultados

A validação comparativa compreendeu na aplicação das possíveis soluções, estas sendo as características de qualidade semântica, nos problemas relatados pelos tipos de ambiguidade semântica, sendo validadas por meio da estruturação de informações baseado em um escopo de pastas de usuários, sendo que tanto para ontologias de fundamentação quanto para tesouros conceituais foi utilizado o mesmo domínio de conhecimento e os mesmos dados, possibilitando assim a representação de conhecimento por ambas as formas representacionais.

Para construção da estrutura informacional e o próprio código fonte, foi utilizada as linguagens OWL para ontologias de fundamentação e SKOS para tesouros conceituais, o ambiente de desenvolvimento foi o gedit, versão 3.4.1 para Linux. O quadro 1 apresenta as ambiguidades como problema e as características de qualidade semântica em ontologias de fundamentação e tesouros conceituais como solução depois da implementação da estrutura informacional e conseqüentemente na validação do uso dos recursos.

Quadro 1 – Solução para ambiguidade em ontologias de fundamentação e tesouros conceituais.

Ambiguidade	Solução	
	Característica em Ontologias de Fundamentação	Característica em Tesouros Conceituais
Homografia	Categorias	Categorização, Rótulo alternativo
Polissemia	Tipos e sub-tipos	Agregação de conceitos
Vaguidade	Rótulos	Definição de conceitos, Rótulo preferencial
Ambivalência	Relação disjuntiva	Relação quantitativa
Indeterminação	Relação completa	Relação qualitativa

Fonte: Do Autor.

Como resultados foi possível verificar que para cada tipo de ambiguidade encontrada na representação de conhecimento do domínio estabelecido foi identificado uma ou mais características de qualidade semântica dentro de ontologias de fundamentação e tesouros conceituais que reduzem e/ou solucionam estas ambiguidades.

Ontologias de fundamentação e tesouros conceituais por meio de suas características de qualidade semântica conseguiram reduzir e/ou solucionar as ambiguidades encontradas dentro da estruturação de informações com foco na representação do conhecimento, sendo que em alguns casos, ambas as formas de representação proporcionaram que uma mesma característica resolvesse mais de uma ambiguidade. Entretanto, em dois dos cinco tipos de ambiguidade delimitados tesouros conceituais dispuseram mais de uma característica para solucionar a mesma ambiguidade, isso se deve principalmente as características de rótulos preferenciais e alternativos e das definições de conceitos, que não são características estruturais e sim conceituais, diferentemente das características de qualidade obtidas em ontologias de fundamentação, onde são todas características estruturais, importantes para a interpretação por máquinas, mas não tão alinhadas ao entendimento humano perante a uma conceituação representada em linguagem de máquina.

4.5. Critérios de qualidade semântica para uso em ontologias de fundamentação e tesouros conceituais

Ontologias de fundamentação e tesouros conceituais são evoluções de ontologias e tesouros respectivamente, que agregaram aos seus antecessores um alinhamento entre o modelo estrutural e o modelo conceitual, permitindo que a representação de conhecimento seja clara tanto para pessoas quanto para máquinas. Esta evolução como já mencionado, baseou-se principalmente na utilização de categorias, definições de conceitos e generalizações, que aumentam a qualidade na representação de conhecimento, principalmente no que tange a redução da ambiguidade.

No quadro 2 é apresentado os tipos de ambiguidades e os locais onde as mesmas se apresentam em uma representação de conhecimento, assim como os critérios de qualidade semântica para uso em ontologias de fundamentação e tesouros conceituais que solucionam as respectivas ambiguidades. Estes critérios foram elencados a partir da realização da validação comparativa, onde se definiu quais critérios apresentam qualidade semântica para avaliação de uso de recursos em ontologias de fundamentação e tesouros conceituais, estes critérios se tornam soluções para a redução de ambiguidade na estruturação de informações com finalidade de representação de conhecimento em páginas da Web.

Quadro 2 – Critérios de qualidade semântica para solucionar a ambiguidade.

Local	Ambiguidade	Critérios de qualidade semântica	
		Ontologias de Fundamentação	Tesouros Conceituais
Conceitos	Duplo sentido, Homonímia, Homografia, Policategorização, Polissemia, Vaguidade	Categorias, Fases, Funções, Rótulos, Tipos e Sub-tipos	Agregação de conceitos, Categorização, Definição de conceitos, Rótulos preferenciais e alternativos
Relações conceituais	Ambivalência, Anáfora, Indeterminação, Metáfora, Sinonímia	Relação disjuntiva e completa, Restrições, Superordenação e/ou Subordinação	Organização alfabética, Organização sistemática, Relação conceitual quantitativa e qualitativa, Verbalização.

Fonte: Do Autor.

5. Conclusão

A constante comparação entre ontologias e tesouros dentro da área de Ciência da Computação e da Ciência da Informação objetivou a esta nova comparação entre ontologias e tesouros, que neste caso tratou-se de uma comparação entre suas evoluções: ontologias de fundamentação e tesouros conceituais.

