

**ADEMIR ROBERTO FREDDO**

**FOLKONCEPT: MÉTODO DE SUPORTE À MODELAGEM  
CONCEITUAL DE ONTOLOGIAS A PARTIR DA AQUISIÇÃO DE  
CONHECIMENTOS DE FOLKSONOMIAS**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, como requisito parcial para obtenção do título de “Doutor em Ciências” – Área de Concentração: Informática Industrial.

Orientador: Prof. Dr. Cesar Augusto Tacla.

CURITIBA

2010

F852 Freddo, Ademir Roberto

Folkconcept : método de suporte à modelagem conceitual de ontologias a partir da aquisição de conhecimentos de folksonomias / Ademir Roberto Freddo. — 2010.  
225 p. : il. ; 30 cm

Orientador: César Augusto Tacla

Tese (Doutorado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial. Área de concentração: Informática Industrial, Curitiba, 2010.

Bibliografia: f. 193-209

1. Ontologia. 2. Aquisição de conhecimentos (Sistemas especialistas). 3. Sistemas de recuperação da informação. 4. Web semântica. 5. Web 2.0 (Sistema de recuperação da informação). 6. Engenharia elétrica – Teses. I. Tacla, César Augusto, orient. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial. III. Título.

CDD (22. ed.) 621.3

---

Título da Tese N° 57:

**“Desenvolvimento de ontologias pelo aprendizado  
semiautomático a partir de folksonomias”**

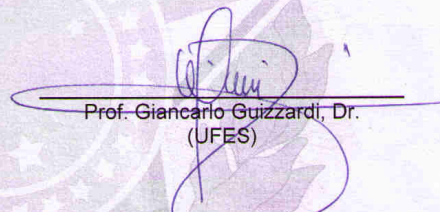
por

**Ademir Roberto Freddo**

Esta tese foi apresentada como requisito parcial à obtenção do título de DOUTOR EM CIÊNCIAS – Área de Concentração: Informática Industrial, pelo Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial – CPGEI – da Universidade Tecnológica Federal do Paraná – UTFPR – Campus Curitiba às 9h do dia 01 de outubro de 2010. O trabalho foi aprovado pela Banca Examinadora, composta pelos professores:



Prof. Cesar Augusto Tacla, Dr.  
(Presidente – UTFPR)



Prof. Giancarlo Guizzardi, Dr.  
(UFES)



Prof. Alessandro Lameira Koerich, Dr.  
(PUC-PR)

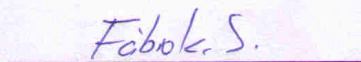


Prof.ª Andréia Malucelli, Dr.  
(PUC-PR)



Prof. Gustavo Alberto Giménez Lugo, Dr.  
(UTFPR)

Visto da coordenação:



Prof. Fábio Kurt Schneider, Dr.  
(Coordenador do CPGEI)



*Dedido esta tese a minha esposa Elizabeti  
Aparecida Gomes Pereira Freddo e ao meu filho  
Eduardo Izeppi Freddo.*



## AGRADECIMENTOS

Agradeço a DEUS por toda força e espiritualidade.

Agradeço a minha esposa Elizabeti que sempre me apoiou, pelo amor e dedicação. Agradeço pela força nos momentos mais difíceis, pela compreensão em momentos de ausência, pelos incentivos e pelas palavras de motivação.

Agradeço ao meu filho que ainda não sabe o que é um doutorado, mas que trouxe alegrias na reta final e, por dar um propósito à minha vida.

Agradeço ao meu orientador, Dr. Cesar A. Tacla que tornou possível a realização deste trabalho. Agradeço pela oportunidade, pelas sugestões, pela dedicação, pela motivação, pelos direcionamentos, pelos conselhos, sendo mais que um orientador, mas um verdadeiro amigo.

Agradeço a minha sogra Adélia, por sempre acompanhar minha esposa quando precisei me ausentar, pela maneira alegre de como encara a vida e supera as dificuldades que se apresentam.

Ao meu cunhado (Pico) pelos conselhos, palavras de apoio e pela força.

Aos meus pais que me incentivaram a cursar o doutorado.

A Professora Andreia Malucelli pelas sugestões e contribuições ao trabalho.

Aos colegas de trabalho da coordenação de informática da UTFPR Pato Branco por terem acumulado tarefas e por apoiarem minha liberação para cursar o doutorado.

Aos colegas do laboratório LSIP pelos momentos de descontração, pelas conversas e ajudas.

Ao Professor Gustavo Gimenez-Lugo pelas idéias e pela colaboração.

Aos componentes da banca, pelas contribuições e sugestões que possibilitaram a melhora do trabalho.

Aos estagiários Anderson e Lucas pelas implementações realizadas durante o doutorado.

A todas aquelas pessoas, que mesmo não sendo citadas contribuíram de alguma forma para elaboração deste trabalho.





## RESUMO

FREDDO, Ademir Roberto. **Folkconcept: Método de Suporte à Modelagem Conceitual de Ontologias a partir da Aquisição de Conhecimentos de Folksonomias**. 2009. 225f. Tese – Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2009.

Neste trabalho, apresenta-se um método para o desenvolvimento de ontologias a partir de folksonomias. O objetivo do método é auxiliar os atores envolvidos no processo de desenvolvimento na elicitación de termos a serem representados na ontologia e na tomada de decisão de como modelar tais termos. Busca-se, pela aplicação do método, reduzir o gargalo na aquisição de conhecimentos empregando-se técnicas de aprendizado de ontologias a partir de folksonomias. O método atinge três atividades do processo de desenvolvimento de ontologias: aquisição de conhecimentos, modelagem conceitual e avaliação das ontologias, sendo este último integrado à modelagem conceitual. Na aquisição de conhecimentos, o método trata da recuperação, representação e enriquecimento das etiquetas (termos) presentes nas folksonomias originadas de um processo social de etiquetagem realizado pelos atores envolvidos no desenvolvimento da ontologia. Na modelagem conceitual, auxilia o projetista a transformar as etiquetas das folksonomias em elementos da ontologia em desenvolvimento, ou seja, na modelagem de novos elementos. Na avaliação de ontologias, o método auxilia os projetistas na validação dos novos elementos que são sugeridos pelo método de aprendizado. Além disso, o método diminui a dificuldade em utilizar a metodologia *OntoClean* tornando sua aplicação transparente ao projetista. A avaliação do método foi realizada por meio de experimentos de desenvolvimento de ontologias em um ambiente controlado. Participaram dos experimentos equipes compostas por projetistas da área da computação, sendo que algumas equipes trabalharam com um protótipo que implementa o método e outras não. A avaliação foi realizada por meio de métricas que comprovaram que o sistema auxiliou os projetistas a desenvolverem ontologias mais descritivas e com número menor de erros de formalismo.

**Palavras-chave:** Ontologias. Desenvolvimento de Ontologias. Aprendizado de Ontologias. Folksonomias. Aquisição de Conhecimentos.



## ABSTRACT

FREDDO, Ademir Roberto. **Folkconcept: a Method for Supporting Conceptual Modeling of Ontologies starting with Knowledge Acquisition from Folksonomies**. 2010. 225p. Thesis – Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, Federal University of Technology of Paraná. Curitiba, 2010.

In this work, we present a method called Folkconcept for supporting conceptual modeling of ontologies starting with knowledge acquisition based on folksonomies. The method aims at helping actors enrolled in the development process in eliciting terms and in the modeling choice of how to represent these terms in the ontology. The objective of applying the Folkconcept method is to reduce the knowledge acquisition bottleneck through ontology learning techniques based on folksonomies. Folkconcept reaches three activities of the development process: knowledge acquisition, conceptual modeling, and evaluation, the latter being integrated into the conceptual modeling activity. With relation to the knowledge acquisition, Folkconcept deals with the retrieval, representation, and enrichment of terms (tags) coming from a folksonomy resulting from a social process of tagging performed by the actors involved in the ontology development process. In the conceptual modeling activity, Folkconcept helps the ontology designer to transform folksonomy's tags into elements of the ontology being developed. In the ontology evaluation activity, the method helps ontology designers to validate the new elements that are suggested by the ontology learning method. In addition, the Folkconcept reduces the difficulty in using the OntoClean methodology making its use transparent to the ontology designer. Folkconcept was evaluated by means of ontology development experiments realized in a controlled environment by teams composed by ontology designers coming from the area of computing. Some teams worked with a prototype system that implements the Folkconcept. Results obtained by these teams were compared with the results from teams working without the system. The comparison was performed through metrics that show that the Folkconcept helped ontology designers to develop more descriptive ontologies with fewer errors with relation to the idealized taxonomies of OntoClean.

**Keywords:** Conceptual Modeling. Ontology. Ontology Development. Ontology Learning. Folksonomy. Knowledge Acquisition.



## LISTA DE FIGURAS

Figura 1 – Relacionamento entre Ontologia, Linguagem e Conceitualização.....	39
Figura 2 - Dois sistemas A e B utilizando a mesma linguagem L com sobreposição dos modelos pretendidos $I_a(L)$ e $(I_b(L))$ .....	40
Figura 3 - Sobreposição de ontologias (em cinza claro) sem a sobreposição dos pretendidos	40
Figura 4 – Classificação de Ontologias.....	41
Figura 5 – Classificação de Ontologias.....	42
Figura 6 – Metodologias para o Desenvolvimento de Ontologias. ....	43
Figura 7 – Desenvolvimento e Ciclo de Vida de uma Ontologia na Metodologia METHONTOLOGY .....	45
Figura 8 – Artefatos produzidos pela METHONTOLOGY na Conceitualização .....	52
Figura 9 – Avaliação da Ontologia pela Análise do Conteúdo. ....	54
Figura 10 – Avaliação do conteúdo de ontologias baseada em evolução, regras e métricas. ..	55
Figura 11 – Avaliação de Ontologias pela Análise do Conteúdo Estrutural ou Funcional .....	56
Figura 12 – Formas de Avaliação de uma Ontologia .....	57
Figura 13 – Essência e Rigidez.....	61
Figura 14 – Exemplo de Hierarquia.....	62
Figura 15 – Taxonomia das Propriedades.....	67
Figura 16 – Taxonomia idealizada pela <i>OntoClean</i> .....	67
Figura 17 – Metodologia DILIGENT .....	69
Figura 18 – Metodologia NEON .....	70
Figura 19 – Engenharia Reversa no Aprendizado de Ontologias.....	81
Figura 20 – Triângulo da Significação.....	82
Figura 21 – Divisão das Abordagens de Aprendizado de Ontologias. ....	83
Figura 22 – Camadas/Etapas no Aprendizado de Ontologias .....	85
Figura 23 – Motivação no Processo de Anotação .....	90
Figura 24 – Grupos de Entidades Envolvidas na Anotação .....	91
Figura 25 – Exemplo de um Processo de Anotação Coletiva. ....	91
Figura 26 - Nuvem de <i>Tags</i> para Visualização das <i>Tags</i> Utilizadas no Processo.....	92
Figura 27 – Classificação dos Trabalhos de Aprendizado de Ontologias a partir de Folksonomias.....	97
Figura 28 – Ontologia SCOT .....	102
Figura 29 – Relação entre os <i>Synsets</i> das palavras <i>Java</i> , <i>Prog Language</i> e <i>Prolog</i> .....	104
Figura 30 – Exemplo de Relações na <i>WordNet</i> resultante da Pesquisa pela Palavra <i>Bike</i> . ...	107
Figura 31 – Algoritmo de <i>Levensthein</i> .....	109
Figura 32 – Exemplo do Cálculo da Distância de <i>Levensthein</i> .....	109
Figura 33 – Exemplo de uma Taxonomia. ....	110
Figura 34 – Exemplo de Conjuntos de Termos. ....	111
Figura 35 – Exemplos de Descrições Textuais após Lematização .....	112
Figura 36 – Exemplos de Taxonomias.....	114
Figura 37 – Hierarquia de Conceitos Formais (Atributos e Objetos) obtida pela FCA.....	116
Figura 38 – Fluxo do Método Desenvolvido para o Desenvolvimento de uma Ontologia. ..	120
Figura 39 – Atividades de Aquisição de Conhecimentos e Conceitualização .....	121
Figura 40 – Subatividades na Aquisição de Conhecimentos. ....	121
Figura 41 – Construção das Anotações na Aquisição de Conhecimentos. ....	124
Figura 42 – Algoritmo para a Construção das Anotações na Aquisição de Conhecimentos. ....	124



Figura 43 – Algoritmo para Recuperação de uma Anotação. ....	125
Figura 44 – Algoritmo para Enriquecimento da Anotação. ....	127
Figura 45 – Hierarquia obtida pela Correlação de <i>Tags</i> das Anotações. ....	129
Figura 46 – Ontologia para Representação das Anotações. ....	130
Figura 47 – Subatividades da Atividade de Conceitualização. ....	133
Figura 48 – Fluxo da Atividade de Conceitualização que inclui Alinhamento e Interação...	134
Figura 49 – Pseudo-código do Algoritmo para o Alinhamento de <i>Tags</i> . ....	136
Figura 50 – Refinamento do Algoritmo para o Alinhamento das <i>Tags</i> . ....	137
Figura 51 – Algoritmo para o Cálculo da $sim_2$ . ....	138
Figura 52 – Resultado da Subatividade de Alinhamento e Dados ....	140
Figura 53 – Fluxo do Questionário na Interação com o Projetista. ....	144
Figura 54 – Ontologia Raiz Inicial. ....	153
Figura 55 – Posicionamento da Avaliação dos Experimentos na Avaliação de Ontologias..	155
Figura 56 – Ontologia desenvolvida pelo Grupo 1 para exemplificar a Métrica <i>Attribute Richness</i> . ....	159
Figura 57 – Ontologia desenvolvida pelo Grupo 2 para exemplificar a Métrica <i>Attribute Richness</i> . ....	159
Figura 58 – Ontologia desenvolvida pelo Grupo 1 para exemplificar as Métricas <i>Relationship Diversity</i> e <i>Schema Deepness</i> . ....	160
Figura 59 – Ontologia desenvolvida Grupo 2 para exemplificar as Métricas <i>Relationship Diversity</i> e <i>Schema Deepness</i> . ....	161
Figura 60 – Ontologia desenvolvida pelo Grupo 1 para exemplificar as Métricas <i>Average Population</i> e <i>Class Utilization</i> . ....	162
Figura 61 – Ontologia desenvolvida pelo Grupo 2 para exemplificar as Métricas <i>Average Population</i> e <i>Class Utilization</i> . ....	163
Figura 62 – Exemplo de Relações Multihierárquica. ....	164
Figura 63 – Gráfico Comparativo da Avaliação Quantitativa do Primeiro Experimento. ....	165
Figura 64 – Gráfico do Desvio Padrão dos Elementos Criados pelos Grupos no Primeiro Experimento. ....	166
Figura 65 – Largura e Profundidade das Ontologias dos Grupos no Primeiro Experimento. ....	166
Figura 66 – Gráfico do Aproveitamento de <i>Tags</i> no Primeiro Experimento. ....	167
Figura 67 – Gráfico do Aproveitamento de <i>Tags</i> para o Grupo 1 no Primeiro Experimento. ....	168
Figura 68 – Gráfico Comparativo da Avaliação Primeiro Experimento utilizando-se das Métricas da <i>OntoQA</i> . ....	169
Figura 69 – Erros Taxonômicos no Desenvolvimento das Ontologias pelos Grupos no Primeiro Experimento. ....	169
Figura 70 – Gráfico Comparativo dos Elementos Criados no Segundo Experimento. ....	174
Figura 71 – Gráfico Comparativo da Avaliação do Segundo Experimento em relação as Métricas <i>OntoQA</i> . ....	174
Figura 72 – Gráfico do Aproveitamento de <i>Tags</i> no Segundo Experimento. ....	175
Figura 73 – Gráfico do Aproveitamento de <i>Tags</i> para o Grupo 1 no Segundo Experimento. ....	176
Figura 74 – Escopo do Trabalho Desenvolvido. ....	189
Figura 75 – Figura da Questão <i>Q2C</i> . ....	213
Figura 76 – Figura da Questão <i>Q3C</i> . ....	214
Figura 77 – Figura da Questão <i>Q5C</i> . ....	215
Figura 78 – Figura da Questão <i>Q2I</i> . ....	217
Figura 79 – Figura da Questão <i>Q3I</i> . ....	218
Figura 80 – Figura da Questão <i>Q4I</i> . ....	219
Figura 81 – Figura da Questão <i>Q5I</i> . ....	219





## LISTA DE TABELAS

Tabela 1 – Possíveis Classificações das Propriedades.....	66
Tabela 2 – Exemplo de uma Tabela de Contexto.....	115
Tabela 3 – Tabela de Contexto ou Correlação de <i>Tags</i> nas Anotações.....	128
Tabela 4 – Elementos criados nas Ontologias do Primeiro Experimento.....	156
Tabela 5 – Número de Conceitos Populados e Relações Taxonômicas das Ontologias do Primeiro Experimento.....	156
Tabela 6 – Desvio Padrão do Total de Tags e Elementos Criados pelos Atores no Primeiro Experimento.....	157
Tabela 7 – Relação de <i>Tags</i> , Elementos, <i>Tags</i> Ignoradas, <i>Tags</i> Aceitas e Índice Elementos/ <i>Tags</i> Aceitas no Primeiro Experimento.....	157
Tabela 8 – Relação de <i>Tags</i> Ignoradas, <i>Tags</i> Aceitas e Sugestões Aceitas no Grupo 1 do Primeiro Experimento.....	158
Tabela 9 – Avaliação das Ontologias do Primeiro Experimento utilizando-se das Métricas <i>OntoQA</i> no Primeiro Experimento.....	158
Tabela 10 – Largura e Profundidade das Ontologias Desenvolvidas pelos Grupos no Primeiro Experimento.....	163
Tabela 11 – Erros de Modelagem relacionados a Taxonomia Idealizada <i>OntoClean</i> no Primeiro Experimento.....	164
Tabela 12 – Elementos criados nas Ontologias desenvolvidas no Segundo Experimento.....	170
Tabela 13 – Número de Conceitos Populados e Relações Taxonômicas do Segundo Experimento.....	171
Tabela 14 – Relação de <i>Tags</i> , Elementos, <i>Tags</i> Ignoradas, <i>Tags</i> Aceitas e Índice de Elementos Criados por <i>Tags</i> Aceitas no Segundo Experimento.....	171
Tabela 15 – Relação de <i>Tags</i> Aceitas, <i>Tags</i> Ignoradas e Sugestões Aceitas no Grupo 1 do Segundo Experimento.....	172
Tabela 16 – Avaliação das Ontologias do Segundo Experimento com relação as Métricas da <i>OntoQA</i> .....	172
Tabela 17 – Questionário sobre a criação de um Novo Elemento no Segundo Experimento grupo 1.....	173



## LISTA DE QUADROS

Quadro 1 - Artefatos produzidos pelas Metodologias na Aquisição de Conhecimento. ....	49
Quadro 2 - Técnicas de Aquisição de Conhecimentos das Metodologias. ....	49
Quadro 3 - Artefatos produzidos pelas Metodologias de Desenvolvimento de Ontologias na Conceitualização. ....	52
Quadro 4 - Técnicas de Conceitualização das Metodologias de Desenvolvimento de Ontologias. ....	53
Quadro 5- Quadro Comparativo das Metodologias com base nas Atividades de Gerenciamento da METHONTOLOGY. ....	76
Quadro 6 - Quadro Comparativo das Metodologias com base nas Atividades de Desenvolvimento da METHONTOLOGY. ....	77
Quadro 7 - Quadro Comparativo das Metodologias com base nas Atividades de Suporte da METHONTOLOGY. ....	79
Quadro 8 - Descrição dos Conceitos da Ontologia para Representação das Anotações. ....	130
Quadro 9 - Descrição das Propriedades da Ontologia para Representação das Anotações. ..	131
Quadro 10 - Resumo das Questões do Questionário Eletrônico. ....	144
Quadro 11 – Comparação do Métodos em relação as Metodologias nas Técnicas utilizadas e Artefatos produzidos. ....	184
Quadro 12 – Resumo das Contribuições de Folkconcept em relação aos Métodos de Aprendizado a partir de Textos. ....	187
Quadro 13 – Comparação em relação ao desenvolvimento de ontologias a partir de folkonomias. ....	187
Quadro 14 - Combinação das Classificações <i>OntoClean</i> com Justificativas e Sugestões. ....	221



## LISTA DE ABREVIATURAS E SIGLAS

ABox	<i>Assertion Box</i>
API	<i>Aplication Programming Interface</i> (Interface de Programação de Aplicativos)
CPGEI	Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial
CYC	<i>enCYClopaedia</i>
DL	<i>Description Logic</i> (Lógica de Descrições)
DOLCE	<i>Descriptive Ontology for Linguistic and Cognitive Engineering</i>
FCA	<i>Formal Concept Analysis</i> (Análise Formal de Conceitos)
FOL	<i>First Order Logic</i> (Lógica de Primeira Ordem)
GUM	<i>Generalized Upper Model</i>
IC	<i>Identity Condition</i>
KEA	<i>Keyphrase Extraction Algorithm</i>
NLP	<i>Natural Language Processing</i> (Processamento de Linguagem Natural)
ODE	<i>Ontology Development Environment</i>
OntoQA	<i>Metric-Based Ontology Quality Analisys</i>
OTA	<i>Open Travel Alliance</i>
POAM	<i>Partial Ontology Alignment Method</i> (Método de Alinhamento Parcial de Ontologias)
RDF	<i>Resource Description Framework</i>
SC	<i>Semantic Cotopy</i>
SCOT	<i>Social Semantic Cloud of Tags</i>
SUMO	<i>Suggested Upper Merged Ontology</i>
TBox	<i>Terminological Box</i>
TF/IDF	<i>Term Frequency/ Inverse Document Frequency</i>
TOVE	<i>TOronto Virtual Enteprise</i>
UFO	<i>Unified Foundational Ontology</i>
UML	<i>Unified Modelling Language</i>
UPON	<i>Unified Process for Ontology</i>
URL	<i>Uniform Resource Locator</i>
WSD	<i>Word Sense Desambiguation</i>



# SUMÁRIO

<b>1.</b>	<b>INTRODUÇÃO .....</b>	<b>25</b>
1.1	PROBLEMAS E JUSTIFICATIVAS.....	25
1.2	HIPÓTESES.....	31
1.3	OBJETIVOS.....	32
1.4	CONTRIBUIÇÕES .....	32
1.5	ESTRUTURA DO TRABALHO .....	36
<b>2.</b>	<b>FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>37</b>
2.1	ONTOLOGIAS .....	37
2.1.1	Definições.....	37
2.1.2	Classificação das Ontologias.....	40
2.2	DESENVOLVIMENTO DE ONTOLOGIAS .....	42
2.2.1	METHONTOLOGY .....	44
2.2.2	Aquisição de Conhecimentos .....	47
2.2.3	Conceitualização .....	50
2.2.4	Avaliação de Ontologias .....	53
2.2.5	Metodologias para Desenvolvimento Colaborativo.....	68
2.2.6	Desenvolvimento de Ontologias pelo Reuso.....	71
2.2.7	Comparação das Metodologias .....	73
2.2.8	Aprendizado de Ontologias .....	80
2.3	ANOTAÇÃO .....	88
2.4	FOLKSONOMIAS E PERSONOMIAS.....	92
2.4.1	Folksonomias e Taxonomias .....	95
2.4.2	Folksonomias e Ontologias .....	95
2.4.3	Aprendizado de Ontologias a partir de Folksonomias .....	97
2.4.4	Ontologias para Representação das Folksonomias.....	101
2.5	PROCESSAMENTO DE LINGUAGEM NATURAL .....	102
2.5.1	Ontologias Linguísticas.....	105
2.6	MEDIDAS DE SIMILARIDADE.....	107
2.6.1	Análise Formal de Conceitos .....	114
2.6.2	Cálculo da Dissimilaridade .....	116
<b>3.</b>	<b>MÉTODO PROPOSTO: FOLKONCEPT .....</b>	<b>119</b>
3.1	AQUISIÇÃO DE CONHECIMENTOS .....	121
3.1.1	Anotação de Páginas Web.....	122
3.1.2	Construção das Anotações .....	124
3.2	CONCEITUALIZAÇÃO .....	132
3.2.1	Alinhamento das <i>Tags</i> .....	136
3.2.2	Interação com o Projetista.....	142
3.3	PRIMEIRA VERSÃO DA ONTOLOGIA RAIZ .....	147
<b>4.</b>	<b>EXPERIMENTOS.....</b>	<b>149</b>
4.1	METODOLOGIA DOS EXPERIMENTOS.....	149
4.1.1	Primeiro Experimento .....	150
4.1.2	Segundo Experimento .....	150
4.1.3	Aquisição de Conhecimentos nos Experimentos .....	151
4.1.4	Conceitualização nos Experimentos.....	151
4.1.5	Desenvolvimento da Ontologia Raiz.....	151
4.1.6	Avaliação dos Experimentos .....	154
4.2	RESULTADOS DO EXPERIMENTO INDIVIDUAL.....	155
4.2.1	Síntese do Primeiro Experimento.....	165

4.3	RESULTADOS DO SEGUNDO EXPERIMENTO	170
4.3.1	Síntese do Segundo Experimento.....	173
4.4	DISCUSSÕES DOS EXPERIMENTOS .....	176
<b>5.</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS.....</b>	<b>181</b>
5.1	TRABALHOS FUTUROS .....	189
	<b>REFERÊNCIAS.....</b>	<b>193</b>
	<b>APÊNDICE A .....</b>	<b>211</b>
	<b>APÊNDICE B.....</b>	<b>221</b>



# 1. INTRODUÇÃO

## 1.1 PROBLEMAS E JUSTIFICATIVAS

Segundo Guizzardi (2005) e Guarino (2008), na inteligência artificial, ontologia se refere a um artefato de engenharia, compreendendo um vocabulário de termos estruturados em uma taxonomia, suas definições e um conjunto de axiomas formais utilizados para criar novas relações e para restringir as interpretações dos termos. No sentido filosófico, de acordo com Guarino (2008), ontologia é um sistema de categorias que representa uma certa visão do mundo designada conceitualização. Uma conceitualização não depende de uma linguagem em particular e, desta forma, pode ser expressa por linguagens e/ou termos diferentes.

Guizzardi (2005, p. 2) adota o nome **conceitualização** para designar o conjunto de conceitos utilizados para articular abstrações do estado das coisas num domínio. **Modelo** é uma abstração de uma porção da realidade articulada segundo uma conceitualização de um domínio. Por exemplo (Guizzardi, 2005, p. 26), no mundo real há relações genealógicas. Uma conceitualização deste domínio pode ser construída pelos conceitos Pessoa, Homem, Mulher, Pai, Mãe, Filho, Irmão, entre outros. Com estes conceitos, pode-se construir vários modelos que representam fatos da realidade, tal como, *o homem chamado João é pai da mulher chamada Carla*. Ainda, para Guizzardi (2005), tanto conceitualizações como modelos existem somente nas mentes das pessoas. O que há de concreto são **especificações de modelos conceituais** feitas em uma **linguagem de modelagem** que permite expressar (representar) conceitualizações.

Um dos primeiros esforços da comunidade de desenvolvimento de ontologias, de acordo com Gómez-Pérez, Corcho e Fernández-López (2004), foi o de definir as atividades que compõem o processo de desenvolvimento. Fernández-López, Gómez-Pérez e Juristo (1997) definiram um conjunto de atividades para o processo de desenvolvimento de ontologias que inclui especificação, conceitualização, formalização, integração, implementação, manutenção, aquisição de conhecimentos, documentação e avaliação. Dentre estas atividades, o foco desta tese está nas atividades de conceitualização e de aquisição de conhecimentos.

Guizzardi e Halpin (2008) definem modelagem conceitual como uma disciplina cujo objetivo principal é o de identificar, analisar e descrever os conceitos essenciais e restrições

de um universo de discurso com o auxílio de uma linguagem de modelagem baseada num conjunto básico de conceitos de modelagem. Guizzardi (2005) ressalta dois aspectos importantes das especificações conceituais: *i)* é independente das escolhas de projeto ou de tecnologias; *ii)* são usadas para dar suporte à compreensão (aprendizado), resolução de problemas e comunicação entre **atores humanos** sobre um domínio. Neste trabalho, conceitualização significa construir uma especificação de um modelo conceitual por meio de uma linguagem a partir de um modelo embasado numa conceitualização, sendo que a construção mental de conceitualização é considerada parte da atividade de mesmo nome.

A **aquisição de conhecimentos**, classificada como atividade de suporte ao processo de desenvolvimento por (GÓMEZ-PÉREZ; CORCHO; FERNÁNDEZ-LÓPEZ, 2004), é realizada em praticamente todas as atividades do processo de desenvolvimento de ontologias, porém com intensidade variável. A aquisição de conhecimentos consiste na captura e na representação de uma conceitualização dando suporte à construção do modelo conceitual e da especificação do modelo conceitual.

Um problema bem conhecido pela comunidade de engenharia de conhecimentos é o do **gargalo de aquisição de conhecimentos** (*knowledge acquisition bottleneck*). O gargalo está na dificuldade dos projetistas de ontologias em capturar e representar o conhecimento do domínio. Este problema é relatado em diversas publicações na área de engenharia de ontologias (USCHOLD, 1996; FERNÁNDEZ-LÓPEZ et al., 1999; FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JURISTO, 1997; SURE, 2003; PINTO; TEMPICH; CIMIANO, 2006; GÓMEZ-PÉREZ; CORCHO; FERNÁNDEZ-LÓPEZ, 2004; MAEDCHE; STAAB, 2001; HOLSAPPLE; JOSHI, 2002; PINTO; MARTINS, 2004).

Na aquisição de conhecimentos, frequentemente os projetistas não conhecem o domínio e utilizam diferentes fontes de informação, incluindo nestas os especialistas do domínio, para aprender e elicitare conceitos, relações e axiomas a serem representados na ontologia. Ao longo do processo de desenvolvimento, o projetista passa gradualmente da conceitualização para a especificação do modelo conceitual. Este processo é longo, pois o conhecimento sobre o domínio é aprendido de maneira incremental. Mesmo após conhecer o domínio, os projetistas necessitam, muitas vezes, revisar a ontologia, pois os próprios especialistas do domínio possuem dificuldades em explicitar conhecimentos tácitos e em entender a representação construída. Deste modo, devido ao tamanho, complexidade e dinamicidade de um determinado domínio, e com o objetivo de reduzir os esforços necessários no processo de aquisição de conhecimento, vários métodos de aprendizado de ontologias têm sido desenvolvidos (GÓMEZ-PÉREZ; MANZANO-MACHO, 2003;

AUSSENAC-GILLES; BIÉBOW; SZULMAN, 2000; KIETZ; MAEDCHE; VOLZ, 2000; MAEDCHE; STAAB, 2001; SÁNCHEZ; MORENO, 2008).

Entende-se aprendizado de ontologias como um conjunto de métodos e técnicas utilizados para desenvolver novas ontologias, enriquecer ou adaptar ontologias existentes com novos elementos, de maneira semiautomática, utilizando diversas fontes de informação (GÓMEZ-PÉREZ; MANZANO-MACHO, 2003). O desafio do aprendizado de ontologias é extrair elementos lógicos dos dados de entrada (*e.g.* textos, bases de dados, dicionários) de maneira semiautomática (MAEDCHE; STAAB, 2000; CIMIANO, 2006). O aprendizado de ontologias não exclui outras técnicas, tais como a realização de entrevistas com especialistas do domínio, *brainstormings* e utilização de fontes diversas de informação.

Há várias abordagens para o aprendizado de ontologias que combinam processamento de linguagem natural, extração de informação, aprendizado de máquina e mineração de textos em *corpus* (MAEDCHE; STAAB, 2001) para extrair conceitos e relações entre os conceitos. Um dos problemas de se utilizar textos como entrada, é que muitas das informações necessárias para interpretá-los não estão explícitas. Algumas abordagens para suprir esta deficiência utilizam conhecimentos provenientes de outras fontes que incluem conhecimentos não ontológicos (*e.g.* dicionários, glossários, padrões) e ontológicos (ontologias). Além disso, o conhecimento formalizado por tais métodos pode estar incompleto, incorreto e inconsistente. O aprendizado de ontologias é suscetível a erros e numa ontologia gerada por estes métodos, há bastante trabalho manual de inspeção e validação da ontologia gerada (CIMIANO, 2006; WANG; VÖLKER; HAASE, 2006).

No desenvolvimento de ontologias, tanto manual pela utilização de alguma metodologia, quanto semiautomático pela utilização de métodos de aprendizado, o problema do gargalo na aquisição de conhecimentos persiste. O projetista precisa, de alguma forma, aprender sobre o domínio para ser capaz de compartilhar uma conceitualização com os especialistas do domínio e de validar a criação de elementos ontológicos sugeridos pelos métodos de aprendizado. Mesmo as metodologias colaborativas mais recentes, como NEON (SUÁREZ-FIGUEROA et al., 2008) e DILIGENT (PINTO; TEMPICH; STAAB, 2004), se baseiam na aquisição de conhecimentos tal como especificada em outras metodologias (notadamente, na METHONTOLOGY), bem como preconizam a utilização de métodos e técnicas consolidadas de aquisição de conhecimentos como leitura e entrevistas. Em geral as metodologias, com exceção da *On-To-Knowledge*, não propõem um processo de aprendizado de ontologias com o objetivo de reduzir esforços na aquisição e na conceitualização, porém, a *On-To-Knowledge* não especifica como deve ser este processo.

No propósito de atenuar o problema de aquisição de conhecimentos, folksonomias são uma alternativa interessante (WAL, 2007; GOLDER; HUBERMAN, 2005; JÄSCHKE et al., 2008; HOTH0 et al., 2006). As folksonomias são resultantes de um processo social de anotação de recursos Web com etiquetas (*tags*) e, eventualmente, com breves descrições dos mesmos. A anotação é uma atividade que não onera os usuários em termos de tempo e de curva de aprendizagem, já que consiste em atribuir palavras indexadoras a um recurso eletrônico sem regras de sintaxe rígidas. A leitura e anotação destes recursos são integráveis às atividades cotidianas dos atores e podem auxiliar na construção de uma conceitualização compartilhada do domínio. A dificuldade está em fixar o significado das *tags* em razão dos problemas de ambiguidade (sinonímia e polissemia) e em descobrir relações taxonômicas e outros tipos de relações devido à estrutura plana das anotações. Anotações realizadas por comunidades de usuários se estabilizam ao longo do tempo significando que um número razoável de usuários concorda que um conjunto particular de *tags* descreve bem um recurso (GOLDER et al., 2006). Nesta situação, diz-se que há emergência de conhecimento coletivo.

No contexto do desenvolvimento de ontologias, atores<sup>1</sup> com o objetivo comum de desenvolver uma ontologia podem realizar anotações em um conjunto de recursos que consideram representativos do domínio a ser modelado. Com o passar do tempo, há convergência para um vocabulário comum composto por *tags* que, em última análise, são termos conectados a exemplos de uso (*i.e.* aos recursos anotados) que explicitam o significado pretendido para os termos. Tal vocabulário pode auxiliar no desenvolvimento de ontologias, principalmente na atividade de conceitualização, onde as *tags* podem ser consideradas conceitos elicitados pelos atores e a correlação das *tags* (o fato de anotarem o mesmo recurso) pode ser indicativa da existência de relações e axiomas. Desta forma, a criação de folksonomias pode complementar as técnicas de aquisição de conhecimentos já existentes com a vantagem de colocar os especialistas do domínio como parte integrante do processo de desenvolvimento de ontologias.

Para entendimento do especialista como parte integrante do processo de desenvolvimento, deve-se distinguir desenvolvimento centralizado e colaborativo, tal como adotados pelo autor desta tese. No desenvolvimento colaborativo os atores que desempenham os papéis de projetistas, especialistas do domínio e de usuários atuam diretamente na especificação do modelo conceitual, logicamente, com ações correspondentes aos papéis

---

<sup>1</sup> Atores designa de forma genérica especialistas do domínio, projetistas e usuários (interessados na utilização da ontologia).

desempenhados. No desenvolvimento centralizado, todas as modificações na especificação do modelo conceitual, mesmo aquelas que dão apenas suporte ao desenvolvimento de tal artefato e que cabem aos especialistas do domínio (*e.g.* construir uma lista de termos elicitados), são sempre mediadas pelos projetistas da ontologia.

O processo de desenvolvimento de ontologias é incremental, pois a conceitualização compartilhada pelos atores do processo é construída de maneira incremental. Especialistas do domínio explicitam seus conhecimentos gradualmente e projetistas se apropriam dos mesmos também de maneira gradual. Além disso, como o conhecimento é modificado com o tempo, ontologias precisam de atualizações periódicas. Segundo Noy e McGuinness (2001), falta flexibilidade nas metodologias e nos métodos de aprendizado para atualizar, evoluir ou enriquecer ontologias existentes. Projetistas não precisam apenas adicionar novos elementos à ontologia, mas integrá-los coerentemente aos já existentes. Folksonomias são uma alternativa natural para dar início ao desenvolvimento e para evoluir ontologias, pois as folksonomias evoluem à medida que uma comunidade aprende e revisa seus conhecimentos sendo portanto úteis na detecção de mudanças de vocabulário ao longo do tempo. Tais mudanças, podem refletir mudanças na conceitualização da comunidade acerca de um domínio que podem ser incorporadas à ontologia.

No desenvolvimento de ontologias, seja este manual ou auxiliado por algum método de aprendizado semiautomático, técnicas de avaliação são necessárias. Avalia-se uma ontologia a fim de selecionar ontologias existentes de repositórios, avaliar o impacto de uma ontologia em aplicações de recuperação de informação, verificar a corretude da estrutura taxonômica da ontologia, analisar a estrutura e verificar a cobertura da ontologia em relação à alguma referência (*e.g.* corpus, ontologia padrão) (HARTMAN et al., 2005). Corretude da estrutura taxonômica da ontologia, nesta tese, se refere à verificação da natureza ontológica atribuída aos elementos da ontologia frente às regras de subsunção estabelecidas por *OntoClean* (Guarino e Welty, 2004).

Trabalhos relacionados ao desenvolvimento de ontologias reconhecem a importância da avaliação no desenvolvimento de ontologias (GRUBER, 1993; FERNÁNDEZ-LÓPEZ et al., 1999; STAAB et al., 2001; GRÜNINGER; FOX, 1995; PINTO; MARTINS, 2004; SURE et al., 2004; GÓMEZ-PÉREZ; CORCHO; FERNÁNDEZ-LÓPEZ, 2004). No entanto, a aplicação de métodos de avaliação (BURTON-JONES et al., 2005; GÓMEZ-PÉREZ, 1999; GÓMEZ-PÉREZ; JURISTO; PAZOS, 1995; GUARINO; WELTY, 2002; WELTY; GUARINO, 2001; TARTIR et al., 2005) requer experiência e esforços dos projetistas.

Guarino e Welty (2002) argumentam que a avaliação de ontologias é complexa, pois deve ser baseada em aspectos filosóficos do conhecimento acerca do domínio a ser representado. Por definir aspectos filosóficos, o projetista deve capturar mais do que a ontologia significa de uma forma concisa e formal (VÖLKER et al., 2008), deve capturar a natureza ontológica de um novo elemento, entender o significado de suas escolhas e, por conseguinte, quais serão as consequências ontológicas de suas escolhas de modelagem (RECTOR et al., 2004).

Por exemplo, *Pessoa* e *Professor* podem ser definidos como conceitos em uma ontologia, porém podem ser considerados diferentes do ponto de vista filosófico. As instâncias ou indivíduos de *Pessoa* sempre serão instâncias em todos os “mundos”. No entanto, os indivíduos de *Professor* podem ser instâncias de outro conceito em algum domínio (*i.e* podem ser *Estudantes* no domínio da capacitação) (GUARINO; WELTY, 2002; WELTY; GUARINO, 2001). Na literatura há métodos de avaliação que exploram as consequências ontológicas de certas escolhas de modelagem, porém estes métodos não exploram como identificar a natureza ontológica dos elementos tais como, essência, dependência e identidade (GUARINO; WELTY, 2002). Este, justamente, é um dos pontos abordados nesta tese quando se fala em integração das atividades de conceitualização e avaliação.

Concisamente, o problema abordado nesta tese é que embora existam metodologias de desenvolvimento e métodos de aprendizado de ontologias é preciso melhorá-los para que as atividades de **aquisição de conhecimentos**, **conceitualização** e **avaliação** sejam mais facilmente realizáveis pelos atores.

A deficiência das metodologias na aquisição de conhecimentos se deve principalmente à natureza consensual entre um grande número de pessoas que supostamente deve haver num desenvolvimento de uma ontologia. As técnicas de aquisição existentes são, na sua maioria, oriundas dos sistemas especialistas, onde se pretendia capturar o conhecimento de alguns poucos especialistas para aplicações particulares. Por este motivo, tais técnicas não são totalmente adaptadas às necessidades do desenvolvimento de ontologias. Por outro lado, as folksonomias representam emergência de conhecimento consensual coletivo e que são ainda pouco exploradas no desenvolvimento de ontologias.

A deficiência das metodologias na conceitualização e avaliação se refere à integração destas atividades no processo de desenvolvimento. A idéia é que as metodologias de avaliação sirvam para guiar o processo mental de conceitualização dos atores. Quanto mais cedo estas metodologias forem incorporadas ao processo, menos erros serão propagados para as fases posteriores de desenvolvimento.

Portanto, defende-se a tese que as metodologias de desenvolvimento de ontologias existentes não estão suficientemente equipadas para auxiliar os atores do processo nestas atividades e, portanto, precisam ser melhoradas para que todos possam participar ativamente perfazendo um desenvolvimento colaborativo, para facilitar a captura de uma conceitualização numa especificação do modelo conceitual de forma que esta reflita os significados pretendidos pela conceitualização, para alcançar a corretude da perspectiva do modelo e para verificar se a ontologia é válida de uma perspectiva formal.

## 1.2 HIPÓTESES

No contexto apresentado, faz-se as seguintes hipóteses:

1. A utilização de folksonomias pelos atores do processo de desenvolvimento de ontologias diminui o gargalo de aquisição de conhecimentos pois ajuda na capacidade cognitiva facilitando a conceitualização mental e explicitação de tal conceitualização;
2. A integração das atividades de conceitualização e de avaliação permite aos atores do processo de desenvolvimento elaborarem ontologias mais descritivas (*i.e.* com maior número de axiomas e de relações não taxonômicas) e com corretude da estrutura taxonômica em relação às consequências ontológicas pretendidas.

A primeira hipótese pressupõe que as *tags* de folksonomias auxiliam os atores a elicitarem conhecimentos a serem representados na ontologia, *i.e.* que grande parte das *tags* presentes numa folksonomia gerada pelos atores envolvidos no desenvolvimento da ontologia são transformadas em elementos da ontologia. Não se trata de eliminar outras técnicas de aquisição de conhecimentos, mas sim, de complementá-las auxiliando os especialistas do domínio a explicitarem e exemplificarem suas conceitualizações individuais e auxiliando o aprendizado dos projetistas sobre tais conceitualizações. Também está embutido o pressuposto que o processo social de etiquetagem aumenta a colaboração entre atores e permite que participem de maneira ativa no desenvolvimento da ontologia.

A segunda hipótese pressupõe que a integração da atividade de conceitualização com a de avaliação permite aos atores do processo de desenvolvimento realizarem conceitualizações explícitas corretas e consistentes e justificarem suas escolhas de

modelagem, o que é importante para que a comunidade definidora/usuária da ontologia atinja consenso baseando-se em justificativas bem fundamentadas. Se a avaliação for convenientemente combinada com a conceitualização, os atores serão capazes de produzir ontologias mais descritivas, com conceitos com um maior número de instâncias e com relações taxonômicas corretas em relação a natureza ontológica dos elementos segundo o preconizado pela metodologia *OntoClean*.

### 1.3 OBJETIVOS

O objetivo principal deste trabalho, no âmbito do desenvolvimento de ontologias, é desenvolver um método que permita reduzir o problema do gargalo de aquisição de conhecimentos e, também, melhorar as ontologias resultantes em relação à correteza da taxonomia<sup>2</sup> e ao conteúdo estrutural da ontologia. São objetivos específicos:

- realizar aquisição de conhecimentos a partir de folksonomias englobando a recuperação, representação e enriquecimento das mesmas;
- auxiliar aos projetistas a incluírem conhecimento adquirido das folksonomias na ontologia em desenvolvimento de forma ontologicamente correta.

### 1.4 CONTRIBUIÇÕES

Perante os benefícios do uso de ontologias e pela diversidade de aplicações é importante fornecer suporte metodológico, seja no desenvolvimento centralizado ou colaborativo de ontologias. No desenvolvimento centralizado, um projetista ou um grupo restrito de projetistas centraliza e insere/modifica elementos da ontologias a partir de sugestões de outros. No colaborativo, a participação não é mediada por um grupo restrito, isto é, o ator pode interagir diretamente com o artefato, embora esta interação possa ser limitada. Relacionado às **metodologias** (GRÜNINGER; FOX, 1995; USCHOLD; KING, 1995; FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JURISTO, 1997; SURE; STAAB; STUDER, 2002; PINTO; TEMPICH; STAAB, 2004; SUÁREZ-FIGUEROA et al., 2008), o método

---

<sup>2</sup> consistência ontológica em relação as noções filosóficas de essência, identidade e dependência preconizadas na *OntoClean* com o objetivo de melhor entender a natureza das relações taxonômicas.



desenvolvido pode ser utilizado pelos projetistas nas atividades de modelagem conceitual e de aquisição de conhecimentos. Na aquisição de conhecimentos difere dos métodos existentes pela utilização de folksonomias. Na modelagem conceitual, por induzir os atores a elicitarem coletivamente novos elementos e pela indagação da natureza ontológica do elemento ao utilizar a metodologia *OntoClean* de forma transparente.

O método proposto aborda pontos não explorados nas **metodologias** de desenvolvimento nas atividades de aquisição de conhecimentos, conceitualização e avaliação, entre eles:

- utilizar recursos e conhecimentos existentes (folksonomias) resultantes de um processo consensual;
- fornecer diretrizes de como construir uma ontologia a partir de folksonomias;
- incluir semântica em termos provenientes da anotação de recursos e utilizar os mesmos para induzir elementos na ontologia em desenvolvimento;
- utilizar um processo de avaliação na modelagem conceitual a fim de evitar a propagação de erros na ontologia em atividades subsequentes no ciclo de vida;
- fornecer suporte ao desenvolvimento colaborativo, pelos fatos seguintes: *i*) por utilizar folksonomias resultantes de um processo social de anotação; *ii*) por permitir enriquecer ou continuar o desenvolvimento de uma ontologia denominada raiz, também desenvolvida em consenso; e *iii*) por associar os elementos ontológicos (i.e conceitos, propriedades e instâncias) às fontes que lhes deram origem, mais precisamente às *tags* presentes na folksonomia.
- fornecer diretrizes de como enriquecer (incluir conhecimento na ontologia em desenvolvimento, ou enriquecer uma ontologia existente a partir de um processo de aprendizado de ontologias. Tanto a ontologia em desenvolvimento quanto a ontologia existente são consideradas nesta tese como ontologia raiz. ;
- reduzir esforços na aquisição de conhecimentos a partir de um processo de aprendizado;
- fornecer suporte metodológico no desenvolvimento centralizado ou colaborativo de ontologias, pelo fato das atividades de aquisição, modelagem conceitual e avaliação estarem presentes em metodologias de desenvolvimento tanto individuais, quanto colaborativas;

- desenvolver ontologias pelo nível de conhecimento ao utilizar as folksonomias e pelo nível ontológico pela utilização de características ontológicas provenientes da *OntoClean*;
- desenvolver ontologias que reflitam o significado pretendido, dado um contexto ontológico, acessar a corretude da perspectiva do modelo e verificar se a ontologia é válida de uma perspectiva formal (regras ontológicas).

Metodologias como a CYC, SENSUS e ENTERPRISE procuram fornecer suporte para o desenvolvimento de uma ontologia pela extensão de uma ontologia de nível superior. Estas metodologias são específicas a este tipo de desenvolvimento (extensão), sendo falhas no desenvolvimento iterativo de ontologias. Neste ponto a metodologia NEON (SUÁREZ-FIGUEROA et al., 2008), procura tabalhar o desenvolvimento colaborativo juntamente com a evolução da ontologia. Porém, mesmo a NEON preconiza a utilização de outros métodos de aquisição de conhecimentos (e.g. leitura, anotação, entrevistas, entre outros descritos por metodologias de desenvolvimento), bem como métodos semiautomáticos de aprendizado. Logo o método desenvolvido, torna-se uma alternativa na aquisição de conhecimentos e também no aprendizado semiautomático.