A comparação entre ontologias de fundamentação e tesouros conceituais teve como objetivo uma validação das características de qualidade semântica na estruturação de informações com finalidade de representação de conhecimento em páginas da Web, sendo o fator qualidade semântica as características encontradas nas referidas formas de representação de conhecimento que reduzem a ambiguidade semântica e/ou conceitual de um determinado domínio de conhecimento

O resultado da validação comparativa pode esclarecer os motivos que fazem com que ontologias e tesouros possuam várias pesquisas comparativas tanto na Ciência da Computação quanto na Ciência da Informação, haja vista que são formas de representação de conhecimento relativamente diferentes e que resolvem quase os mesmos problemas. Entretanto as soluções propostas por cada forma de representação são distintas, ontologias melhoram a interoperabilidade e interpretação por máquinas e tesouros melhoram a representação e recuperação da informação para humanos. No que se diz respeito a ontologias de fundamentação e tesouros conceituais quanto a qualidade semântica na estruturação de informações e representação de conhecimento em páginas da Web os resultados finais são semelhantes, sendo que em comparação com as suas respectivas origens, ontologias e tesouros, os mesmos apresentam maior qualidade semântica, porém ontologias de fundamentação se especializam em máquinas e tesouros conceituais em humanos. A definição dos critérios de qualidade semântica para avaliação de uso de recursos em ontologias de fundamentação se aplica justamente como base teórica e técnica para a escolha de uma das formas de representação discutidas, visto que os critérios identificam qual forma possui mais qualidade na solução de problemas relativos a ambiguidades geradas pela linguagem natural imposta por humanos e não compreendidas por máquinas.

Referências

BERNERS-LEE, Tim et al. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, New York, p.29-37, 17 maio 2001.

BECHHOFFER, Sean; MILES, Alistair. SKOS Simple Knowledge Organization System Reference. 2009. Disponível em: <<http://www.w3.org/TR/skos-reference/>>. Acesso em: 13 mai. 2012.

BREITMAM, Karin K. *Web Semântica: a internet do futuro*. Rio de Janeiro, 2005.

CAMPOS, Maria L. A. *Linguagem documentária: teorias que fundamentam sua elaboração*. Niterói: UFF, 2001. 133 p.

CAMPOS, Maria L. A.; GOMES, Hagar E. Metodologia de Elaboração de Tesouro Conceitual: A Categorização como Princípio Norteador. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 11, n. 3, p. 346-359, setembro de 2006. Disponível em

<<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/273/66>>. Acesso em: 22 jun. 2012.

CARLAN, Eliana. Sistemas de Organização do Conhecimento: uma reflexão no contexto da Ciência da Informação. 2010. 195 f. Dissertação – Programa de Pós Graduação em Ciência da Informação, UB, Brasília, 2010.

DEVEDZIC, V. Semantic Web and Education. Springer, 2006.

FLEMMING, Annika. Qualitätsmerkmale von Linked Data-veröffentlichenden Datenquellen. 2011. 173f. Tese (Doutorado em Ciência da Computação) – Universidade Humboldt, Berlin, Alemanha, 2011.

FUCHS, C. Les ambiguïtés du français. Paris: Orphys, 1996. 183p.

GUIZZARDI, G. Ontological foundations for structural conceptual models. 416 f. Tese (PhD em Computer Science) – Twente University of Technology, Twente, Holanda, 2005.

ISO 8000-102. Master data: Exchange of characteristic data: Vocabulary. 2009. Disponível em: <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=50799>. Acesso em: 19 jun. 2012.

ISOTANI, Seiji et al. Web 3.0: Os Rumos da Web Semântica e da Web 2.0 nos Ambientes Educacionais. In: XIX SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 2008, Fortaleza. Anais do Simpósio Brasileiro de Informática na Educação. Itajaí: Sbie, 2008. p. 785 – 795.

KDE. Nepomuk. 2012. Disponível em: <<http://nepomuk.kde.org/>>. Acesso em: 10 nov. 2012.

LUFT, Celso P. Minidicionário Luft. 20ª ed. São Paulo: Ática, 2001.

MCGUINNESS, Deborah L.; HARMELEN, Frank V. OWL Web Ontology Language Overview. 2004. Disponível em: <<http://www.w3.org/TR/owl-features/>>. Acesso em: 30 abr. 2012.

MEDEIROS, Jackson S. Tesouros conceituais e ontologias de fundamentação: análise comparativa entre as bases teórico-metodológicas utilizadas em seus modelos de representação de domínios. 2011. 145 f. Monografia (Mestrado) – Programa de Pós Graduação em Ciência da Informação, UFF, Niterói, 2011.

PAGANI, Luiz Arthur. Avaliação epistemológica de um exemplo de análise de ambigüidade num manual de introdução à semântica. 2009. Disponível em <<http://people.ufpr.br/~arthur/textos/ambig.pdf>>. Acesso em: 25 out. 2012.

RAMALHO, Rogério A. S. Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação. 2006. 120 f. Dissertação (Mestrado) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2006.

RAMALHO, Rogério A. S. et al. Web semântica: uma investigação sob o olhar da Ciência da Informação. Revista de Ciência da Informação - v. 8, n. 6, dezembro de 2007. Disponível em <http://www.dgz.org.br/dez07/Art_04.htm>. Acesso em: 6 abr. 2012.

SALES, Rodrigo de. Tesouros e ontologias sob a luz da Teoria Comunicativa da Terminologia. 2008. 164f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação. Universidade Federal de Santa Catarina, UFSC, Florianópolis, 2008.

SCHNEIDER, L. How to build a foundational ontology: the object-centered high-level reference ontology. 2003. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.1.3440>>. Acesso em: 21 jun. 2012.

SILVA, Lúcio B. Ambiguidades da língua portuguesa: recorte classificatório para a elaboração de um modelo ontológico. 2006. 135 f. Dissertação (Mestrado) - Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2006.

W3C BRASIL. Web Semântica. 2011. Disponível em: <<http://www.w3c.br/Padroes/WebSemantica>>. Acesso em: 11 jun. 2012.

WANG, R. Y.; STRONG, D. M. Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems - v. 12, n. 4, março de 1996. Disponível em: <<http://dl.acm.org/citation.cfm?id=1189572>>. Acesso em: 20 jun. 2012.