Em relação aos **métodos de aprendizado de ontologias** a partir de textos (SÁNCHEZ; MORENO, 2008; CIMIANO; WENDEROTH, 2005; YAMADA; BALDWIN, 2004; POESIO; ALMUHAREB, 2005; VILLAVERDE et al., 2009; SINTEK; BUITELAAR; OLEJNIK; 2004; VELARDI et al., 2005; MORIN, 1999; KIETZ; MAEDCHE; VOLZ, 2000; AUSSENAC-GILLES; BIÉBOW; SZULMAN, 2000), visto que o conhecimento formalizado por tais métodos pode estar incompleto e incorreto, procura-se gerar uma estrutura taxonômica consistente, coerente com as consequências ontológicas da *OntoClean*. Outra questão levantada nos métodos de aprendizado é que muitas das informações utilizadas para interpretar textos e identificar elementos e relações não estão explícitas. No método procura-se suprir esta deficiência através do enriquecimento das folksonomias buscando-se significados em outras fontes (e.g. *WordNet*, acrônimos), pela análise formal de conceitos (FCA – *Formal Concept Analysis*) e pela utilização de um padrão, por meio de um questionário eletrônico, que torna transparente a utilização da *OntoClean*. Este padrão procura identificar a natureza ontológica dos novos elementos e das relações entre os mesmos, natureza esta que não está explícita nos textos. Em relação aos métodos de **aprendizado de ontologias** a partir de **folksonomias** (MIKA, 2005; WU; ZHANG; YU, 2006; BEGELMAN; KELLER; SMADJA, 2006; SCHMITZ, 2006; WU; ZUBAIR; MALY, 2006; LUX;

DSINGER, 2007; ANGELETOU et al., 2007; SPECIA; MOTTA, 2007; LANIADO; EYNARD; COLOMBETTI, 2007; SILVA, 2009; DAMME; HEPP; SIORPAES, 2007), a contruição está na geração de uma estrutura mais rica do que simplesmente correlação e agrupamento de *tags*. Nos métodos existentes, são identificados conceitos e relações entre os mesmos com base em cálculos estatísticos ou em alguma relação existente em outra fonte. Logo geram-se, no máximo, conceitos e taxonomias, sendo que as taxonomias são resultantes de grupos de conceitos ou de termos identificados nos dicionários. Além disso o propósito dos métodos atuais é utilizar a geração de uma estrutura a partir das folksonomias que facilite a navegação e indexação de recursos, sendo que o método proposto procura desenvolver uma ontologia, ou seja, identificam-se conceitos, relações de taxonomia e de propriedades entre os conceitos, bem como a identificação de instâncias.

Relacionado ao desenvolvimento de ontologias a partir de padrões ontológicos (GUIZZARDI, 2005; BLOMQVIST, 2005), o método desenvolvido utiliza a taxonomia idealizada pela *OntoClean* (WELTY; GUARINO, 2001; GUARINO; WELTY, 2002) como padrão de modelagem. No *OntoCase* (BLOMQVIST, 2005), há a necessidade a priori de criar os padrões. A *OntoUML* (GUIZZARDI, 2005) enriquece a semântica de diagramas de classes UML com distinções ontológicas da UFO (*Unified Foundational Ontology*). Logo, as diferenças em relação às abordagens baseadas em padrões residem na técnica de aquisição de conhecimentos utilizada (leitura e anotação), no artefato de aquisição de conhecimento (folksonomias) e no padrão de modelagem utilizado para conceitualizar os elementos na ontologia. Guizzardi (2005) utiliza a UFO como padrão, a *OntoCase* elabora a priori primitivas descritas em uma ontologia. O trabalho proposto pretende utilizar as diretivas da *OntoClean* como um padrão na identificação de elementos e da relação entre os mesmos.

Além do desenvolvimento de ontologias iniciais, o método pode ser utilizado por projetistas na evolução de ontologias, com o objetivo de manter e enriquecer suas ontologias enquanto adquirem novos conhecimentos pela anotação de recursos e consequente produção de folksonomias. Uma ontologia é desenvolvida por sucessivas iterações. A cada nova iteração, novos conhecimentos são incluídos na ontologia provenientes do processo de anotação dos atores. Neste processo, novos conhecimentos são explicitados e estes podem ser incluídos ou modificar conhecimentos já existentes na ontologia.

## **1.5 ESTRUTURA DO TRABALHO**

Esta tese contém cinco capítulos além desta introdução, das referências bibliográficas e dos apêndices. No capítulo 2 são apresentados os conhecimentos fundamentais que constituem o alicerce para elaborar este trabalho e os trabalhos relacionados. No capítulo 3 é apresentado o método de desenvolvimento de ontologias que inclui a aquisição de conhecimentos e a conceitualização. Neste capítulo também é apresentada a metodologia dos experimentos realizados. Encontram-se no capítulo 4, os experimentos, resultados obtidos e discussões. As conclusões e discussões sobre o trabalho desenvolvido, assim como as perspectivas dos possíveis desdobramentos em trabalhos futuros são encontrados no capítulo 5.

## 2. FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem por objetivo descrever conceitos, fundamentos, definições e trabalhos da literatura que se relacionam com o desenvolvimento desta tese. Assim, neste capítulo define-se ontologia e descreve-se o processo de desenvolvimento de ontologias que incluem metodologias e métodos de aprendizado, além das atividades abordadas nesta tese para o desenvolvimento, entre elas, aquisição, conceitualização e avaliação. Também descreve-se neste capítulo o processo de anotação, folksonomias, técnicas estas que subsidiam o desenvolvimento do método, entre elas, processamento de linguagem natural e medidas de similaridade.

### 2.1 ONTOLOGIAS

Nesta seção, são apresentadas definições da literatura acerca de ontologias bem como temas relacionados, tais como, classificação de ontologias, formalismos, processo de desenvolvimento, aprendizado e avaliação.

#### 2.1.1 Definições

O termo ontologia tem diversas definições na filosofia. O dicionário *Webster* (GUIZZARDI, 2005 *apud* MERRIAM-WEBSTER, 2004) apresenta três definições de ontologia: *i*) um ramo da metafísica que denota uma teoria sobre a natureza do ser ou existência; *ii*) uma teoria particular sobre a natureza do ser e tipos existentes; *iii*) uma teoria sobre os tipos de entidades abstratas admitidas na linguagem. Etimologicamente, a palavra Ontologia (*ont + logia*) pode ser traduzida como o estudo da existência (GUIZZARDI, 2005). No sentido filosófico, ontologia visa desenvolver teorias sobre, por exemplo, existência, essência, identidade, causalidade, entre outras. Destas teorias surgem questões filosóficas, entre elas: que tipos de entidades existem? O que é essência? Os conceitos existem ou estão presentes apenas na imaginação? Como as entidades do mundo podem ser classificadas? Qual a relação entre as entidades? A essência precede a existência. No final do século 20, ontologias emergem como uma importante área de pesquisa na ciência da computação sendo

utilizada em três importantes áreas (SMITH; WELTY, 2001): *i*) sistemas de informação e base de dados; *ii*) engenharia de software; *iii*) inteligência artificial.

Na inteligência artificial, ontologia possui uma interpretação mais relacionada à modelagem de conhecimento, ou seja, considera-se ontologia como um artefato, compreendendo um vocabulário de termos estruturado em uma taxonomia, suas definições e um conjunto de axiomas formais utilizados para criar novas relações e para restringir as suas interpretações (GUIZZARDI, 2005; GUARINO, 1998).

Em Inteligência Artificial, ontologia pode ser interpretada como um conjunto de entidades com suas relações, restrições, axiomas e vocabulário. Uma das primeiras definições de ontologias foi de Neches et al. (1991): “... *an ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary.*” Esta definição identifica condições básicas para construir uma ontologia: termos, relações entre termos e regras. Alguns anos depois, Gruber (1993) definiu ontologia como uma especificação explícita de uma conceitualização. Por especificação explícita entende-se que as entidades e suas restrições estão claramente definidas. O termo conceitualização refere-se a um modelo abstrato de algum fenômeno no mundo, *i.e.* uma área de conhecimento que se deseja representar.

A definição de Gruber (1993) é complementada por Studer, Benjamins e Fensel (1998) para os quais uma ontologia é uma especificação explícita e formal de uma conceitualização compartilhada de um domínio de interesse. Nesta definição, a palavra ‘compartilhada’ significa que uma ontologia deve capturar o conhecimento consensual e formal, que a ontologia deve ser declarativamente definida e interpretável por máquina.

Guarino e Giaretta (1995) propõem considerar ontologia como “*a logical theory which gives an explicit, partial account of a conceptualization*”. Os autores incluem a teoria lógica como base para desenvolver uma ontologia no contexto de sistemas computacionais. Esta definição é refinada por Guarino (1998) para “*a set of logical axioms designed to account for the intended meaning of a vocabulary*”.

As relações entre vocabulário, conceitualização e ontologia estão ilustradas na Figura 1, extraída de (GUARINO, 1998). Dada uma linguagem  $L$  com um compromisso ontológico  $K$  (toda linguagem se compromete com a existência de certas entidades do mundo por meio de suas primitivas), uma ontologia em  $L$  é um conjunto de axiomas concebidos de modo que o conjunto de seus modelos se aproxima o melhor possível do conjunto de modelos pretendidos de  $L$  de acordo com  $K$ . Na figura, o comprometimento  $K$  é definido por

$K = \langle C, \mathfrak{I} \rangle$ , sendo  $C$  uma conceitualização definida pela tripla  $\langle D, W, \mathfrak{R} \rangle$ , e  $\mathfrak{I}$  uma interpretação intencional dada pela função  $\mathfrak{I}: V \rightarrow D \cup \mathfrak{R}$ . Nestas definições, tem-se:

- $D$  é um domínio;
- $W$  é o conjunto máximo do estado das coisas do domínio  $D$  (também chamado de mundos possíveis);
- $\mathfrak{R}$  é o conjunto de relações conceituais no espaço  $\langle D, W \rangle$ ;
- $V$  é o vocabulário do domínio constituído por símbolos de constantes e de predicados.

A Figura 1 mostra que os modelos pretendidos  $I_k(L)$  de uma linguagem lógica  $L$  refletem o comprometimento desta linguagem com uma conceitualização particular do mundo. Uma ontologia reflete indiretamente este comprometimento e pode ser considerada uma tentativa de se aproximar dos modelos pretendidos. Por este motivo, diz-se que uma ontologia é uma representação parcial de uma conceitualização.

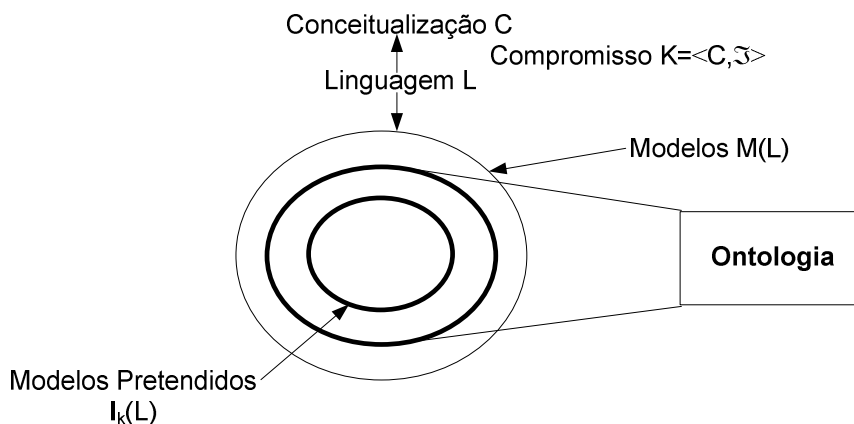


Figura 1 – Relacionamento entre Ontologia, Linguagem e Conceitualização

Fonte: Guarino (1997).

Quando dois sistemas de informação adotam o mesmo vocabulário, não há garantias de entendimento se eles não utilizam a mesma conceitualização. Assumindo que cada sistema possui sua própria conceitualização, uma condição necessária para um entendimento é que os modelos pretendidos das conceitualizações originais se sobreponham. A Figura 2 mostra dois sistemas que utilizam a mesma linguagem  $L$  com sobreposição dos modelos pretendidos (GUARINO, 1998). Neste caso, há um entendimento entre os processos.

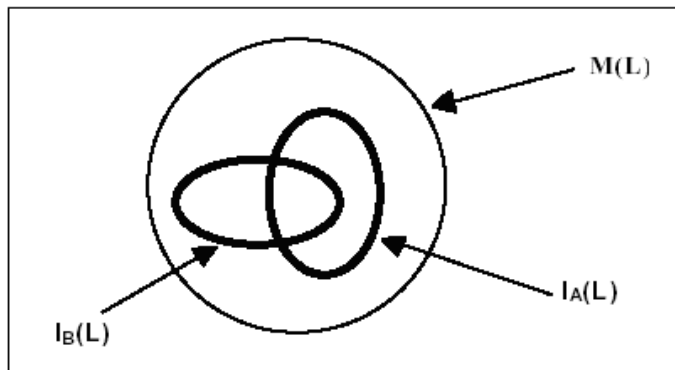


Figura 2 - Dois sistemas A e B utilizando a mesma linguagem L com sobreposição dos modelos pretendidos  $I_A(L)$  e  $I_B(L)$

Fonte: Guarino (1998).

Supõe-se agora que os modelos pretendidos são aproximados por duas ontologias diferentes (em cinza claro). Observa-se que neste caso as duas ontologias se sobrepõem, enquanto que os modelos pretendidos não (Figura 3). Isto significa que uma abordagem *bottom-up* na integração de sistemas de informação baseada na integração de ontologias pode não funcionar.

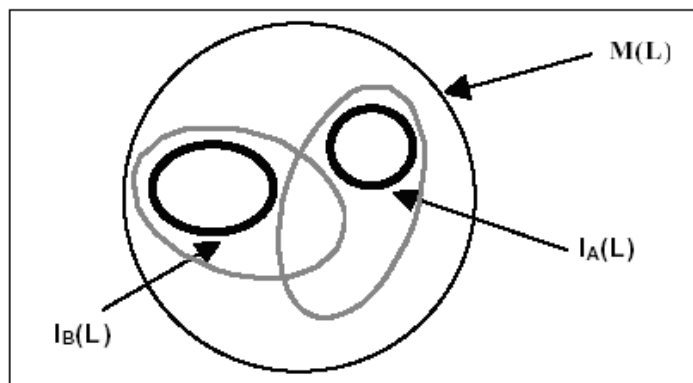


Figura 3 - Sobreposição de ontologias (em cinza claro) sem a sobreposição dos modelos pretendidos

Fonte: Guarino (1998).

### 2.1.2 Classificação das Ontologias

Baseado nesta teoria, Guarino (1998) sugere o desenvolvimento de diferentes tipos de ontologias de acordo com o nível de generalidade e propõe a seguinte classificação: nível



superior (*i.e.* ontologia de alto nível, ontologia de topo, ontologia genérica), domínio, tarefa e aplicação (Figura 4).

Em uma ontologia de nível superior descrevem-se conceitos como objeto, evento, ação, *i.e.* conceitos que são gerais e independentes de um domínio particular. Guizzardi, Falbo e Guizzardi R. (2008) nomeiam as ontologias de nível superior como ontologias de fundamentação e as definem como sistemas de categorias filosoficamente bem fundamentados e independentes de domínio. Algumas ontologias de nível superior são: DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*) (BOTTAZZI; FERRARIO, 2006), SUMO (*Suggested Upper Merged Ontology*) (NILES; PEASE, 2001), UFO (*Unified Foundational Ontology*) (GUIZZARDI, 2005) e CYC (LENAT; GUHA, 1990).

As ontologias de domínio descrevem o vocabulário de um domínio (*e.g.* medicina, turismo) através da especialização de conceitos definidos nas ontologias de nível superior. As ontologias de tarefas definem processos ou tarefas de um domínio pela especialização de conceitos introduzidos nas ontologias de nível superior proporcionando uma visão mais funcional, embora declarativa de um domínio (FREITAS, 2003). As ontologias de aplicação solucionam problemas específicos de um domínio, definindo conceitos que dependem tanto de um domínio específico, como de tarefas específicas e, geralmente, são uma especialização de ambas.

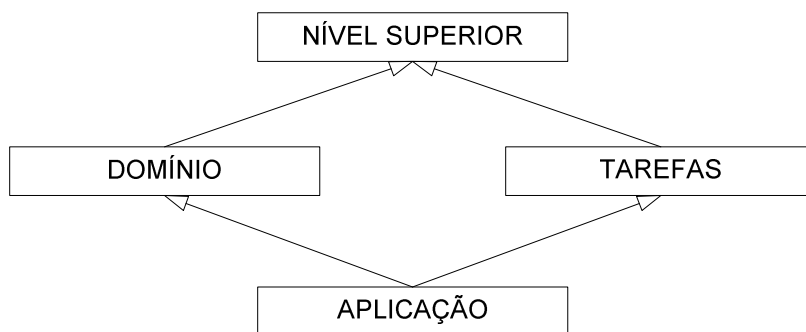


Figura 4 – Classificação de Ontologias

Fonte: Guarino (1998).

McGuinness (2002) propõe uma classificação baseada na estrutura interna e no conteúdo das ontologias (Figura 5). Essa classificação segue uma linha em que as ontologias são dispostas desde a mais informal como catálogos de termos (linguagem desestruturada) até a mais formal com o máximo de expressividade (restrições lógicas). A linha de corte na

Figura 5, segundo McGuinness (2002), indica que tudo que está do lado direito pode ser chamado de ontologia e deve satisfazer as seguintes condições: vocabulário controlado, interpretação inequívoca de classes e relacionamentos, relacionamento hierárquico estrito entre as classes. A linha de corte permite classificar as ontologias à esquerda em leves (*lightweight*) e à direita ontologias mais pesadas (*heavyweight*) (MCGUINNESS, 2002).

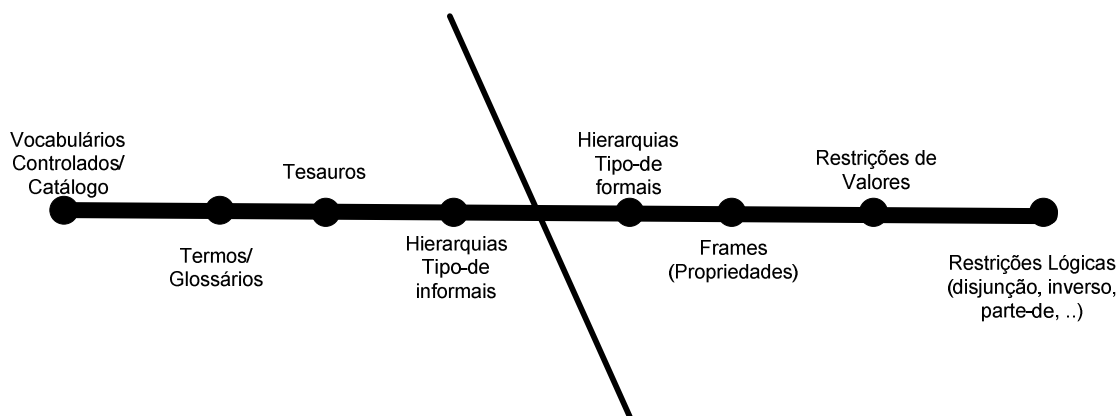


Figura 5 – Classificação de Ontologias

Fonte: McGuinness (2002).

Nesta tese, assume-se que uma ontologia é tanto mais **descritiva** quanto mais pesadas forem. Segundo (GUARINO, 1998), construir uma ontologia é restringir as interpretações possíveis e isto é alcançado pela definição de propriedades e axiomas de definições de classes.

## 2.2 DESENVOLVIMENTO DE ONTOLOGIAS

O IEEE (1990) define metodologia como “*a comprehensive, integrated series of techniques or methods creating a general systems theory of how a class of thought intensive work ought be performed*”. Logo, uma metodologia é composta de métodos e técnicas. Um método é um conjunto de “*orderly process or procedure used in the engineering of a product or performing a service*” (IEEE,1990). Uma técnica é “*a technical and managerial procedure used to achieve a given objective*” (IEEE, 1990). De agora em diante, os termos metodologia, método e técnica são utilizados com o significado exposto.

Há diversas metodologias para desenvolvimento de ontologias oriundas, na visão do autor desta tese, de experiências de desenvolvimento ou da combinação de metodologias de engenharia de software e de conhecimentos (Figura 6). No primeiro grupo, daquelas provenientes de experiências de desenvolvimento geralmente relacionadas ao desenvolvimento de grandes ontologias, encontram-se CYC (LENAT; GUHA, 1990), TOVE (*TO*ronto *V*irtual *E*nterprise) (GRÜNINGER; FOX, 1995), *Enterprise Ontology* (USCHOLD; KING, 1995; USCHOLD, 1996) e SENSUS (SWARTOUT et al., 1997). CYC, TOVE, Enterprise e SENSUS preconizam o desenvolvimento de ontologias a partir de ontologias de nível superior.

No segundo grupo, combinação de engenharia de software e de conhecimentos, encontram-se a METHONTOLOGY (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JURISTO, 1997) e a *On-To-Knowledge* (SURE; STAAB; STUDER, 2002; SURE; STUDER, 2002; SURE; STAAB; STUDER, 2003) foram desenvolvidas com a influência das metodologias de desenvolvimento de software, porém focadas em uma nova disciplina, a engenharia de conhecimento, mais especificamente, engenharia de ontologias.

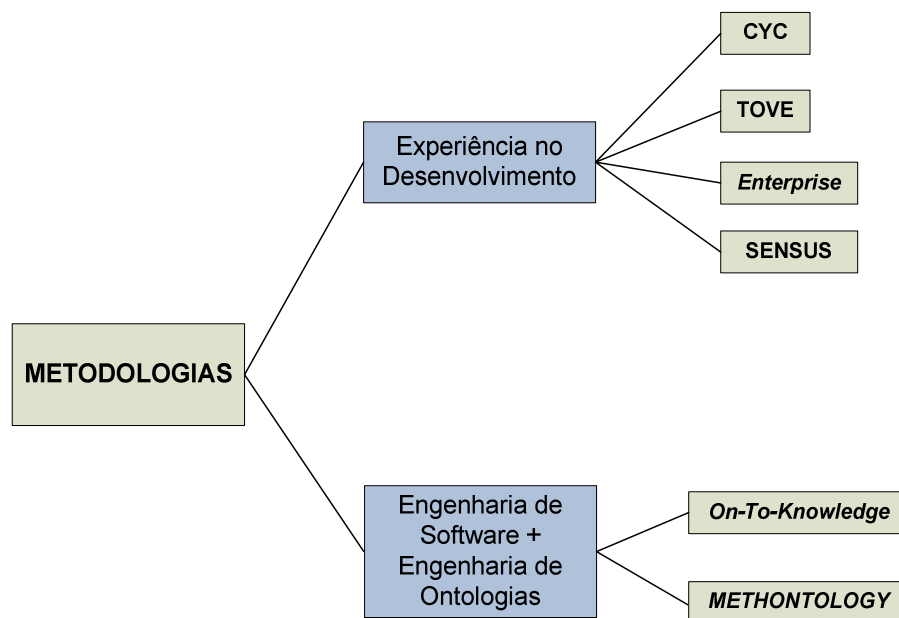


Figura 6 – Metodologias para o Desenvolvimento de Ontologias.

Neste trabalho, adotou-se a METHONTOLOGY como referência ao se fazer um sobrevoo nas atividades de conceitualização, aquisição de conhecimentos e avaliação nas diversas metodologias. METHONTOLOGY foi umas das metodologias precursoras e, na

literatura, encontra-se um maior detalhamento das suas atividades do que para as demais, além de ter sido fonte inspiradora para muitas outras. Pretende-se, por meio deste sobrevoo, demonstrar que as metodologias de desenvolvimento de ontologias apresentadas não estão suficientemente equipadas para auxiliar os atores do processo nas atividades de aquisição de conhecimentos, conceitualização e validação, conforme a tese colocada na introdução deste documento. A seção seguinte oferece uma visão geral da METHONTOLOGY enquanto as seções subsequentes (2.2.2 e 2.2.3) apresentam comparações entre a METHONTOLOGY e as demais no que se refere às atividades de aquisição de conhecimento e conceitualização. No que se refere à avaliação, primeiro apresenta-se uma visão geral de avaliação em ontologias para, em seguida, detalhar-se as metodologias *OntoQA* e *OntoClean* utilizadas nesta tese.

### 2.2.1 METHONTOLOGY

A METHONTOLOGY foi criada no Laboratório de Inteligência Artificial da Universidade Politécnica de Madrid em 1996, sendo complementada em trabalhos seguintes (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JURISTO, 1997; FERNÁNDEZ-LÓPEZ et al., 1999). Para o desenvolvimento de ontologias a METHONTOLOGY propõe uma série de atividades categorizadas em **gerenciamento**, **desenvolvimento** e **suporte** (Figura 7). As atividades de desenvolvimento são executadas sequencialmente, enquanto que as atividades de **suporte** e **gerenciamento** são executadas concomitantemente às atividades de **desenvolvimento**.

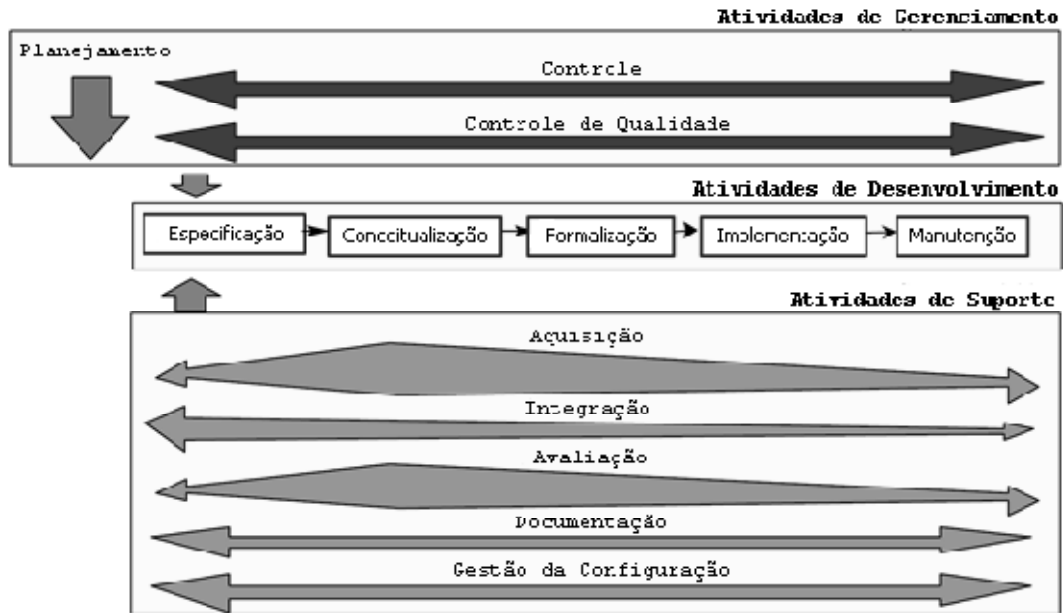


Figura 7 – Desenvolvimento e Ciclo de Vida de uma Ontologia na Metodologia METHONTOLOGY

Fonte: Adaptado de Fernández-López, Gómez-Pérez, Juristo (1997).

O **gerenciamento** possui as seguintes atividades:

- **Planejamento:** identifica as tarefas a serem executadas, o tempo e os recursos necessários para sua execução.
- **Controle:** procura garantir que as tarefas sejam executadas da maneira prevista no planejamento.
- **Garantia de Qualidade:** assegura a qualidade da documentação e da ontologia desenvolvida.

As atividades de **desenvolvimento** fazem parte do ciclo de vida da ontologia que indica a sequência das atividades a serem executadas. Levando em conta que o desenvolvimento de ontologias requer um processo iterativo (com revisões e evoluções constantes), o ciclo de vida do desenvolvimento de uma ontologia identifica a ordem de execução das atividades. O desenvolvimento possui as seguintes atividades:

- **Especificação:** define o escopo e os objetivos da ontologia, questiona a construção da ontologia (por que a ontologia está sendo construída?) e sua utilização, define usuários que a utilizarão e a finalidade da mesma (os propósitos e usos esperados da ontologia). As respostas a esses questionamentos são descritas em um documento chamado de especificação de requisitos da ontologia.

- **Conceitualização:** captura e estrutura o conhecimento do domínio em um modelo conceitual. A captura ou aquisição de conhecimentos é realizada na atividade de suporte de mesmo nome.
- **Formalização:** transforma o modelo conceitual em um modelo computável por máquina.
- **Implementação:** torna a ontologia computável pela tradução em uma linguagem de representação (codificação) de ontologias (*e.g* OWL<sup>3</sup> – *Web Ontology Language*).
- **Manutenção:** corresponde às atualizações e correções na ontologia.

O **suporte** inclui uma série de atividades desenvolvidas simultaneamente às atividades de desenvolvimento. São atividades do suporte:

- **Aquisição de conhecimentos:** compreende adquirir conhecimentos do domínio com base em documentos ou especialistas do domínio. A aquisição pode ser realizada manualmente ou através de algum processo semiautomático denominado aprendizado de ontologias (KIETZ; MAEDCHE; VOLZ, 2000).
- **Integração:** requerida quando uma nova ontologia é construída a partir de outras ontologias disponíveis. A combinação (*merging*) obtém uma ontologia a partir de combinações de ontologias de mesmo domínio (NOY; MUSEN, 2001; STUMME; MAEDCHE, 2001; PINTO; GÓMEZ-PÉREZ; MARTINS, 1999). A ontologia é resultante da unificação entre definições de todas as ontologias de origem. O alinhamento identifica mapeamentos entre as ontologias envolvidas preservando as ontologias originais (KLEIN, 2001; EHRIG, 2006).
- **Avaliação:** julga tecnicamente a ontologia e sua documentação durante e entre as fases do ciclo de vida de um sistema. Verifica se a ontologia atende os requisitos e propósitos planejados.
- **Documentação:** descreve as atividades concluídas e os produtos gerados visando facilitar futuro reuso e manutenção.
- **Configuração:** mantém registros de todas as versões da documentação e da ontologia gerada.

---

<sup>3</sup> <http://www.w3.org/TR/owl-features/>

As atividades de suporte são executadas em paralelo às atividades de desenvolvimento, sendo que, aquisição de conhecimentos e avaliação requerem maiores esforços dos atores quando da realização da atividade de conceitualização. As razões são que a maior parte do conhecimento para a construção da ontologia é obtida no início do desenvolvimento e a conceitualização deve ser avaliada para evitar futuros erros no ciclo de desenvolvimento (GÓMEZ-PÉREZ; CORCHO; FERNÁNDEZ-LÓPEZ, 2004).

A principal contribuição da METHONTOLOGY é a identificação de um processo de desenvolvimento, identificação do ciclo de vida e diretrizes de como construir uma nova ontologia. A METHONTOLOGY também ressalta a avaliação de ontologias durante a atividade de conceitualização a fim de evitar a propagação de erros na ontologia em atividades seqüentes no ciclo de vida.

### **2.2.2 Aquisição de Conhecimentos**

Segundo Kendal e Creen, (2007) apud Debenham (1997) define-se conhecimento: “... *the explicit functional association between items of information and/or data*”. Milton (2007) relaciona outras definições existentes na literatura para conhecimento, entre elas: conhecimento é uma forma estruturada de informação; conhecimento é o que é necessário para pensar como um especialista; conhecimento é o que separa especialistas de não especialistas; conhecimento é o que é requerido para fazer tarefas complexas. A partir destas definições existentes na literatura, Milton (2007) apresentou uma definição múltipla de conhecimento: “... é a habilidade/competência/perícia para manipular/transformar/criar dados/informações/idéias com o objetivo de fazer algo/tomar decisões/resolver problemas”.

Na disciplina da engenharia de conhecimentos, conhecimento é algo que pode ser expresso como uma regra ou que seja útil para uma tomada de decisão (KENDAL; CREEN, 2007).

A aquisição de conhecimentos envolve obter conhecimento de várias fontes incluindo especialistas, livros, vídeos, bases de dados e Internet (KENDAL; CREEN, 2007). Rolston (1988) define a aquisição de conhecimentos como a explicitação e captura de conhecimento em um formato estruturado.

Para aquisição de conhecimentos há algumas técnicas, entre elas, imersão na literatura, construção de glossários, construção/simulação de cenários, técnicas de observação, classificação de conceitos e *brainstorming* (*i.e.* tempestade de idéias).

Milton (2007) descreve três grupos de técnicas para a aquisição de conhecimentos com suas respectivas variantes (subitens da lista abaixo):

1. **de entrevista**
  - a. **não estruturada:** conversa livre com o especialista;
  - b. **semiestruturada:** conjunto de questões pré-definidas e questões suplementares;
  - c. **estruturada:** apenas questões pré-definidas.
2. **de modelagem:** compreendem o desenvolvimento de modelos de conhecimento com ajuda de especialistas do domínio, nesta tese, denominados de especificações de modelos conceituais. As especificações conceituais são uma forma de visualizar o conhecimento e de compreender o domínio.
3. **especializadas:** de maneira geral, estas técnicas são elaboradas por psicólogos tendo por objetivo extrair o conhecimento tácito dos especialistas do domínio.

Na engenharia de ontologias a aquisição de conhecimento compreende a eliciação do conhecimento necessário sobre o domínio para a construção da ontologia (PINTO; MARTINS, 2004). Na METHONTOLOGY a aquisição de conhecimentos é uma atividade independente no desenvolvimento da ontologia, porém coincide com outras atividades de desenvolvimento, sendo mais presente na conceitualização. A METHONTOLOGY dita uma aquisição em três estágios. No primeiro estágio, faz-se uma análise geral do conhecimento obtido dos especialistas do domínio. No segundo estágio, realiza-se um estudo detalhado da documentação. O último estágio compreende a observação de detalhes dos conhecimentos adquiridos para refiná-los. Nestes estágios, a METHONTOLOGY propõe as seguintes técnicas de aquisição de conhecimentos:

- **Entrevistas não estruturadas com especialistas:** entrevista livre com especialistas do domínio. Nesta técnica é produzido um esboço inicial do documento de especificação, bem como uma lista inicial de termos e conceitos.
- **Análise informal de textos:** imersão na literatura para estudo dos principais conceitos em livros e dicionários. O resultado é uma definição e uma classificação dos conceitos.
- **Análise formal de textos:** identificação de padrões no texto pela análise de sentenças. Nesta técnica as sentenças com determinados padrões são analisadas formalmente para extração de regras, atributos e possíveis valores de atributos.



- **Entrevistas estruturadas com especialistas:** entrevista com questões pré-definidas para obter detalhes dos conceitos e relacionamentos. Esta técnica é utilizada para avaliar a especificação do modelo conceitual, visto que na realização da mesma, a atividade de conceitualização deve ter sido finalizada.
- **Análise das tabelas de domínio:** verificação das tabelas de domínio construídas na conceitualização, principalmente dos valores de atributos dos conceitos.
- **Análise dos grafos do domínio:** verificação dos relacionamentos entre conceitos para a criação de novos relacionamentos.
- **Análise das unidades de medida:** determinação dos atributos e valores dos mesmos presentes em fórmulas.
- **Revisão detalhada do especialista:** análise detalhada pelo especialista dos artefatos (tabelas, textos, modelos) produzidos na conceitualização. Nesta técnica o especialista inclui sugestões e correções aos artefatos.
- **Análise de fórmulas:** verificar se as fórmulas e seus atributos estão corretas.

O Quadro 1 apresenta uma lista dos artefatos produzidos pelas metodologias em atividades que correspondem a aquisição de conhecimentos na METHONTOLOGY.

Quadro 1 - Artefatos produzidos pelas Metodologias na Aquisição de Conhecimento.

Artefatos para a Aquisição de Conhecimentos	
<b>TOVE</b>	Ausente.
<i>Enterprise</i>	Lista de termos relevantes.
<b>SENSUS</b>	Lista inicial de conceitos.
<i>On-To-Knowledge</i>	Modelos do domínio.
<b>METHONTOLOGY</b>	Lista inicial de termos e conceitos; definição inicial de termos e conceitos; lista de regras, atributos e possíveis valores para os atributos; detalhamento de conceitos e relacionamentos.

No Quadro 2 identificam-se as técnicas de aquisição de conhecimentos utilizadas pelas metodologias. Pelo fato da METHONTOLOGY ter uma fase específica na aquisição de conhecimentos, há uma superioridade da mesma no número de técnicas em comparação a outras metodologias.

Quadro 2 - Técnicas de Aquisição de Conhecimentos das Metodologias.

Técnicas de Aquisição de Conhecimentos	
<b>TOVE</b>	Elaboração das questões de competência.
<i>Enterprise</i>	Ausente
<b>SENSUS</b>	Processo manual de integração a SENSUS.
<i>On-To-Knowledge</i>	Elaboração de modelos.
<b>METHONTOLOGY</b>	Entrevistas não estruturadas, análise informal de textos, análise formal de textos, entrevistas estruturadas, análise de tabelas do domínio, análise de grafos, análise das unidades de medida, revisão detalhada do especialista, análise de fórmulas.

### 2.2.3 Conceitualização

A conceitualização organiza e estrutura o conhecimento capturado na aquisição de conhecimentos utilizando representações externas independentes de linguagens de implementação. Nesta fase, as informações obtidas na aquisição de conhecimentos são transformadas em uma especificação do modelo conceitual do domínio pelos atores do processo. Para tal, os especialistas do domínio e os projetistas utilizam um conjunto de representações intermediárias (ou artefatos) que permitem a eles atingir consenso sobre o modelo conceitual representado. Simultaneamente à atividade de conceitualização, de acordo com a METHONTOLOGY, são realizadas as atividades de suporte de aquisição de conhecimentos e de avaliação.

A METHONTOLOGY propõe um conjunto de artefatos para guiar a conceitualização a atividade de conceitualização. São eles (Figura 8):

- **Glossário de termos:** identificação de termos (conceitos, instâncias e atributos) do domínio com as descrições.
- **Classificação de conceitos:** definição da taxonomia de conceitos e conceitos disjuntos.
- **Diagrama de relações binárias:** definição de relacionamentos não taxonômicos (propriedades) entre os conceitos da taxonomia gerada na etapa anterior. Utilizado também para integrar conceitos de outras ontologias.
- **Dicionário de conceitos:** criação de uma tabela com todos os conceitos do domínio, instâncias desses conceitos, atributos de conceitos (nome do atributo),

atributos de instâncias (atributos definidos nos conceitos e valorados nas instâncias) dos conceitos, sinônimos, antônimos e relações não taxonômicas.

- **Tabela de relações binárias:** descrição para cada relação binária do nome, imagem, escopo, relação inversa e cardinalidade da relação.
- **Tabela de atributos de instância:** definição para atributo de instância das seguintes informações: nome, tipo de valor, unidade de medida para valores numéricos, precisão dos valores, faixa de valores, valor padrão, cardinalidade máxima e mínima, atributos que podem ser inferidos deste valor, fórmulas ou regras para inferir o atributo e referências utilizadas para preencher o atributo.
- **Tabela de atributos de conceito:** definição para cada atributo de conceito das seguintes informações: nome do atributo, tipo de valor, unidade de medida e precisão para valores numéricos, cardinalidade do atributo e atributos de instância que podem ser inferidos.
- **Tabela de axiomas:** definição de expressões lógicas para os conceitos. Para cada axioma define-se: nome do conceito, descrição em linguagem natural, conceitos e atributos aos quais o axioma se refere e a expressão lógica que define o axioma.
- **Tabela de constantes:** especificação para cada constante do nome, descrição em linguagem natural, tipo do valor, valor da constante, unidade de medida e atributos que podem ser inferidos pela constante.
- **Tabela de fórmulas:** descrição para cada fórmula dos seguintes itens: nome, atributos inferidos, expressão matemática, descrição em linguagem natural, atributos e constantes utilizados e precisão dos valores calculados pela fórmula.
- **Arvore de classificação de atributos:** construção gráfica dos atributos e constantes relacionados na sequência de fórmulas utilizadas.
- **Tabela de instâncias:** definição das instâncias do domínio pelo nome, os atributos de instâncias e valores para os mesmos.

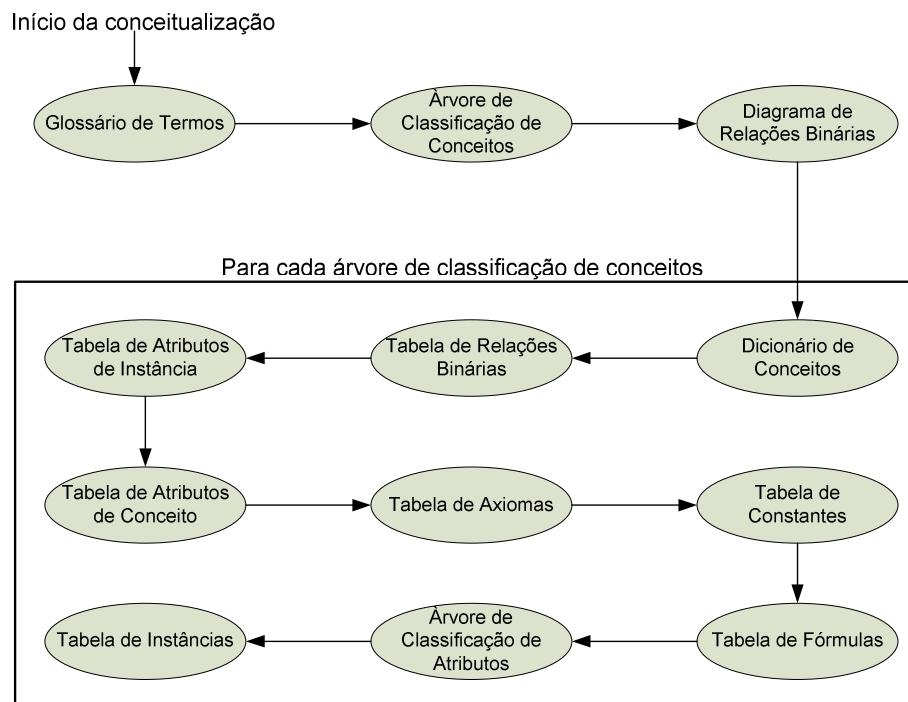


Figura 8 – Artefatos produzidos pela METHONTOLOGY na Conceitualização

O Quadro 3 apresenta os artefatos produzidos pelas metodologias na atividade de conceitualização.

Quadro 3 - Artefatos produzidos pelas Metodologias de Desenvolvimento de Ontologias na Conceitualização.

Artefatos produzidos na Conceitualização	
<b>TOVE</b>	Terminologia inicial e axiomas.
<b>Enterprise</b>	Dicionário e glossário de conceitos.
<b>SENSUS</b>	Não específica.
<b>On-To-Knowledge</b>	Não específica.
<b>METHONTOLOGY</b>	Glossário de termos, árvore de classificação de conceitos, dicionário de conceitos, diagrama de relações binárias, tabela de relações binárias, tabela de atributos de instância, tabela de atributos de conceito, tabela de axiomas, tabela de constantes, tabela de fórmulas, árvore de classificação de atributos e tabela de instâncias.

No Quadro 4 identificam-se as técnicas de conceitualização utilizadas pelas metodologias.

Quadro 4 - Técnicas de Conceitualização das Metodologias de Desenvolvimento de Ontologias.

	<b>Técnicas de Conceitualização</b>
<b>TOVE</b>	Definição da terminologia, dos axiomas e formalização das questões de competência.
<i>Enterprise</i>	Identificação inicial de conceitos e relacionamentos, construção de glossários e dicionários.
<b>SENSUS</b>	Processo manual de integração a SENSUS.
<i>On-To-Knowledge</i>	Não específica.
<b>METHONTOLOGY</b>	Criação de glossários e tabelas com elementos da ontologia.

#### 2.2.4 Avaliação de Ontologias

As metodologias para desenvolvimento de ontologias reconhecem a importância da avaliação e preconizam diversas formas de avaliação. A comunidade de engenharia de ontologias reconhece ao menos dois tipos de avaliação (PINTO; MARTINS, 2004): avaliação técnica e do usuário:

- **Técnica:** a estrutura e o conteúdo da ontologia são analisados para verificar se estão de acordo com os requisitos de especificação e se a ontologia foi modelada corretamente de acordo com o entendimento sobre o domínio.
- **Usuário:** a avaliação é feita da perspectiva do usuário, isto é, o mesmo avalia a ontologia pela usabilidade, utilidade e sua documentação.

Quanto ao momento da avaliação, segundo Hartmann et al. (2005), uma ontologia pode ser avaliada na(o):

- **pré-modelagem:** corresponde à atividade de especificação segundo a METHONTOLOGY;
- **modelagem:** corresponde às atividades de conceitualização, formalização e implementação;
- **pós-desenvolvimento/aplicação:** quando a ontologia está efetivamente em uso ou em aplicação. No pós-desenvolvimento, uma ontologia é avaliada, segundo Noy (2004), para selecionar, dentre as ontologias existentes, a mais adequada para determinada tarefa e aplicação ou para fins de reuso.

Ao avaliar e validar e uma ontologia durante seu **desenvolvimento**, que compreende os momentos de pré-modelagem e modelagem na visão de Hartmann, verifica-se o conteúdo da ontologia, isto é, se há a erros lógicos, inconsistências, redundâncias e a completude da ontologia de acordo com as questões de competência e cenários elaborados na especificação. Como esta tese aborda primordialmente a conceitualização, interessa particularmente descrever a avaliação durante o desenvolvimento conforme ilustra a Figura 9.

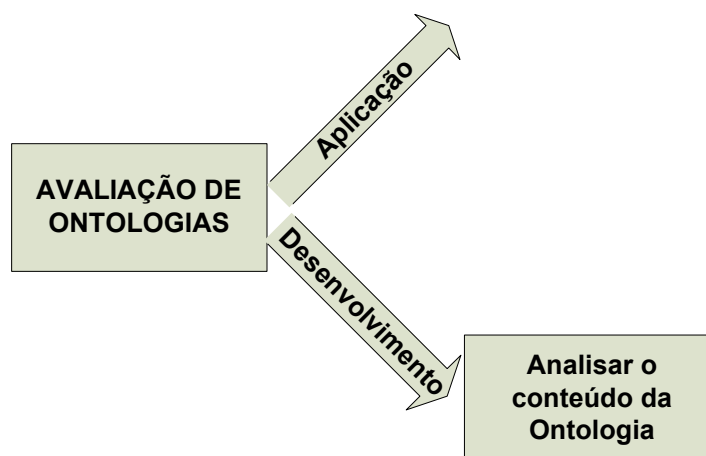


Figura 9 – Avaliação da Ontologia pela Análise do Conteúdo.

Hartman et al. (2005) propõe avaliar uma ontologia, pelo conteúdo segundo três abordagens: baseada em evolução, baseadas em regras, baseadas em métricas. A Figura 10 ilustra o posicionamento destas abordagens.

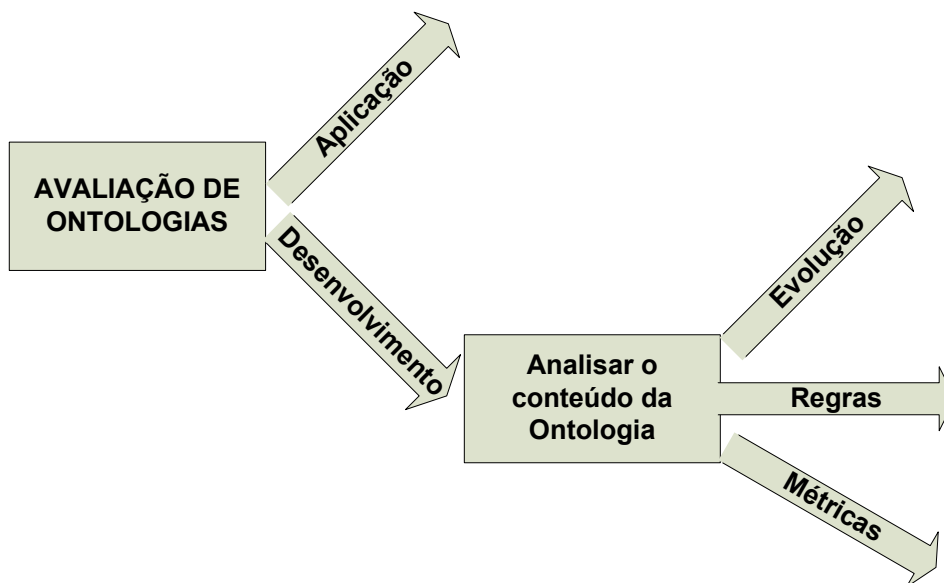


Figura 10 – Avaliação do conteúdo de ontologias baseada em evolução, regras e métricas.

Na avaliação baseada na evolução (PLESSER, 2005; HAASE, 2006), verifica-se se há alterações inválidas em diferentes versões de uma ontologia, ou seja, que não estão de acordo com as alterações no domínio e se há inconsistências ao adicionar novos elementos na ontologia.

A avaliação baseada em regras (ARPINAR, 2006), utiliza regras construídas *a priori* pelos projetistas para detectar conflitos (*e.g.* propriedades, conceitos disjuntos, taxonomia) entre as regras e o que existe na ontologia.

Na avaliação baseada em métricas (SUPEKAR, 2004; LOZANO-TELLO, 2004; ALANI, 2006; GANGEMI et al., 2006; TARTIR et al., 2005; TARTIR; BUDAK ARPINAR, 2007) uma ontologia é avaliada sob a perspectiva quantitativa da qualidade da ontologia. Nesta abordagem, obtêm-se diferentes medidas sobre o conteúdo da ontologia. Considera-se, por exemplo, a localização das classes na estrutura taxonômica, a descrição dos classes por propriedades, a população da ontologia (distribuição das instâncias nas classes), entre outras características que dão indícios sobre o conhecimento presente na ontologia.

Gangemi et al. (2006) categorizou a avaliação das ontologias em três dimensões:

1. **estrutural:** a ontologia é vista como um grafo, assim a sintaxe e a semântica do formalismo da ontologia são analisadas de forma independente do contexto de aplicação;

2. **funcional**: avalia-se a ontologia em relação ao uso pretendido, i.e., sua função num contexto. Nesta dimensão, o foco está na conceitualização especificada na ontologia;
3. **usabilidade e anotações**: avalia o nível de anotações, i.e., de informações disponíveis acerca da própria ontologia para que possa ser facilmente utilizada pelos usuários.

A Figura 11, complementa a linha de raciocínio da Figura 10 pela inclusão da análise estrutural e funcional ao utilizar métricas que analisam o conteúdo da ontologia. A dimensão *usabilidade e anotações* foi descartada por englobar medidas utilizadas primordialmente no pós-desenvolvimento (embora deva haver uma preocupação já no desenvolvimento com estas características no desenvolvimento).

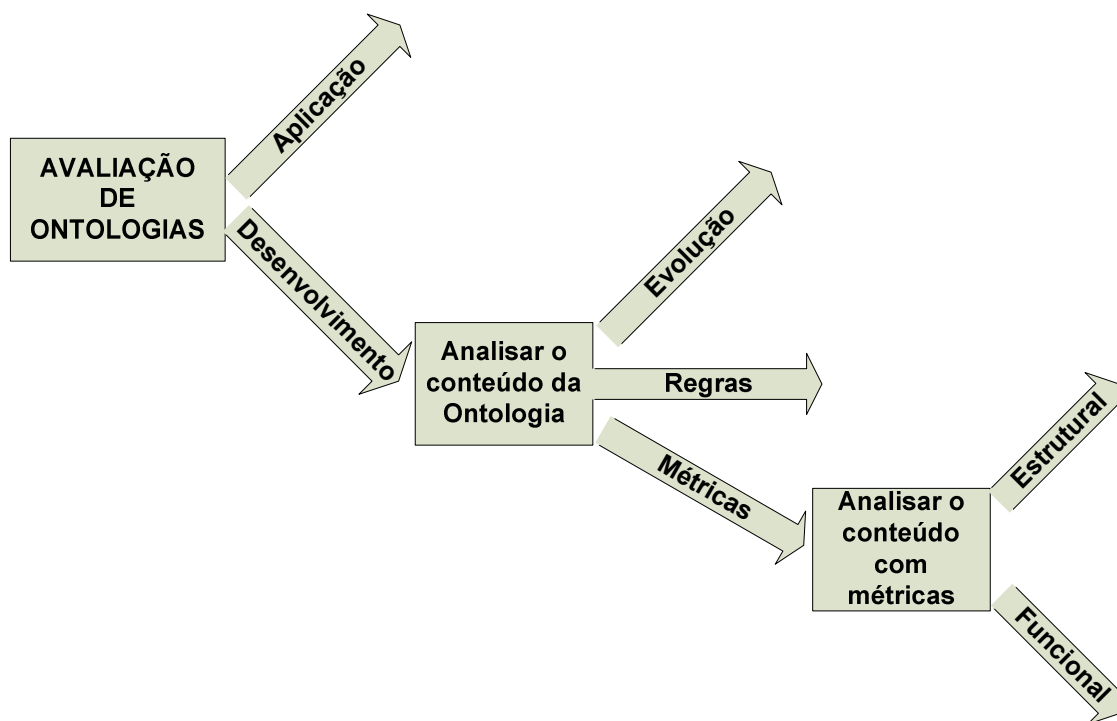


Figura 11 – Avaliação de Ontologias pela Análise do Conteúdo Estrutural ou Funcional por utilizar Métricas.

A Figura 12 ilustra algumas técnicas utilizadas tanto na avaliação estrutural quanto na funcional. Ainda não apresentada, a técnica de comparação a uma fonte ou ontologia padrão, também chamada de *golden standard* é utilizada nos trabalhos de Maedche e Staab (2002) e Dellschaft e Staab (2006). Estes autores comparam a ontologia desenvolvida com um



padrão existente ou com fontes de informação com base em métricas de recuperação de informação, tais como, *precision* e *recall*, ou métricas sintáticas (e.g. *edit distance*) ou semântico-estruturais (e.g. *semantic cotopy*). *Precision* e *recall* são medidas comumente utilizadas na área de recuperação de informação. *Precision* calcula o valor das correspondências encontradas (verdadeiros positivos) sobre o total de correspondências retornadas (verdadeiros positivos e falsos positivos). *Recall* calcula os corretamente encontrados (verdadeiros positivos) sobre o número de correspondências esperadas (verdadeiros positivos e verdadeiros negativos).

Dentre as mostradas na Figura 12, utilizam-se nesta tese aquelas ligadas às noções filosóficas dos elementos ontológicos (metodologia *OntoClean*) e estruturais (que analisam o grafo da ontologia independente do contexto) baseando-se nas métricas provenientes da *OntoQA*.

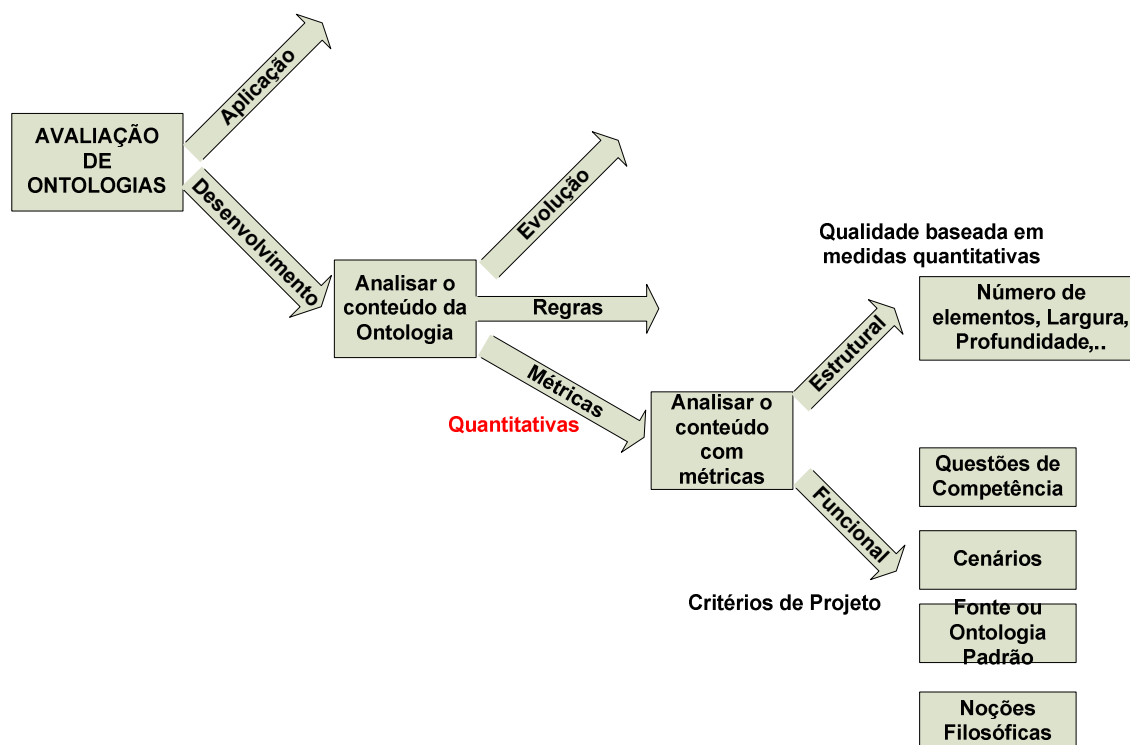


Figura 12 – Formas de Avaliação de uma Ontologia e posicionamento de algumas Abordagens.

### 2.2.4.1 OntoQA

O *OntoQA* avalia a qualidade das ontologias desenvolvidas através de um conjunto de métricas estruturais. Dentre as métricas de *OntoQA*, as seguintes são utilizadas na avaliação dos experimentos que analisam o método proposto no capítulo 4: *attribute richness*, *relationship diversity*, *schema deepness*, *class utilization* e *average population*. Estas medidas podem ser categorizadas em *schema metrics* (projeto da ontologia e seu potencial para representar o conhecimento) e *instance metrics* (avalia o posicionamento das instâncias na ontologia e a capacidade de representar o conhecimento modelado). A *schema metrics* inclui as medidas *attribute richness*, *relationship diversity* e *schema deepness*. As métricas *average population* e *class utilization* fazem parte da *instance metrics*.

A métrica *attribute richness* indica a qualidade da ontologia relacionada à quantidade de informação. Quanto mais atributos por conceitos, mais conhecimento está representado na ontologia. *Attribute richness* é definida pela fórmula:

$$\text{Attribute Richness} = \frac{|Att|}{C}, \text{ sendo:}$$

- *Att*: número de atributos de todos os conceitos;
- *C*: número de conceitos.

O resultado será um número real que representa o número médio de atributos por conceito. Se uma ontologia A apresenta um valor *attribute richness* maior do que uma ontologia B, os conceitos de A tem em média mais atributos do que os conceitos de B, o que pode ser interpretado como A é mais descritiva no que se refere aos conceitos.

A métrica *relationship diversity* reflete a diversidade de relações em uma ontologia. Uma ontologia que contém relações taxonômicas e de propriedades é mais descritiva do que uma que contém apenas relações taxonômicas. *Relationship diversity* é definida pelo número de propriedades dividido pela soma do número de relações taxonômicas mais o número de propriedades. *Relationship diversity* é definida pela fórmula:

$$\text{Relationship diversity} = \frac{|P|}{|H| + |P|}, \text{ sendo:}$$

- *P*: número de propriedades;
- *H*: número de relações taxonômicas.

A métrica *schema deepness* descreve a distribuição de conceitos entre os diferentes níveis de uma ontologia, ou seja, número de relações taxonômicas por conceito. Nesta medida, quanto mais alto o valor, mais conceitos relacionados por relações taxonômicas, e quanto menor, maior o número de conceitos que formam outras hierarquias e conectam-se através de propriedades. *Schema deepness* é definida pela fórmula:

$$\text{Schema Deepness} = \frac{|H|}{|C|}, \text{ sendo:}$$

- *H*: número de relações taxonômicas;
- *C*: número de conceitos.

A métrica *average population* é o número médio de instâncias por conceito. *Average population* é definida pela fórmula:

$$\text{Average population} = \frac{|I|}{|C|}, \text{ sendo:}$$

- *I*: número de instâncias;
- *C*: número de conceitos.

A métrica *class utilization* verifica como as instâncias são distribuídas entre os conceitos pela divisão do número de conceitos populados pelo número total de conceitos. *Class utilization* é definida pela fórmula:

$$\text{Class Utilization} = \frac{|C'|}{|C|}, \text{ sendo:}$$

- *C*: número de conceitos;
- *C'*: número de conceitos populados.

#### 2.2.4.2 OntoClean

*OntoClean* permite que se façam escolhas de modelagem embasadas em noções filosóficas definidas formalmente, tais como, **essência**, **dependência** e **identidade** (GUARINO; WELTY, 2002). A *OntoClean* define um conjunto de metapropriedades que

permitem ao projetista definir características relevantes do significado pretendido para as propriedades da ontologia.

No trabalho de Guarino e Welty (2004), uma propriedade equivale a um predicado unário na lógica de primeira ordem. Dado um mundo possível, uma classe ou conceito corresponde às instâncias que exibem tal propriedade. As metapropriedades da *OntoClean*, que definem a natureza ontológica das propriedades, são suficientemente gerais para serem usadas em qualquer ontologia, independente de um domínio particular.

Embora, segundo seus autores, a *OntoClean* não deve ser vista como uma forma de auxiliar nas decisões sobre a natureza ontológica de certa propriedade, mas sim, de auxiliar na exploração das consequências ontológicas de se fazer certas escolhas (GUARINO; WELTY, 2004), acredita-se, nesta tese, que esta metodologia pode auxiliar os atores na conceitualização de um domínio (cf. seção 3.2.2).

A *OntoClean* requer um esforço de aprendizagem muito grande, o que acaba por dificultar a sua aplicação. Além disso, *OntoClean* é definida formalmente por um sistema lógico, havendo diferenças nos termos empregados na sua definição e na computação em geral. Por outro lado, a qualidade das ontologias desenvolvidas tende a melhorar com a utilização da *OntoClean*. A título de exemplo, (GANGEMI et al., 2002) relatam que *OntoClean* tem sido utilizada para redefinir (limpar) as classes de nível superior da *WordNet*.

Os parágrafos que seguem são baseados em Guarino e Welty (2004). Nas definições que segem utilizam-se propriedades arbitrárias representadas pelos símbolos  $\langle \phi, \psi \rangle$  e a notação  $\phi^M$  para indicar que a propriedade  $\phi$  possui a metapropriedade  $M$ .

### *Essência e Rigidez*

A metapropriedade de rigidez é baseada na noção de essência. Uma propriedade é dita essencial se o seu conjunto de instâncias é invariável em todos os mundos possíveis. Por exemplo, a propriedade *ter um cérebro* é essencial para todos os seres humanos, porque todos os seres humanos devem *ter um cérebro*. A essência apresenta formas variantes com relação à sua ocorrência, representada pela noção de rigidez (Figura 13): rígida e não rígida. Por sua vez, não rígida subdivide-se em semirrígida e antirrígida.

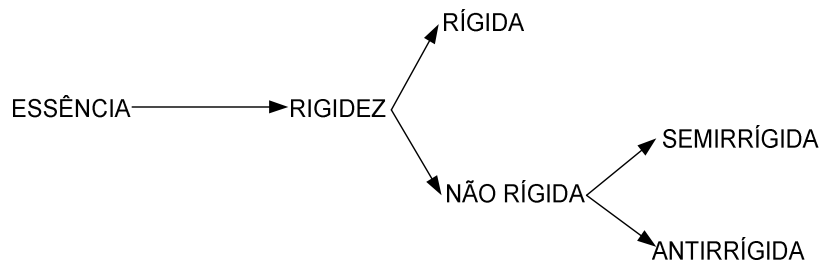


Figura 13 – Essência e Rigidez

Fonte: Guarino e Welty (2002).

As noções de rigidez, não rigidez e antirrigidez foram definidas inicialmente em Guarino, Carrara e Giaretta (1994) como segue:

**Definição 1 (rigidez):** uma propriedade rígida é uma propriedade que é essencial para todas as instâncias, i.e. uma propriedade  $\phi$ , tal que:  $\forall x\phi(x) \rightarrow \phi(x)$ . Representa-se uma propriedade rígida pela notação  $\phi^{+R}$ . O símbolo modal  $\Box$  representa a noção de obrigatoriedade e formaliza a expressão “necessário que...”. Por exemplo, a expressão  $\Box(P(x) \rightarrow Q(x))$  formaliza a expressão “é necessário que  $P(x)$  implique em  $Q(x)$ ”.

**Definição 2 (não rigidez):** uma propriedade não rígida é uma propriedade que não é essencial para algumas das instâncias, i.e.  $\exists x\phi(x) \wedge \neg \phi(x)$ . Representa-se uma propriedade não rígida pela notação  $\phi^{-R}$ .

**Definição 3 (antirrigidez):** uma propriedade é antirrigida quando não é essencial para todas as suas instâncias, i.e.  $\forall x\phi(x) \wedge \neg \phi(x)$ . Representa-se uma propriedade antirrigida pela notação  $\phi^{\sim R}$ .

Uma propriedade é rígida se ela é imprescindível para todas as suas instâncias, ou seja, uma instância de uma propriedade rígida não pode deixar de ser uma instância daquela propriedade em qualquer mundo possível. Na Figura 14 ilustra-se um exemplo de rigidez. Na Figura 14 *ser humano* é a propriedade principal que subsume *professor* e *estudante*. No exemplo, um *estudante* pode deixar de ser *estudante*. O mesmo ocorre com o *professor*. Porém tanto *professor* quanto *estudante* não deixariam de ser *seres humanos*. Logo *ser humano* é uma propriedade rígida.

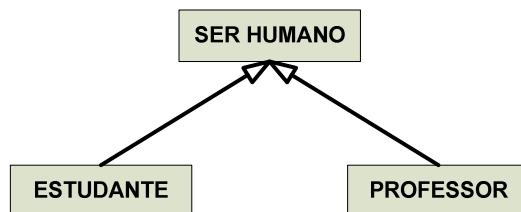


Figura 14 – Exemplo de Hierarquia.

Uma propriedade é não rígida se ela puder ser adquirida ou perdida por algumas de suas instâncias, dependendo da conjuntura considerada. Por exemplo, um *ser humano* pode deixar de ser *estudante*, mas sempre será um *ser humano*. Logo *estudante* é uma propriedade não rígida enquanto *ser humano* é rígida. As propriedades não rígidas podem ser semirrígidas e antirrígidas. Semirrígidas são aquelas essenciais para alguns conceitos e não essenciais para outros. Antirrígidas são aquelas que são não essenciais para todas as suas instâncias (*e.g* todos os *estudantes* podem deixar de serem *estudantes*).

### *Identidade*

A identidade refere-se à capacidade de reconhecer se indivíduos em um mundo são os mesmos ou diferentes. Um conceito com identidade é aquele onde as instâncias podem ser identificadas como sendo as mesmas, a qualquer tempo e em qualquer mundo. Isso significa que o conceito carrega um critério de identidade (*Identity Condition* - IC). Este é utilizado para responder questões como “este é o meu carro” (*e.g* a placa do veículo é um IC).

Um IC subdivide-se em *diachronic* (diacrônica) e *synchronic* (sincrônica). Um IC que é o mesmo ao longo do tempo é diacrônico. Como reconhecer uma pessoa depois de um longo tempo. Como reconhecer se é a mesma pessoa. O critério sincrônico identifica um indivíduo no exato momento.

Um IC para uma propriedade  $\phi$  é definida como uma relação  $\rho$  satisfazendo a seguinte fórmula:  $\phi(x) \wedge \phi(y) \rightarrow (\rho(x, y) \leftrightarrow x = y)$ . Por exemplo,  $Pessoa(x) \wedge Pessoa(y) \rightarrow (impressão\_digital\_idêntica(x, y) \leftrightarrow x = y)$ ,  $x$  e  $y$  serão os mesmos indivíduos se possuem os mesmos valores para a relação *impressão\_digital*.

Tendo em vista que o IC pode ser herdado ao longo da hierarquia de relacionamentos, através da subsunção, outra distinção que deve ser feita é se as propriedades carregam ou fornecem ICs.

Propriedades não rígidas podem somente carregar ICs. Os ICs de propriedades não rígidas são herdados de propriedades rígidas. Por exemplo, *estudante* carrega seus ICs herdando os mesmos de propriedades rígidas como *pessoa*. Uma propriedade  $\phi$  carrega ICs se e somente se é subsumida por uma propriedade  $\psi$  que fornece um IC. Uma propriedade que carrega ICs é identificada como  $\phi^{+I}$  (caso contrário  $\phi^{-I}$ ).

Uma propriedade  $\phi$  fornece ICs se é rígida, há um IC necessário e suficiente para ela (possui sua própria identificação), e o mesmo IC não é carregado por todas as propriedades que subsumem  $\phi$ . Isto significa que se a propriedade  $\phi$  herda diferentes ICs de diferentes propriedades, ela ainda fornece ICs. Por exemplo, instâncias do conceito *pessoa* possuem a impressão digital como um critério de identificação, pois pode-se identificar qualquer *pessoa* pela impressão digital. Instâncias do conceito *estudante* podem herdar de *pessoa* o critério de identificação, porém podem possuir outros critérios como número de matrícula. Uma propriedade que fornece ICs é identificada como  $\phi^{+O}$  (caso contrário  $\phi^{-O}$ ). Uma propriedade que fornecem ICs implica em ser rígida e carregar ICs. Logo  $\phi^{+O}$  é também  $\phi^{+R}$  e  $\phi^{+I}$ .

### *Dependência*

A dependência pode ser intrínseca ou extrínseca. É intrínseca quando algo é inerente a uma instância, como *ter um cérebro* ou *impressão digital* e não depende de outras instâncias. É extrínseca quando instâncias possuem uma natureza relacional. Por exemplo, para exibir a propriedade de *pai* é necessário que uma instância tenha uma relação de paternidade com outra instância. Enfim, um conceito C1 é dependente de outro conceito C2, se para toda a instância em C1 deve existir uma instância em C2. Welty e Guarino (2001) definiram dependência:

**Definição 4 (*dependência*):** Uma propriedade  $\phi$  é externalmente dependente de uma propriedade  $\psi$  se, para todas as instâncias de  $x$ , necessariamente alguma instância de  $\psi$  deve existir a qual não é parte e nem constituinte de  $x$ :

$\forall x (\phi(x) \rightarrow \exists y \psi(y) \wedge \neg P(x, y) \wedge \neg C(x, y))$ , sendo:

- $P(x, y)$ :  $x$  é parte de  $y$ ;
- $C(x, y)$ :  $x$  é constituído de  $x$ .

Por exemplo, *estudante* é dependente de *escola*. Para ser *estudante* é necessário uma relação com *escola*. *Escola* não é parte de *estudante* e, *estudante* não é constituído de *escola*.

### Restrições

As metapropriedades impõem restrições nos relacionamentos de subsunção (implicação), que podem ser usadas para checar as consistências ontológicas das ligações taxonômicas. As seguintes restrições são impostas na metodologia *OntoClean* para as metapropriedades rigidez (*R*), identidade (*I*) e dependência (*D*) (GUARINO; WELTY, 2002):

- $\phi^{\sim R}$  não pode subsumir  $\psi^{+R}$ ;
- $\phi^{+I}$  não pode subsumir  $\psi^{-I}$ ;
- $\phi^{+D}$  não pode subsumir  $\psi^{-D}$ .

Uma propriedade não rígida (*-R*) não pode subsumir uma propriedade rígida (*+R*). Por exemplo, a propriedade *estudante* não pode subsumir *ser humano*, considerando-se que *estudante* é antirrígida e *ser humano* é rígida. Assim, encontra-se uma inconsistência, pois se considera que as entidades de ser *estudante* podem deixar de o ser, enquanto que as entidades de *ser humano* não o podem. Conseqüentemente, se todos os *humanos* são necessariamente *estudantes* (este é o significado da subsunção), então nenhuma pessoa poderia deixar de ser *estudante*. Como *estudante* é antirrígida, todas as propriedades subsumidas pela mesma também devem ser antirrígidas.

Uma propriedade que possui um critério de identidade (*+I*) não pode subsumir uma propriedade que não possui esse critério (*-I*) porque, neste caso, as instâncias desta última deveriam ter identidade. Por exemplo se *ser humano* possui um critério de identidade, então *estudante* que é subsumido por *ser humano* deve possuir o mesmo critério. Se essa regra for ignorada, significa que as instâncias de um conceito subsumido não podem ser identificadas.

Na dependência, uma propriedade dependente (*+D*) não pode subsumir uma propriedade não dependente (*-D*). Por exemplo, *estudante universitário* é um subconceito de *estudante*. Este possui uma relação de dependência com *instituição de ensino*. Logo, não se pode definir um indivíduo como *estudante* que não esteja relacionado com o conceito *instituição de ensino*. Esta dependência que *estudante* possui é herdado para os conceitos que o mesmo subsume.

### Combinação das metapropriedades



Ao definir as características filosóficas das propriedades com a utilização das metapropriedades de identidade, rigidez e dependência, uma propriedade recebe uma classificação que inclui: *Type*, *Quasi-Type*, *Material role*, *Phased sortal*, *Mixin*, *Category*, *Formal Role* e *Attribution*. Estas propriedades podem ser classificadas em *Sortal* quando carregam ICs, caso contrário em *Non-Sortal*.

Ainda considerando o critério de identidade, de acordo com Guarino e Welty (2000) e adaptado de Lowe (1989) faz-se a seguinte definição para uma propriedade:

- *Sortal*: são propriedades que carregam ou fornecem ICs. Seguem o princípio que indivíduos em um domínio possuem um IC, ou seja, não há indivíduos sem identificação (QUINE, 1969). Correspondem a substantivos. *Pessoa*, *estudante*, *automóvel* são exemplos de *sortal*. Este tipo de propriedade possui grande utilidade para a criação de uma taxonomia.
- *Non-Sortal*: são propriedades que não possuem um IC. Correspondem a adjetivo, possuem limitação na organização taxonômica, mas alto valor semântico. *Grande*, *vermelho*, *bonito* são exemplos de *non-sortals*.

A Tabela 1 ilustra todas as possíveis combinações de metapropriedades (*R*, *I*, *D* e *O*). A Figura 15 ilustra a taxonomia das propriedades. A seguir uma breve explicação das classificações pela combinação das metapropriedades (WELTY; GUARINO, 2001):

- *Type*: são propriedades rígidas que possuem seu próprio critério de identificação. São propriedades importantes em uma ontologia porque fornecem ICs a outras propriedades. Por exemplo, o conceito *pessoa* pode ser classificado como um *type* porque possui o IC impressão digital que pode ser herdado pelos conceitos subsumidos por *pessoa*. A impressão digital é um critério de identificação próprio da *pessoa*.
- *Quasi-Type*: são propriedades rígidas e *sortals* que não fornecem ICs. Este tipo de propriedade são sempre subsumidas por pelo menos um *type*. *Animal herbívoro* é um exemplo de *quase-type* porque herda ICs de um *type*, neste caso *animal* (todo animal possui algum IC).
- *Material Role*: são propriedades antirrígidas (cada entidade não necessariamente carrega um IC) e dependentes (dependem de um evento ou relacionamento). Geralmente são subsumidas por *types* (por que carregam ICs). Por exemplo, *estudante* se subsumido por *pessoa* (*type*) herda o IC impressão digital de *pessoa* e possui dependência com *instituição de ensino*.

- *Phased Sortal*: é um único tipo de propriedade que combina a obrigatoriedade de carregar IC com antirrígidez e independência. Elas não fornecem um IC global, fornecem um IC local. Deve ser subsumido por um *type* porque deve-se determinar qual é o indivíduo em diferentes tempos. Por exemplo, uma *vila* pode tornar-se *cidade* depois de um determinado tempo.
- *Mixins*: são propriedades que carregam a identidade e não rígidas. Geralmente utilizadas para organizar informações em uma maneira não rígida.
- *Categories*: propriedades rígidas que não carregam IC. Estas propriedades não podem ser subsumidas por *sortals* e são geralmente o mais alto nível da taxonomia em rigidez na ontologia.
- *Formal Role*: são propriedades antirrígidas e dependentes. Estas propriedades não tem IC e não são subsumidas por *sortals*.
- *Attribution*: são propriedades *non-sortal*, não rígidas que representam valores de atributos como cor, gênero.

Tabela 1 – Possíveis Classificações das Propriedades  
Fonte: Guarino e Welty (2000).

+O	+I	+R	+D	Type	Sortal	
			-D			
-O	+I	+R	+D	Quasi-type		
			-D			
-O	+I	~R	+D	Material role		
-O	+I	~R	-D	Phased Sortal		
-O	+I	~R	+D	Mixin		
			-D			
-O	-I	+R	+D	Category		Non-sortal
			-D			
-O	-I	~R	+D	Formal role		
-O	-I	~R	-D	Attribution		
		-R	+D			
		-D				

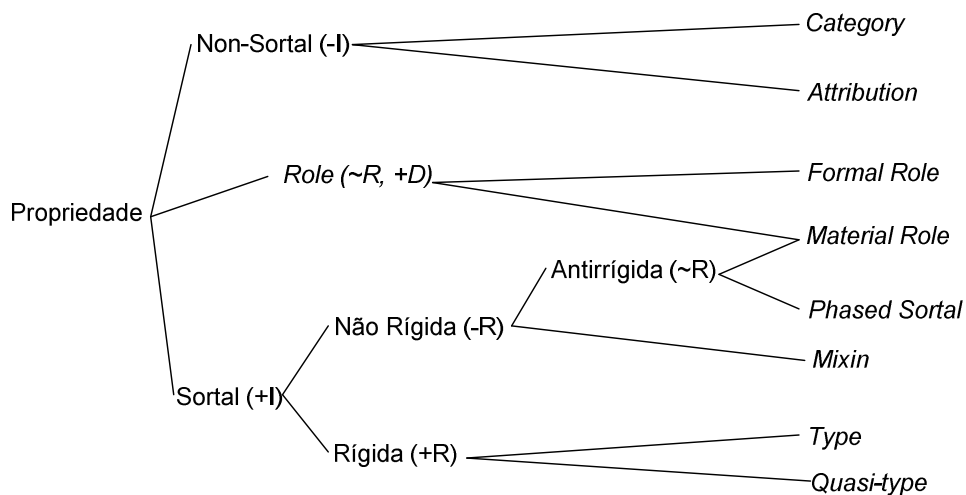


Figura 15 – Taxonomia das Propriedades

Fonte: Guarino, Welty (2000).

A combinação de metapropriedades além de gerar uma classificação para o conceito, também possibilita a criação da taxonomia da ontologia. *OntoClean* idealiza uma taxonomia que pode ser visualizada na Figura 16. Nesta figura percebe-se que *Sortals* podem ser subsumidos por *Non-Sortals*. Ao contrário não é possível devido ao fato que *Sortals* possuem critérios de identificação que devem ser subsumidos por seus descendentes.

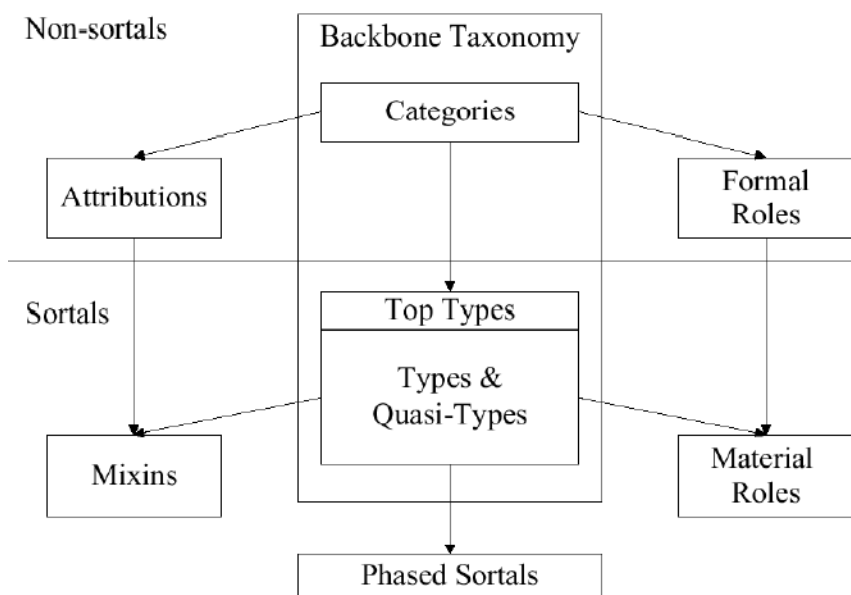


Figura 16 – Taxonomia idealizada pela *OntoClean*

Fonte: Guarino e Welty (2001).

### 2.2.5 Metodologias para Desenvolvimento Colaborativo

A DILIGENT (PINTO; TEMPICH; STAAB, 2004) é uma metodologia para desenvolvimento colaborativo de ontologias destinada a especialistas do domínio, usuários e projetistas. Basicamente, a DILIGENT fornece um ciclo de vida baseada na evolução de protótipos com as seguintes atividades (Figura 17):

- **Desenvolvimento:** atividade onde especialistas do domínio, usuários e engenheiros do conhecimento desenvolvem uma versão inicial da ontologia. A equipe deve ser relativamente pequena a fim de desenvolver a primeira versão em consenso. Nesta primeira versão não considera-se a especificação da ontologia, reuso e reengenharia de recursos de conhecimento existentes. Esta primeira versão é uma versão compartilhada entre todos os participantes do projeto.
- **Adaptação local:** uma vez que a primeira versão foi desenvolvida cada participante faz adaptações à ontologia conforme suas necessidades. Todo o participante, a partir da versão inicial e compartilhada, trabalha individualmente em uma nova versão chamada de local. Esta não afeta a versão compartilhada.
- **Análise:** um controle central analisa todas as ontologias locais, ou seja, todas as alterações em relação à versão compartilhada. Nesta atividade decide-se quais alterações locais serão realizadas na próxima versão compartilhada da ontologia por identificar similaridades entre as ontologias.
- **Revisão:** baseada na etapa anterior, uma nova ontologia compartilhada é criada e distribuída. Isto é para evitar grandes divergências entre as versões locais e as versões compartilhadas.
- **Atualização local:** Nesta atividade, das ontologias compartilhadas, os participantes atualizam suas ontologias locais. A cópia compartilhada pode ter várias alterações que foram introduzidas nas atividades, mas nem todas as alterações locais sugeridas pelo participante em sua ontologia foram introduzidas. Este pode ou não aceitar os novos conceitos criados por outros usuários.

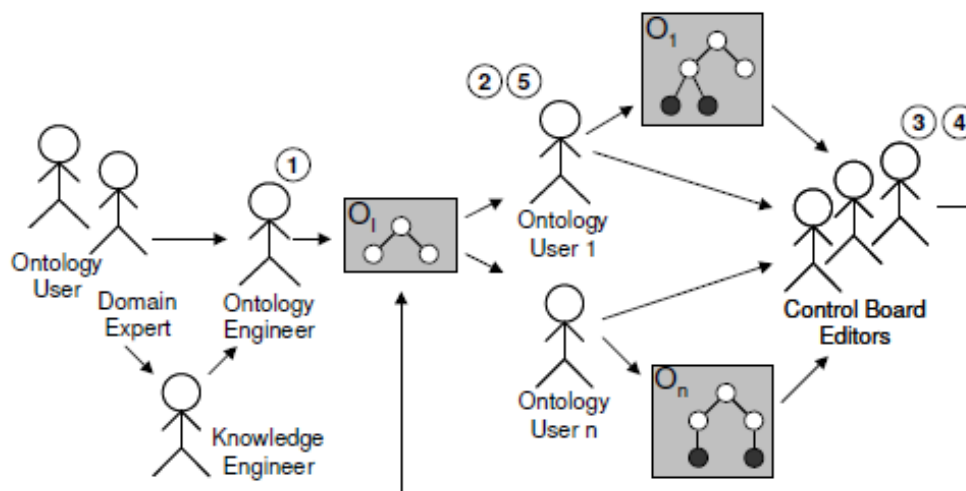


Figura 17 – Metodologia DILIGENT

Fonte: Pinto, Tempich, Staab (2004).

A metodologia NEON (SUÁREZ-FIGUEROA et al., 2008) é uma metodologia para desenvolvimento colaborativo de ontologias baseada em 9 cenários individuais ou combinados (Figura 18):

1. não utilização de recursos de conhecimento;
2. não utilização de recursos ontológicos;
3. reuso de recursos ontológicos;
4. reuso e reengenharia de recursos ontológicos;
5. reuso e combinação de recursos ontológicos;
6. reuso, combinação e reengenharia de recursos ontológicos;
7. reuso de padrões de projeto de ontologias (*i.e ontology design patterns*);
8. reestruturação de recursos ontológicos;
9. localização de recursos ontológicos;

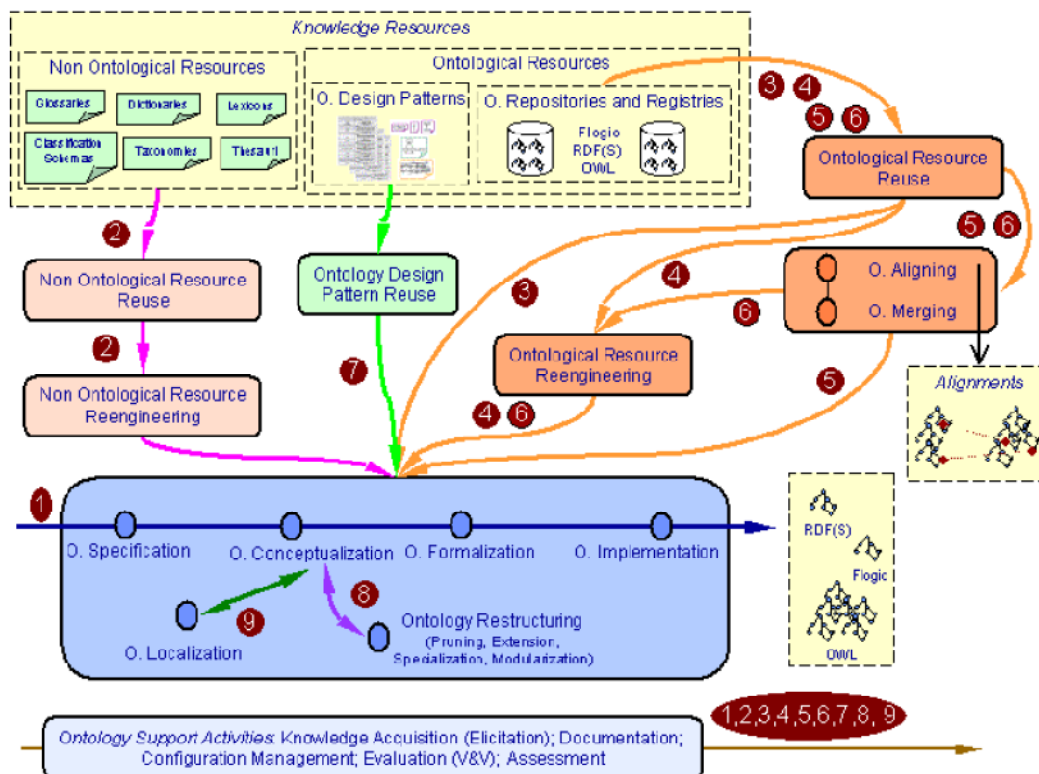


Figura 18 – Metodologia NEON

Fonte: Suárez-Figueroa et al. (2008).

A NEON, além do desenvolvimento colaborativo, trabalha em outros aspectos como a evolução dinâmica da ontologia em ambientes distribuídos, inclusão de informações contextuais no desenvolvimento por especialistas do domínio e projetistas, utilização de outras metodologias, ciclos de vida, métodos, técnicas e ferramentas existentes para o desenvolvimento de ontologias. NEON procura trabalhar em pontos não trabalhados por outras metodologias como dinamicidade (evolução das ontologias), contexto, colaboração e desenvolvimento baseado em padrões de projeto (*ontology design patterns*). Este último é classificado pela NEON como desenvolvimento de ontologias pelo reuso de conhecimentos ontológicos e não ontológicos.

Conhecimentos não ontológicos compreendem artefatos utilizados na aquisição de conhecimentos, padrões linguísticos, mais precisamente, conhecimentos que descrevem o domínio, tais como, taxonomias, glossários e dicionários. O desenvolvimento a partir do reuso é descrito na próxima seção.

## **2.2.6 Desenvolvimento de Ontologias pelo Reuso**

Um dos maiores desafios no desenvolvimento de ontologias é o reuso (PRESUTTI; GANGEMI, 2008). As possíveis razões são tamanho e complexidade das ontologias reutilizáveis, falta de critérios na utilização de recursos de conhecimento e a falta de ferramentas para auxiliar projetistas que possuem dificuldades em se apropriar de ontologias existentes e de revisar conhecimentos representados nestas ontologias. O desenvolvimento de ontologias pelo reuso inclui a utilização de conhecimentos não ontológicos e ontológicos.

Como exemplo de utilização de conhecimentos não ontológicos, a metodologia NEON utiliza glossários, taxonomias, dicionários, entre outros. A METHONTOLOGY utiliza textos que representam o domínio, relatos de entrevistas e tabelas de domínio construídas por especialistas do domínio e projetistas.

A utilização de conhecimento ontológico ocorre pelo reuso de outras ontologias (SIMPERL, 2009; LONSDALE et al., 2010; GÓMEZ-PÉREZ; ROJAS-AMAYA, 1999), por exemplo, especializar uma ontologia de nível superior (GRÜNINGER; FOX, 1995; SWARTOUT et al., 1997; USCHOLD; KING, 1995; USCHOLD, 1996; LENAT; GUHA, 1990) ou pela utilização de padrões ontológicos (GUIZZARDI, 2005).

No desenvolvimento a partir de outras ontologias utilizam-se métodos de integração de ontologias como alinhamento e combinação. Neste ponto a METHONTOLOGY possui uma atividade de integração da ontologias categorizada como suporte. Esta atividade possui o objetivo de integrar a ontologia desenvolvida com outras ontologias existentes. Algumas metodologias (GRÜNINGER; FOX, 1995; SWARTOUT et al., 1997; USCHOLD; KING, 1995; USCHOLD, 1996; LENAT; GUHA, 1990) preconizam o desenvolvimento de ontologias a partir das ontologias de nível superior. A NEON, preconiza a utilização de padrões ontológicos. Na próxima seção descreve-se brevemente o desenvolvimento a partir de padrões.

### **2.2.6.1 Desenvolvimento de Ontologias a partir de Padrões**

Na engenharia de conhecimentos, a modelagem a partir de padrões refere-se à soluções que permitem resolver problemas recorrentes em modelagem ou problemas no projeto das ontologias (CLARK; THOMPSON; PORTER, 2000; PRESSUTI; GANGEMI, 2008, PRESUTTI et al., 2008; RECH, FELDMANN; RAS, 2009). Diferentes tipos de

desenvolvimento de ontologias a partir de padrões são definidos em (PRESSUTTI et al., 2008, RECH, FELDMANN; RAS, 2009):

- **Padrões Lógicos** (*Logical Ontology Design Patterns*): independentes de um domínio, mas dependentes do formalismo lógico utilizado para representação.
- **Padrões Arquiteturais** (*Architectural Ontology Design Patterns*): fornece recomendações sobre a estrutura da ontologia. São definidos em termos de outras ontologias ou da composição das mesmas. Como exemplo, tem-se a utilização de outras ontologias leves ou taxonomias.
- **Padrões de conteúdo** (*Content Ontology Design Patterns*): fornece um modelo conceitual dependente do domínio para resolver problemas relacionados ao conteúdo da ontologia. Exemplificam situações do domínio. Este tipo de padrão é utilizado na maior parte dos trabalhos para o desenvolvimento de ontologias a partir de padrões.

Na utilização de ontologias de nível superior, procura-se estender a ontologia de domínio pela utilização de conceitos e relações existentes nas ontologias de nível superior. Neste caso existe a necessidade de selecionar a área da ontologia de nível superior a ser utilizada.

Na utilização de outras ontologias, estas representam padrões projetados por projetistas. Como exemplo, OntoCase (BLOMQVIST, 2005) os ontologias projetam padrões ontológicos que consideram um padrão base composto por primitivas ontológicas. Estas primitivas são representadas em uma ontologia. O OntoCase extrai conceitos e relações entre os mesmos utilizando técnicas de aprendizado de ontologias. Então, compara o recuperado no método de aprendizado com as primitivas ontológicas e seleciona as primitivas mais apropriadas. Com esta seleção identifica elementos e relações que podem ser formalizados na ontologia.

A OntoUML (GUIZZARDI (2005) é uma linguagem de modelagem conceitual que enriquece a semântica de diagramas de classes UML com distinções ontológicas da UFO (*Unified Foundational Ontology*). A UFO foi desenvolvida com base em teorias, entre elas, Ontologias Formais, Lógica Filosófica, Filosofia da Linguagem, Linguística e Psicologia Cognitiva (GUIZZARDI et al. 2008b).

Na OntoUML é proposto um conjunto de diretivas metodológicas para a criação de ontologias utilizando a linguagem chamada OntoUML. O processo na OntoUML repara o metamodelo em UML com a estrutura da ontologia de fundamentação UFO e transfere



axiomas do metamodelo da linguagem, por meio de restrições formais. Basicamente a OntoUML incorpora um conjunto de padrões de modelagem de ontologias a fim de resolver problemas de modelagem, entre eles: modelagem de papéis (GUIZZARDI et al., 2004), transitividade da relação todo/parte (GUIZZARDI, 2008) e restrições de cardinalidade (Guizzardi & Wagner, 2008).

### 2.2.7 Comparação das Metodologias

Ao analisar o estado da arte das metodologias existentes para desenvolvimento de ontologias, METHONTOLOGY e *On-To-Knowledge* apresentam-se como as metodologias mais completas para o desenvolvimento de ontologias, por suportarem atividades que incluem desde a especificação até a implementação e posterior manutenção da ontologia.

Percebe-se que as metodologias possuem atividades semelhantes, apenas com alterações da nomenclatura, com relação às atividades de aquisição, conceitualização e formalização da ontologia. Algumas metodologias consideram a aquisição de conhecimentos e conceitualização como uma única atividade.

A maioria das metodologias coloca mais ênfase nas atividades de desenvolvimento, desconsiderando as atividades relacionadas ao gerenciamento e suporte, principalmente no que se refere à evolução e avaliação das ontologias. A *On-To-Knowledge* considera a evolução, mas não especifica como e quando a ontologia deve evoluir, bem como o controle de versões deve proceder. Com relação a avaliação, as metodologias reconhecem a importância desta atividade, porém utilizam o mesmo procedimento, ou seja, confronta a ontologia com as questões de competência e/ou utilizam cenários. Logo estas metodologias estão mais preocupadas com a cobertura da ontologia (um dos aspectos funcionais) e deixam de lado aspectos relacionados a corretude e conteúdo da ontologia, como o desenvolvimento de taxonomias limpas preconizado pela *OntoClean*.

Algumas metodologias são mais testadas que outras. A maioria dos grupos de pesquisa cria seu próprio método de desenvolvimento, dependendo das características da aplicação que pretendem desenvolver. De acordo com (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JURISTO, 1997) é extremamente difícil avaliar qual a melhor metodologia. Experimentação é a forma mais adequada, mas há condições (disponibilidade de tempo dos especialistas, especialistas com visões contraditórias) que não podem ser controladas. Algumas metodologias seguem um modelo de ciclo de vida, outras não. Neste sentido, a

METHONTOLOGY é mais completa. A METHONTOLOGY fornece detalhes de como proceder na condução de uma atividade. As outras consideram que o projetista já domina o assunto de desenvolvimento de uma ontologia.

Metodologias como a CYC, SENSUS e Enterprise preconizam o desenvolvimento de ontologias a partir de uma ontologia de nível superior, porém o processo de conectar novos elementos a esta ontologia, bem como, quais partes da ontologia de nível superior a ser utilizada é realizado por um processo manual.

As metodologias de desenvolvimento são descritas nesta tese com base nas atividades da METHONTOLOGY. Esta foi utilizada com referência nesta tese pelo fato de englobar as atividades de aquisição, conceitualização e avaliação apresentadas no método desenvolvido. Além disso, a METHONTOLOGY desenvolve uma ontologia por sucessivas iterações/ciclos. A cada nova iteração, uma nova ontologia é desenvolvida. Esta ontologia pode ser considerada uma ontologia raiz no início de cada ciclo. A medida que novos conhecimentos são adquiridos e conceitualizados, uma nova ontologia é produzida. Esta torna-se uma ontologia raiz na nova iteração.

As metodologias METHONTOLOGY, *On-To-Knowledge* e DILIGENT possuem uma limitação que é a falta de diretrizes para contextualizar uma ontologia existente (como conectar uma ontologia com novas versões, a constante evolução da ontologia durante seu ciclo de vida). Outro problema apontado é como utilizar (diretrizes) recursos (conhecimentos) existentes (resultantes de algum processo consensual) em um domínio particular.

A principal contribuição da METHONTOLOGY é a identificação de um processo de desenvolvimento, identificação do ciclo de vida e diretrizes de como construir uma nova ontologia. A METHONTOLOGY também ressalta a avaliação de ontologias durante a atividade de conceitualização a fim de evitar a propagação de erros na ontologia em atividades seqüentes no ciclo de vida.

A DILIGENT identifica diferentes papéis (usuário, especialista do domínio, engenheiros do conhecimento) para o desenvolvedor de ontologias. Sugere que o desenvolvedor utilize atividades presentes em metodologias para o desenvolvimento centralizado de ontologias. Cada desenvolvedor utiliza métodos e técnicas que melhor convém.

METHONTOLOGY E *On-To-Knowledge* não tratam o desenvolvimento colaborativo de ontologias que é abordado nas metodologias DILIGENT e NEON. A metodologia NEON além de tratar o desenvolvimento colaborativo ataca outros fatores no

desenvolvimento, entre eles: colaboração, contexto e dinâmica. Segundo Suárez-Figueroa et al. (2008) este é o gargalo das metodologias. A NEON procura expandir as atividades da METHONTOLOGY melhorando e estendendo a mesma para trabalhar como colaboração, contexto, dinâmica e reuso. Este último compreende o desenvolvimento de conhecimentos ontológicos e não ontológicos.

As metodologias, com exceção da *On-To-Knowledge* não propõem um processo de aprendizado de ontologias com o objetivo de reduzir esforços na aquisição e conceitualização do conhecimento. A *On-To-Knowledge* propõe um modo de enriquecer uma ontologia em construção por um processo de aprendizado. Porém não especifica como deve ser este processo.

Perante os benefícios do uso de ontologias é importante fornecer suporte metodológico seja no desenvolvimento centralizado ou colaborativo de ontologias. Há metodologias (METHONTOLOGY, *On-To-Knowledge*, DILIGENT), porém estas abordagens são falhas em fornecer diretrizes para o desenvolvimento de ontologias em constante evolução. Neste ponto a NEON, procura trabalhar o desenvolvimento colaborativo juntamente com a evolução da ontologia. Porém, todas as novas metodologias (DILIGENT, NEON) colaborativas procuram adotar atividades especificadas em outras metodologias de desenvolvimento individual, bem como preconizam a utilização de métodos e técnicas existentes na literatura.

Os Quadros (Quadro 5, Quadro 6, Quadro 7) descrevem as metodologias para o desenvolvimento individual comparando as mesmas com a METHONTOLOGY nas respectivas atividades de gerenciamento, suporte e desenvolvimento. Nas tabelas descreve-se “ausente” quando a metodologia não possui atividade semelhante à METHONTOLOGY.

Quadro 5– Quadro Comparativo das Metodologias com base nas Atividades de **Gerenciamento** da METHONTOLOGY.

Atividades da METHONTOLOGY	Metodologias			
	TOVE	<i>Enterprise</i>	SENSUS	<i>On-To-Knowledge</i>
<b>Planejamento</b>	Ausente	Ausente	Ausente	Não há uma etapa específica de planejamento, porém o método indica o que deve ser feito em cada etapa e quando deve ser feito. Logo, preocupa-se com as três atividades (planejamento, controle e qualidade).
<b>Controle</b>	Ausente	Ausente	Ausente	
<b>Qualidade</b>	Ausente	Ausente	Ausente	

Quadro 6 - Quadro Comparativo das Metodologias com base nas Atividades de **Desenvolvimento** da METHONTOLOGY.

Atividades da METHONTOLOGY	Metodologias			
	TOVE	<i>Enterprise</i>	SENSUS	<i>On-To-Knowledge</i>
<b>Especificação</b>	Identifica possíveis problemas que demandam uma nova ontologia, captura cenários motivadores e formula questões informais de competência.	Criação de um documento com a identificação do propósito da ontologia e de termos relevantes.	Propõe, mas não descreve detalhes da atividade.	Realizada na fase chamada de início da ontologia. Antes de realizar esta fase, propõe um estudo de viabilidade.
<b>Conceitualização</b>	Define a terminologia da ontologia, especifica axiomas (condições necessárias e condições necessárias e suficientes), formaliza as questões de competência. Utiliza nesta atividade os cenários e as questões de competência. Propõe ao engenheiro responder as questões de competência para identificação e conceitualização dos termos da ontologia. Confunde-se esta atividade com aquisição por não definir uma atividade específica para aquisição de conhecimentos.	Não existe uma atividade propriamente para a conceitualização. Realizada na atividade de desenvolvimento, mais precisamente na subatividade de captura. Esta subatividade também contempla a aquisição de conhecimentos.  Identifica conceitos principais e relacionamentos entre os mesmos. Propõe a utilização de abordagens como <i>top-down</i> , <i>bottom-up</i> e <i>middle-out</i> para produzir um modelo conceitual. Utiliza dicionários e glossários para identificar diferentes significados para os conceitos.	Apenas relata que existe uma fase de identificação de conceitos. Esta atividade confunde-se com aquisição e formalização.	Não existe uma fase específica de conceitualização. Esta fase é realizada na atividade de refinamento, mais especificamente na subatividade de elicitación de conhecimento.
<b>Formalização</b>	Propõe formalizar a terminologia em FOL.	Não especifica como formalizar. Preocupa-se mais em identificar os elementos e codificá-los em uma linguagem.	Realizada na atividade de identificação de conceitos. Conectada a aquisição e conceitualização.	Realizada na fase de refinamento na subatividade de elicitación de conhecimento. Ressalta a importância de criação de representações para os elementos até a implementação. Na fase de refinamento existe uma subatividade de formalização, porém esta é mais uma

				implementação em alguma linguagem do modelo criado na fase anterior do que uma formalização.
<b>Implementação</b>	Não há uma fase específica de implementação. Não especifica nenhuma linguagem de representação. Apenas a formalização em FOL.	Propõe a utilização de alguma linguagem como Prolog e Ontolingua. Esta atividade está inclusa na atividade codificação que é uma subatividade na atividade de desenvolvimento.	Relata a utilização de ferramentas para implementação.	Realizada na atividade de refinamento, mais especificamente na subatividade chamada de formalização
<b>Manutenção</b>	Ausente	Ausente	Ausente	Propõe esta atividade, mas não descreve em detalhes como é realizada a manutenção.

Quadro 7 - Quadro Comparativo das Metodologias com base nas Atividades de **Suporte** da METHONTOLOGY.

Atividades da METHONTOLOGY	Metodologias			
	TOVE	<i>Enterprise</i>	SENSUS	<i>On-To-Knowledge</i>
<b>Aquisição</b>	Não define uma atividade específica para aquisição de conhecimentos. Os cenários e questões de competência ajudam na aquisição de conhecimento. Confunde-se esta atividade com conceitualização e formalização.	Iniciada na fase da identificação do propósito da ontologia, e continuada na subatividade de captura. Confunde-se com conceitualização e formalização. Utiliza <i>brainstorming</i> e agrupamento de conceitos. Aquisição de conhecimentos e conceitualização são executadas em conjunto.	Apenas descreve que existe uma fase de identificação de conceitos. Esta atividade é realizada em conjunto com a conceitualização e formalização.	Realizada na fase de refinamento. Ressalta a importância de um especialista do domínio e do desenvolvimento de modelos de representação da ontologia para interação com o especialista do domínio.
<b>Integração</b>	Propõe a integração a ontologias de núcleo comum.	Presente na subatividade de integração. Esta subatividade faz parte da atividade desenvolver. Na integração propõe a verificação do reuso de ontologias. Não define detalhadamente esta atividade.	Presente na conexão dos novos conceitos a ontologia SENSUS. Processo manual de integração e poda das subárvores da ontologia.	Propõe a integração para o reuso de ontologias, mas não descreve este processo detalhadamente. Não existe uma atividade específica para a integração.
<b>Avaliação</b>	Verificação da completude da ontologia de acordo com as questões de competência e cenários elaborados na especificação.	Realizada na atividade de avaliação com base no documento elaborado na identificação do propósito da ontologia. Neste documento há questões de competência. Utiliza também os critérios de Gómez-Pérez (1995) para avaliar a ontologia.	Ausente	Há uma fase específica de avaliação. Nesta fase, a ontologia é avaliada conforme requisitos e questões de competência elaboradas na fase início da ontologia. Esta atividade verifica se a ontologia satisfaz os requerimentos e questões de competência e também a utilização da mesma no ambiente de aplicação.
<b>Documentação</b>	Presente no documento de especificação. Este documento possui a descrição das questões de competência e dos cenários.	Presente no documento de especificação na descrição do processo e dos elementos da ontologia.	Ausente	Documento de especificação elaborado no início da ontologia. Não existe uma fase específica de documentação.
<b>Configuração</b>	Ausente	Ausente	Ausente	Propõe um controle, pois as atividades de desenvolvimento.

Embora o crescente número de ferramentas em engenharia de ontologias (GROSSO et al., 1999), o desenvolvimento manual de ontologias é uma tarefa complexa que enfrenta problemas como o gargalo na aquisição de conhecimentos (MAEDCHE; STAAB, 2001). Maedche e Staab (2001) apontam três problemas para o desenvolvimento manual de ontologias: o tempo, a dificuldade e se a ontologia desenvolvida é a adequada/correta.

Neste contexto, o aprendizado de ontologias é uma área emergente que visa ajudar os projetistas a diminuir esforços durante a fase de aquisição de conhecimentos (MAEDCHE; STAAB, 2002; AUSSENAC-GILLES; BIÉBOW; SZULMAN, 2000; KIETZ; MAEDCHE; VOLZ, 2000). Na próxima seção fundamenta-se o aprendizado de ontologias e trabalhos relacionados ao mesmo.

### **2.2.8 Aprendizado de Ontologias**

Aprendizado de ontologias é a aplicação de um conjunto de métodos e técnicas para o desenvolvimento semiautomático de novas ontologias ou para o enriquecimento de ontologias existentes (GÓMEZ-PÉREZ; MANZANO-MACHO, 2003).

Maedche e Staab (2000a) definiram aprendizado de ontologias como a aquisição de um modelo de domínio que consiste em aprender conceitos relevantes a um determinado domínio, suas definições e relações existentes entre os mesmos, a partir de dados, de uma maneira semiautomática.

Cimiano (2006) considera o aprendizado de ontologias com uma engenharia reversa, onde os autores de certos textos ou documentos possuem um modelo conceitual do domínio em mente (Figura 19). Este modelo conceitual molda o conteúdo do texto resultante.



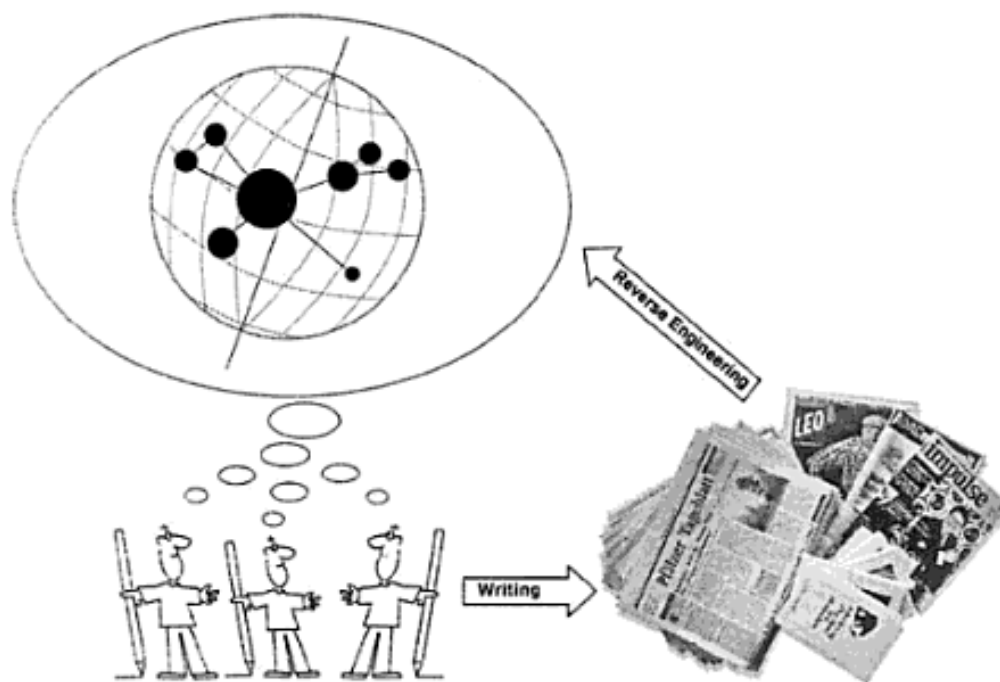


Figura 19 – Engenharia Reversa no Aprendizado de Ontologias

Fonte: Cimiano (2006).

Segundo Cimiano (2006), a tarefa de reconstruir o modelo de domínio de um ou mais autores é um processo de engenharia reversa. Porém, os textos não refletem explicitamente o conhecimento. O mundo está descrito por fatos contidos implicitamente na forma como determinadas palavras ou estruturas linguísticas são utilizadas pelos escritores. Isto se conecta ao triângulo da significação de Ogden e Richards (1923). Segundo Sowa (2000), no triângulo da significação, o vértice esquerdo é o símbolo ou palavra, o superior é o conceito (*i.e* intenção, pensamento, idéia) e o vértice direito o objeto (*i.e* referente, extensão, coisa, realidade). Em toda linguagem formal ou natural há símbolos o qual precisam ser interpretados por conceitos e representam objetos concretos no mundo. Logo, o aprendizado de ontologias trabalha com símbolos ou sinais não interpretados pelo qual o sentido precisa ser identificado por algum processo.

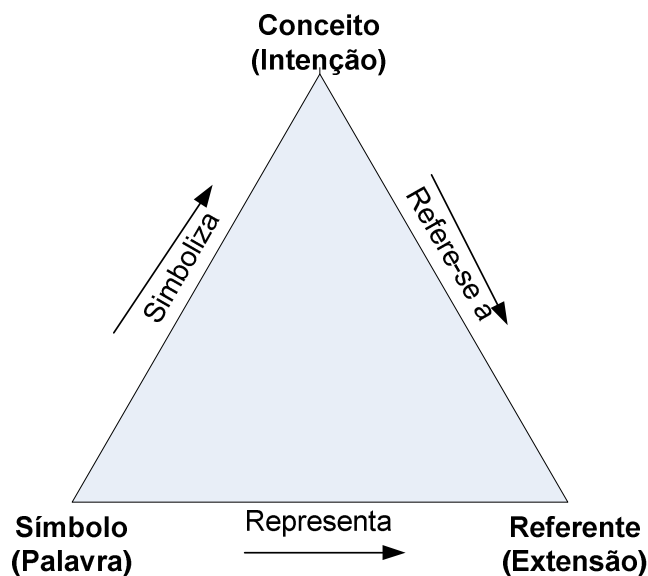


Figura 20 – Triângulo da Significação

Fonte: Sowa (2000) *apud* Ogden, Richards (1923).

Para construir uma ontologia de maneira semiautomática é necessária a automatização do processo de aquisição de conhecimento e, para isso, diversas abordagens foram sugeridas (MAEDCHE; STAAB, 2002). Na próxima seção descreve-se algumas abordagens para o aprendizado de ontologias.

### 2.2.8.1 Estado da Arte no Aprendizado de Ontologias

Maedche e Staab (2001) diferenciam as abordagens para o aprendizado de ontologias focados no tipo de entrada utilizado: **dicionários**, **bases de conhecimento**, **textos**, entre outros. A Figura 21 esquematiza as diferentes abordagens. Na Figura 21 as abordagens são divididas em dicionários, bases de conhecimento e textos. Este último divide-se em agrupamentos, cálculos estatísticos, padrões linguísticos, regras de associação e FCA. Estas técnicas podem ser utilizadas individualmente ou combinadas.

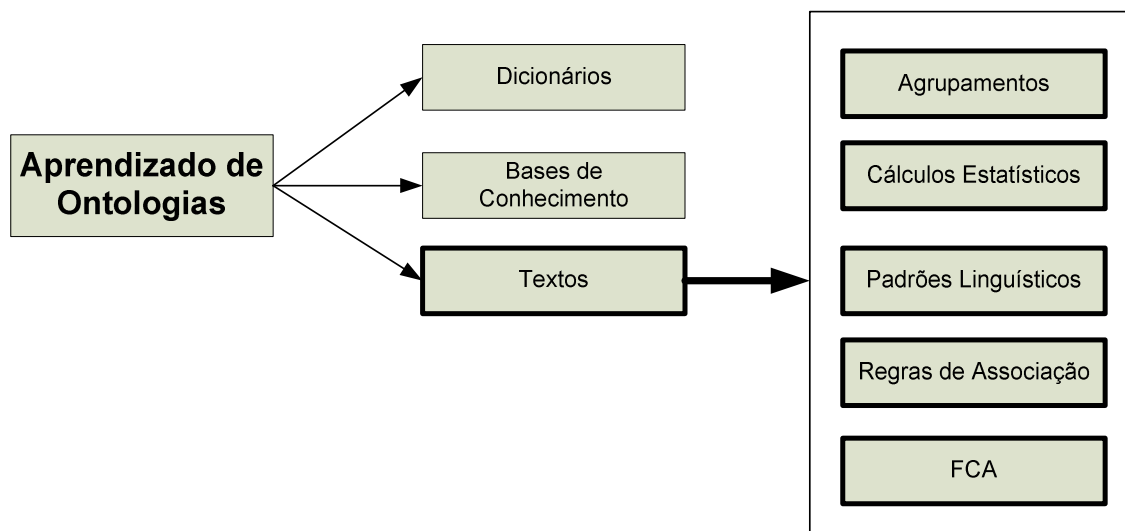


Figura 21 – Divisão das Abordagens de Aprendizado de Ontologias.

O aprendizado a partir de **dicionários** (ALANI et al., 2003; KHAN; LUO, 2002; KIETZ; MAEDCHE; VOLZ, 2000) consiste na utilização de dicionários para identificação de elementos e relações. Por exemplo, a utilização da *WordNet*. As definições e relacionamentos presentes nos dicionários são utilizados para definir a hierarquia e os relacionamentos de conceitos na ontologia. Embora estas abordagens capturem algumas relações, conforme disponibilidade no dicionário, outras relações entre os conceitos podem existir que não estejam presentes no dicionário. Com isso, estes métodos precisam ser combinados com outras técnicas para descobrir relações específicas do domínio. Os dicionários possuem uma definição geral não restrita ao domínio. Além disso, as ontologias desenvolvidas são restritas ao volume e tamanho do dicionário, devido a quantidade de vocábulos existentes no mesmo.

No aprendizado de ontologias a partir de **bases de conhecimento**, as abordagens (ALANI et al., 2003; SURYANTO; COMPTON, 2001) utilizam regras existentes nas bases para desenvolver uma ontologia. Estas abordagens requerem a priori a base de conhecimento com suas regras e instâncias. Logo, as ontologias desenvolvidas são restritas a cobertura da base.

No aprendizado de ontologias a partir de **textos**, as ontologias são desenvolvidas a partir de um conjunto de textos representativos do domínio e selecionados por especialistas/projetistas. Nestes textos são aplicadas técnicas de NLP e aprendizado de máquina. No aprendizado a partir de textos, os métodos são subclassificados de acordo com a técnica utilizada: **agrupamento** (*i.e. clusterização*), **cálculos estatísticos**, **padrões linguísticos**, **regras de associação** e **FCA**.

Nas abordagens (FAURE; NEDELLEC, 1998; ASSADI, 1998; HOTH0; MAEDCHE; STAAB, 2001) que utilizam o **agrupamento** os termos são agrupados de acordo com alguma característica. Esta característica é obtida por algum cálculo de similaridade. Por exemplo, os termos podem ser agrupados de acordo com seu sinônimo. Além disso, os grupos devem ser representados por algum termo que acredita-se ser o mais frequentemente utilizado. No agrupamento há problemas com a definição do cálculo da similaridade mais apropriado e a dificuldade em encontrar um termo que melhor represente todo o grupo (termos podem ter vários significados). Além disso, as abordagens existentes produzem apenas relações de hierarquias, ou seja, taxonômicas.

As abordagens **estatísticas** (MAEDCHE; STAAB, 2000b; MAEDCHE; STAAB, 2001; SRIKANT; AGRAWAL, 1995) confiam em medidas estatísticas como o TF/IDF (THOMPSON; MOONEY, 1999) para identificar elementos e suas relações. Estas abordagens são combinadas com outras técnicas como **regras de associação** (SRIKANT; AGRAWAL, 1995) e **padrões linguísticos**.

Nas **regras de associação**, primeiramente um processamento linguístico identifica as palavras que co-ocorrem. Após calcula-se as medidas de *support* e *confidence* para cada par de co-ocorrência. Ao final, determina-se o nível de abstração mais confiável que descreve o relacionamento conceitual. O desenvolvimento de uma ontologia por regras de associação é limitado porque não se sabe que tipo de relação pode ser construída somente pelas medidas *support* e *confidence*.

Nas abordagens baseadas em **padrões linguísticos** (SÁNCHEZ; MORENO, 2008; CIMIANO; WENDEROTH, 2005; YAMADA; BALDWIN, 2004; POESIO; ALMUHAREB, 2005; VILLAVERDE et al., 2009; SINTEK; BUITELAAR; OLEJNIK; 2004; VELARDI et al., 2005; MORIN, 1999; KIETZ; MAEDCHE; VOLZ, 2000; AUSSENAC-GILLES; BIÉBOW; SZULMAN, 2000) reconhecem-se elementos e relações por padrões de texto ou pela estrutura linguística de uma sentença. Casa-se a sequência de palavras em um texto com padrões previamente estabelecidos (NAVIGLI; VELARDI; GANGEMI, 2003). Um padrão é um formato genérico de uma expressão em linguagem natural para declarar um significado específico. Por exemplo, a ocorrência de verbos antes e após um elemento no texto pode definir um relacionamento de propriedade. Neste caso, os verbos são considerados nomes de propriedades. Nas frases “*causes hypertension*” e “*hypertension is treatate with*”, os verbos *causes* e *treatate* são propriedades relacionadas ao elemento *hypertension*. Outra maneira é estabelecer a priori padrões para a análise de textos. Por exemplo, estabelece-se o seguinte padrão: <plant part><becomes verb><color>. Em um

texto, ao analisar a frase *leaves turn brown*, identifica-se a parte de uma planta e uma cor. Porém, uma expressão em linguagem natural pode ter mais de um padrão e cada padrão denota um significado. A correta identificação dos padrões e argumentos é o ponto chave para este tipo de abordagem. A construção dos padrões deve ser feita por um especialista no domínio, mas não se sabe qual é o melhor padrão a ser utilizado. O projetista precisa criar os padrões ou regras com um conhecimento avançado do domínio. Outra possibilidade neste tipo de abordagem é utilizar padrões existentes.

A **FCA** é utilizada em algumas abordagens de aprendizado de ontologias (STUMME; MAEDCHE, 2001; HOTH; STAAB, 2005; WENG et al., 2006; CIMIANO; STAAB; TANE, 2003). Nesta abordagem os conceitos são obtidos a partir da tabela de contexto formal que consiste da combinação de objetos e atributos formando um reticulado de conceitos. Quando há muitos objetos e atributos, a construção do reticulado é difícil devido ao tempo de processamento. Este tipo de abordagem identifica apenas relações taxonômicas entre conceitos.

Buitelaar, Cimiano e Magnini (2005) afirmam que há uma grande variedade de abordagens e métodos para o aprendizado de ontologias a partir de **textos**. Os autores descrevem as abordagens baseados no objetivo, mais precisamente, que elementos identificam. Assim, Cimiano (2006) organiza os métodos de aprendizado de ontologias em camadas (Figura 22): **termos**, **sinônimos**, **conceitos**, **hierarquia de conceitos (relações taxonômicas)**, relações (**propriedades**), **hierarquia de propriedades**, **axiomas (schemata)** e **axiomas gerais**. As camadas podem ser vistas como etapas dos métodos de aprendizado, sendo que as camadas superiores exigem técnicas mais complexas.

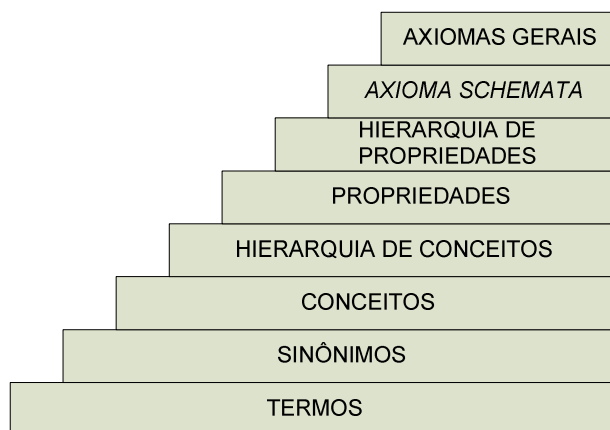


Figura 22 – Camadas/Etapas no Aprendizado de Ontologias

Fonte: Cimiano (2006).

A extração de **termos** é um pré-requisito para os métodos de aprendizado de ontologias a partir de textos (CIMIANO, 2006). Um termo é uma palavra ou conjunto de palavras com algum significado associado ao domínio de interesse. Os termos podem ser recuperados através de técnicas de indexação, recuperação de informação e NLP.

A etapa de identificação dos **sinônimos** envolve a obtenção de variações semânticas dos termos com significados equivalentes entre si. Abordagens nesta etapa integram dicionários como a *WordNet*. Um aspecto importante neste item é o significado a ser extraído, para isso, torna-se necessário um processo de desambiguação.

A descoberta de **conceitos** possui alguma sobreposição com a camada anterior de descoberta de sinônimos. As técnicas destas camadas (sinônimos e conceitos) descobrem palavras similares semanticamente que compartilham algum significado. Os sinônimos devem formar a base para a criação dos conceitos. Na descoberta de sinônimos procuram-se termos diferentes que podem ser substituídos sem alteração do significado. Por exemplo, o termo “cachorro” pode ser substituído por “cão”. Na descoberta de conceitos procuram-se termos que quando substituídos mudam o significado de um contexto. Por exemplo, “cão” e “gato” fazem parte de um contexto “animais de estimação”, mas não são sinônimos e não possuem os mesmos significados.

No entanto, o foco da maioria das abordagens de aprendizado de ontologias a partir de textos é a criação de **relações taxonômicas** (HEARST, 1992; KAVALEC; MAEDCHE; SVÁTEK, 2004).

As abordagens que criam **relações taxonômicas**, a partir da identificação de conceitos, agrupam os mesmos em uma hierarquia usando técnicas estatísticas (MAEDCHE; STAAB, 2000a; MAEDCHE; STAAB, 2001; SRIKANT; AGRAWAL, 1995), padrões linguísticos (HEARST, 1992; KAVALEC; MAEDCHE; SVÁTEK, 2004) e agrupamento (FAURE; NEDELLEC, 1998; ASSADI, 1998; HOTH0; MAEDCHE; STAAB, 2001). Conforme descrito anteriormente, as técnicas baseadas em padrões linguísticos e estatísticas baseiam-se no casamento de padrões de linguagem com análise da ocorrência de termos. O agrupamento cria grupos de termos com base em alguma característica. Há também outros trabalhos que utilizam técnicas de aprendizado de máquina como redes neurais (RUNG-CHING; JUI-YUAN; REN-HAO, 2008) e redes *Bayesianas* (RUI-LING; HONG-SHENG, 2009) para construção de taxonomias. Estas abordagens precisam de grandes quantidades de exemplos e os resultados precisam ser validados por projetistas.

Embora estas abordagens auxiliem o desenvolvimento de ontologias, um grande esforço dos projetistas deve ser dedicado a identificação de **propriedades** entre os conceitos (MAEDCHE; STAAB, 2000a). Nesta camada busca-se uma relação não taxonômica entre dois ou mais conceitos ou um atributo (*i.e.* a relação entre um conceito e um tipo de dado). O aprendizado de ontologias a partir de textos para identificar propriedades (SÁNCHEZ; MORENO, 2008; CIMIANO; WENDEROTH, 2005; YAMADA; BALDWIN, 2004; POESIO; ALMUHAREB, 2005; VILLAVERDE et al., 2009) combina análise estatísticas e linguística. Para identificar propriedades, algumas abordagens (MAEDCHE; STAAB, 2000a, 2000b) apenas identificam a existência de uma relação, ou seja, não nomeiam esta relação. Por exemplo, ao identificar uma relação entre um estudante e uma instituição de ensino pode ser nomeada de “estuda em”. Em (HEARST, 1992; MORIN, 1999; CIMIANO; HANDSCHUH; STAAB, 2004; KIETZ; MAEDCHE; VOLZ 2000) há um esforço para identificar e nomear relações semânticas pela identificação de padrões no texto. Outras abordagens identificam propriedades baseadas em padrões linguísticos (SÁNCHEZ; MORENO, 2008; CIMIANO; WENDEROTH, 2005; YAMADA; BALDWIN, 2004; POESIO; ALMUHAREB, 2005; VILLAVERDE et al., 2009). Estas abordagens analisam a ocorrência dos verbos antes e após os conceitos. Estes verbos são considerados propriedades. Algumas abordagens (MISSIKOFF; NAVIGLI; VELARDI, 2002; MAEDCHE; STAAB, 2000a; SCHUTZ; BUITELAAR, 2005) identificam propriedades pela análise da co-ocorrência (e. g. verbos que co-ocorrem com determinados substantivos – na associação de determinados substantivos). Um dos problemas identificados nestas abordagens é a dificuldade em recuperar propriedades dependentes do domínio (SÁNCHEZ; MORENO, 2008; KAVALEC; MAEDCHE; SVÁTEK, 2004) que são importantes para identificar/caracterizar um domínio (*e.g. company produces products* – domínios diferentes podem ter a mesma relação). Maedche (2002), Sánchez e Moreno (2008) afirmam que é difícil extrair propriedades, ou seja, é difícil descobrir o relacionamento entre pares de conceitos e nomear esta relação – há vários relacionamentos que podem existir entre os conceitos, há sinônimos para estes relacionamentos. Além de identificar as propriedades, também é importante identificar **hierarquias de propriedades**. Poucas abordagens possuem este propósito, entre elas Staab, Erdmann e Maedche (2001)

Cimiano (2006) divide o aprendizado de axiomas em duas camadas: **axioma schemata e axiomas em geral**. A camada de axioma *schemata* possui definições simples como disjunção e similaridade de relações entre conceitos. Na disjunção de conceitos o trabalho de Haase e Völker (2005) procura expressões (*e.g. “homens e mulheres”*) que

indicam uma provável disjunção de conceitos. Na simetria de relações (LIN; PANTEL, 2001), identificam-se relações similares entre conceitos pela ocorrência dos mesmos no texto. Por exemplo, “X escreveu Y” e “X é autor de Y” são duas relações similares entre X e Y. A última camada (**axiomas em geral**), diz respeito a restrições mais específicas como derivar axiomas de expressões condicionais quantificadas (*e.g* “todo homem ama uma mulher”). Nesta área Shamsfard e Barforoush (2004) propõem um trabalho para identificar axiomas.

Considerando a grande quantidade de documentos digitais disponíveis na *Web* e os que cada profissional e organizações possuem, o interesse em desenvolver ontologias a partir de texto vem aumentando. A utilização de textos como fonte de aquisição de conhecimento parece ser um caminho correto, visto que a linguagem é a primeira forma de transferência de conhecimento entre os seres humanos. Porém, as abordagens existentes baseadas em texto geralmente realizam um processamento de texto superficial/raso devido a grande quantidade de textos e também da complexidade dos mesmos (BUITELAAR; CIMIANO; MAGNINI, 2003).

Segundo Maedche e Staab (2004), nos métodos de aprendizado a partir de textos, não existe uma metodologia detalhada que guia o processo de aprendizado, os métodos utilizam principalmente processamento de linguagem natural e precisam de especialistas do domínio e engenheiros do conhecimento para verificar e validar a ontologia.

Segundo Sánchez e Moreno (2008) a *Web* apresenta características interessantes para a aquisição de conhecimentos, devido ao tamanho e heterogeneidade. Porém, há problemas em utilizar a *Web* como fonte de informação, entre eles: falta de estrutura, dinamicidade, informações não confiáveis, ruídos, ambiguidades. Além disso, o senso comum da *Web* não é mais forte que o senso comum dos especialistas do domínio.

## 2.3 ANOTAÇÃO

Anotação é um processo manual de indexação de recursos também conhecido como etiquetagem social (*social tagging*). A anotação, em oposição ao processo de organizar recursos em categorias, não requer experiência dos usuários, não exigem conhecimentos específicos. Os usuários escolhem um conjunto de palavras chaves, denominadas *tags*, do vocabulário individual para anotar um recurso *Web*.



As *tags* são selecionadas a partir de um espaço plano, sem qualquer hierarquia, refletindo o que o usuário pensa no momento da anotação do recurso. Não há controle na criação das *tags*. O usuário escolhe a *tag* mais adequada para descrever um recurso.

Hammond et al. (2005) identificou a motivação para o processo de anotação em quatro regiões, como ilustrado na Figura 23. Nesta, o eixo horizontal indica o criador da anotação e o eixo vertical se refere aos utilizadores das anotações. Descrevem-se as regiões:

- região 1: (eu, eu): representa um usuário que anotou seus recursos para seu próprio benefício (criador e consumidor). O *flickr*<sup>4</sup>, aplicativo para compartilhar fotos, é um exemplo nesta região.
- região 2: (outros, eu): representa um usuário que anotou outros recursos para seu uso. O aplicativo *Del.icio.us*<sup>5</sup> é utilizado nesta região para compartilhar *bookmarks*. Outra aplicação desta região é o *CiteULike*<sup>6</sup> utilizado para organizar e compartilhar referências bibliográficas.
- região 3: (eu, outros) representa um usuário que anota seu recurso em benefício de outras pessoas. Nesta região exemplificam-se páginas pessoais e criação de blogs.
- região 4: (outros, outros) representa usuários que anotam outros recursos para outros usuários. Por exemplo, a aplicação *Wikipedia*<sup>7</sup>.

---

<sup>4</sup> <http://www.flickr.com/>

<sup>5</sup> <http://delicious.com/>

<sup>6</sup> <http://www.citeulike.org/>

<sup>7</sup> <http://pt.wikipedia.org/>

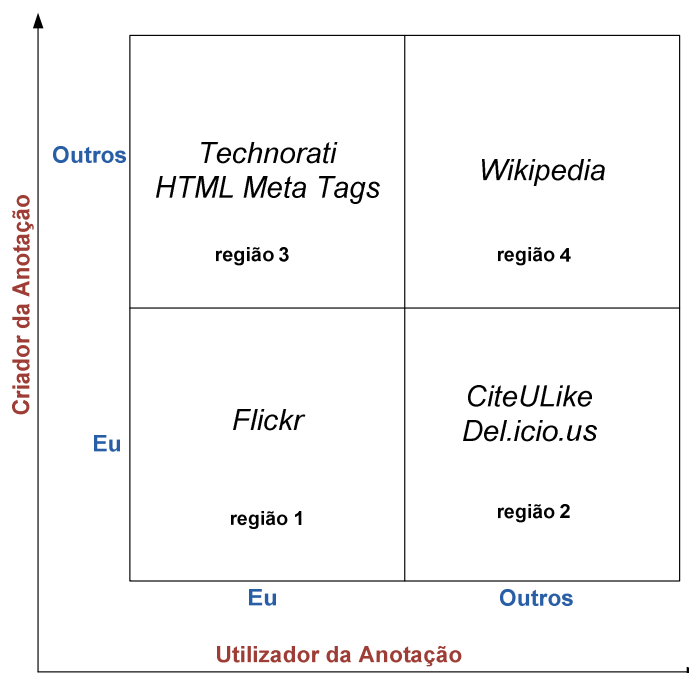


Figura 23 – Motivação no Processo de Anotação

Fonte: Hammond et al. (2005).

Hotho et al. (2006) definem o processo de anotação pela composição de três componentes: o usuário que executa a anotação, o recurso que é anotado e as *tags* que categorizam o recurso. Outros componentes podem ser utilizados – dependendo da disponibilidade do sistema de anotação – que incluem título, descrição textual e data da anotação.

Semelhante a Hotho et al. (2006), Damme, Hepp e Siorpaes (2007) definem que há quatro grupos envolvidos no processo de anotação (Figura 24): *tags*, objetos (recursos), atores (usuários) e os sistemas (*e.g Del.icio.us*) para criação de tais anotações. Os usuários atribuem *tags* aos recursos utilizando um sistema de anotação. Como consequência, as *tags* expressam e refletem o nível subjetivo do usuário e seu interesse no respectivo recurso. Os usuários utilizam *tags* para categorizar os recursos, criam conexões entre estes recursos, provendo significado para seu próprio entendimento e estabelecendo uma forma de como as informações são organizadas (WU; ZHANG; YU, 2006; MATHES, 2004).

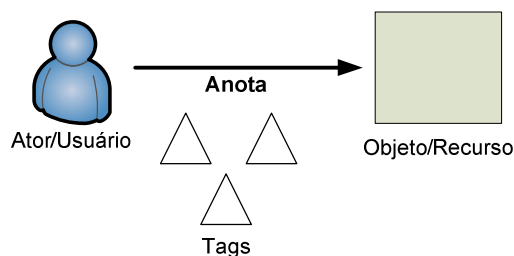


Figura 24 – Grupos de Entidades Envolvidas na Anotação

Fonte: Hotho et. al (2006).

A Figura 25 ilustra um exemplo de anotação de recursos, neste caso, sítios *Web*. Observando a Figura 25 percebe-se que os usuários podem utilizar as mesmas *tags* para categorizar recursos distintos ou semelhantes, além do fato que diferentes *tags* podem ser utilizadas para anotar os mesmos recursos. Usuários estão indiretamente conectados a outros por compartilhar as mesmas *tags* e recursos. Por exemplo, como mostrado na Figura 25, usuários 1 e 2 estão conectados pela *tag* 2 e os usuários 2 e 3 estão relacionados porque ambos anotaram o mesmo recurso. No primeiro caso, a mesma linguagem, no segundo caso, o interesse pelo mesmo recurso.

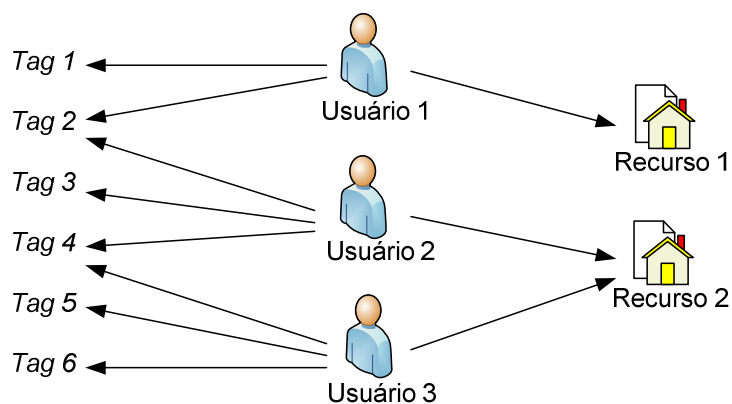


Figura 25 – Exemplo de um Processo de Anotação Coletiva.

Quando duas ou mais *tags* são utilizadas em conjunto em uma mesma anotação ocorre a co-ocorrência entre as *tags* (DAMME; HEPP; SIORPAES, 2007). A co-ocorrência informa quantas vezes um par de *tags* foi utilizado em conjunto para categorizar recursos distintos de um usuário/projetista/anotador. É uma relação semanticamente fraca que gera apenas uma estrutura plana entre as *tags*. A co-ocorrência de *tags* nas anotações pode ser

visualizada pelas nuvens de *tags* (Figura 26). A nuvem de *tags* dá ênfase às *tags* mais utilizadas nas anotações. É um método de apresentação e navegação nas *tags* onde as mais utilizadas são enfatizadas por alguma cor ou com diferentes tamanhos. Na Figura 26 a *tag design* é mais popular do processo de anotação. Nas nuvens de *tags* a única relação existente entre as *tags* é a co-ocorrência entre as mesmas

#### Tag Cloud: Popular

Sort: Alphabetically | [By size](#)

---

.net 2008 3d advertising ajax and animation api apple architecture **art** article articles artist audio **blog** blogging blogs book books browser **business** car cms code collaboration comics community computer converter cooking cool CSS culture data database **design** Design desktop **development** diy documentation download downloads drupal ebooks economics education electronics email entertainment environment fashion fic film finance firefox **flash** flex flickr food forum **free** freeware fun funny gallery game **games** geek google government graphics green guide hardware health history home hosting house **howto** html humor icons illustration images imported information **inspiration** interactive interesting internet iphone japan java **javascript** jobs jquery kids language learning library **linux** list lists literature **mac** magazine management maps marketing math media microsoft mobile money movie movies mp3 **music** network networking news online opensource osx people phone photo **photography** photos photoshop php plugin podcast **politics** portfolio privacy productivity programming psychology python radio rails realestate recipe recipes **reference** religion research resources reviews rss ruby rubyonrails school science search security seo shop **shopping** social socialnetworking **software** statistics streaming teaching tech **technology** tips todo tool tools toread travel tutorial tutorials tv twitter typography ubuntu usability **video** videos vim visualization **web** web2.0 **webdesign** webdev wiki wikipedia windows wishlist wordpress work writing youtube

---

Figura 26 - Nuvem de *Tags* para Visualização das *Tags* Utilizadas no Processo de Anotação.

O resultado do processo de anotação, pelo qual usuários etiquetam recursos com a utilização de *tags*, é denominado de folksonomias. Na próxima seção descreve-se folksonomias.

## 2.4 FOLKSONOMIAS E PERSONOMIAS

O termo folksonomia é a junção das palavras pessoas - *folks* e taxonomia - *taxonomy*. Este termo foi definido primeiramente por Vander Wal (2005). Em 2007, Vander Wal definiu folksonomia como:

*"... the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval. The tagging is done in a social environment (shared and open to others). The act of tagging is done by the person consuming the information. The value in this external tagging is derived from people using their own vocabulary and adding explicit meaning, which may come from inferred understanding of the information/object. People are not so much categorizing, as providing a means to connect items (placing hooks) to provide their meaning in their own understanding."*

O conjunto de anotações e *tags* de um usuário compõe sua personomia (HOTH0 et al., 2006; JÄSCHKE et al., 2008) e um conjunto de personomias disponibilizados socialmente para uma comunidade de usuários caracteriza uma folksonomia (WAL, 2007; JÄSCHKE et al., 2008).

As personomias são extremamente dinâmicas, pois refletem a estrutura mental do usuário no momento da anotação. Uma personomia evolui a medida que o usuário aprende e revisa seus conhecimentos. Vários usuários que utilizam uma aplicação social constroem suas personomias e compartilham, não só os recursos anotados, mas também as *tags* com outros usuários.

Uma das características da folksonomia é que os usuários anotam recursos livremente com diferentes *tags*, dependendo de sua especialidade, experiência e percepção do mundo (BEGELMAN; KELLER; SMADJA, 2006; GOLDER; HUBERMAN, 2005; PETERSON; 2006; WU; ZUBAIR; MALY, 2006). Por exemplo um zoologista pode etiquetar uma foto de um leão como  $\{felidae, phantherinae, mammal\}$ , enquanto um não zoologista pode utilizar as *tags*  $\{lion, king, animal, jungle\}$  para o mesmo propósito. Esta liberdade contribui para o sucesso das folksonomias (usuários sem habilidades, conhecimentos a priori e sem um vocabulário compartilhado) (HOTH0 et al., 2006; WU; ZHANG; YU, 2006). Porém, esta simplicidade introduz o problema onde o mesmo recurso é etiquetado por diferentes *tags*. Logo, uma pesquisa por *mammal* pode ignorar recursos anotados pela *tag lion*. Assim a recuperação é limitada (baixo *recall*).

A folksonomia resultante da atividade de uma comunidade apresenta muito provavelmente sobreposições de *tags*, mostrando a emergência de conhecimento comum e consensual sob a forma de um vocabulário compartilhado. Se por um lado a criação de personomias é fácil e não exige conhecimento especializado, por outro, deixa o usuário introduzir ambiguidades na sua personomia atribuindo aos recursos etiquetas diferentes com o mesmo significado (sinonímia), e mesma etiqueta com significados diferentes (polissemia). Nas folksonomias, a ambiguidade é aumentada pelas diferentes visões que os usuários podem ter de um mesmo recurso, mesmo quando associam *tags* idênticas. Ainda, pelo problema de ambiguidade e falta de formalização, o uso de folksonomias na busca de recursos produz imprecisões e falta de cobertura (*precision e recall*) nos resultados. Define-se folksonomia e personomia:

**Definição 5 (Folksonomia)** (HOTH0 et al., 2006) - Uma folksonomia  $F$  é uma quádrupla  $F = \langle U, T, R, Y \rangle$ , sendo:

- $U$ : conjunto finito de usuários  $u$ ;
- $T$ : conjunto finito de *tags*  $t$ ;
- $R$ : conjunto finito de recursos  $r$ ;
- $Y$ : relação ternária entre os elementos, i.e.,  $Y \subseteq U \times T \times R$

**Definição 6 (*Personomia*)** (HOTH0 et al., 2006) - Uma personomia  $\mathbb{P}$  de um usuário  $u \in U$  é uma restrição de  $F$  para  $u$ , i.e.,  $\mathbb{P} = \langle T_u, R_u, I_u \rangle$ , sendo:

- $I_u = \{(t, r) \in T \times R \mid (u, t, r) \in Y\}$  (relações *tag* x recurso para um usuário  $u$ );
- $t$ : uma *tag*;
- $r$ : um recurso.

Vander Wal (2005) categorizou as folksonomias em dois grupos: extensas (*broad*) e limitadas (*narrow*). Uma folksonomia extensa é o resultado de muitos usuários anotando o mesmo recurso. Neste tipo de folksonomia verificam-se as anotações realizadas em comum pelos anotadores (qual a preferência de anotação de um recurso) e descobre-se o que é comum em uma comunidade de usuários. Folksonomia limitada é o resultado de um usuário anotar um recurso, ou um número menor de usuários anotadores. O grau de visibilidade é maior na folksonomia extensa do que na limitada. A folksonomia limitada perde a riqueza da anotação em massa (QUINTARELLI, 2005).

Kroski (2005), Quintarelli (2005) e Mathes (2004) definiram os pontos fortes das folksonomias, que incluem:

- Folksonomias são inclusivas: representam o vocabulário de todos os anotadores, não como taxonomias ou ontologias que são restritas a um vocabulário controlado.
- Folksonomias são atuais: usuários criam *tags* e anotações rapidamente.
- Folksonomias são não binárias: na anotação um recurso pode se encaixar em várias categorias, ao contrário das taxonomias, onde um item só pode existir em um único lugar.
- Folksonomia ajudam a observar o comportamento dos usuários: pelas folksonomias observa-se como os usuários anotam seus recursos. As *tags* dos usuários são consideradas subjetivas devido a natureza pessoal da anotação. Porém, alguns usuários podem achar interessantes as *tags* dos outros usuários ao utilizarem em suas anotações.

Kroski (2005) e Mathes (2004) identificaram os seguintes pontos fracos das folksonomias: são imprecisas por terem *tags* ambíguas, personalizadas e inexatas; e possuem termos compostos inúteis.

- Faltam controles na criação de sinônimos, homônimos.
- Produzem baixo *recall* na recuperação das anotações devido a falta de controle em sinônimos e homônimos.

Limpens, Gandon e Buffa (2008) identificaram os seguintes problemas nas folksonomias:

- *Tags* especializadas e *tags* mais gerais podem se referir ao mesmo objeto. Por exemplo, cão (especializado) e mamífero (mais geral) podem fazer referência ao mesmo recurso;
- A mesma *tag* pode ser utilizada por diferentes objetos em diferentes contextos. Por exemplo, a *tag* Itália pode ser usada para categorizar pratos típicos da cozinha italiana, bem como categorizar pessoas que vivem na Itália com seus respectivos endereços.

#### 2.4.1 Folksonomias e Taxonomias

De uma perspectiva de anotação, taxonomia e folksonomia podem ser colocadas nos dois extremos opostos (SHIRKY, 2005; QUINTARELLI, 2005). A taxonomia é uma abordagem *top-down*. É um tipo simples de ontologia que fornece vocabulário de domínio hierárquico específico, que descreve os elementos de um domínio e suas relações hierárquicas. Além disso, as taxonomias são criadas por especialistas no domínio. Em contraste, a folksonomia é uma abordagem *bottom-up*. A folksonomia não possui um determinado vocabulário nem tem uma hierarquia explícita. É o resultado do vocabulário dos usuários anotadores. Assim, não tem limites (ou seja, de duração indeterminada) e as anotações não são estáveis nem abrangentes.

#### 2.4.2 Folksonomias e Ontologias

A folksonomia é criada por usuários que utilizam sistemas sociais para anotar recursos com diferentes *tags*. As anotações e escolha das *tags* dependem da origem social, cultural, do conhecimento e da percepção do mundo do usuário (BEGELMAN; KELLER;

SMADJA, 2006; GOLDBER; HUBERMAN, 2005, PETERSON, 2006; WU; ZUBAIR; MALY, 2006).

Uma ontologia é um artefato desenvolvido por um processo de engenharia (GÓMEZ-PÉREZ; CORCHO; FERNÁNDEZ-LÓPEZ, 2004), que inclui as atividades de desenvolvimento, suporte e gerenciamento. Ontologias especificam conceitualizações comuns em um domínio de interesse, independente do modelo de dados.

Segundo Berners-Lee, Hendler e Lassila (2001), quando os usuários precisam de um processamento mais complexo para as folksonomias, são necessárias representações semânticas. As ontologias possuem exatamente o que falta para as folksonomias (HAMMOND, et al., 2005): permitem classificar um conteúdo fazendo referência a um vocabulário controlado (esta referência ou anotação de recursos é feita por especialistas); os metadados fornecidos pelo uso das ontologias podem ser interpretáveis pela máquina.

Em contraste à ontologia, folksonomias não representam explicitamente uma conceitualização compartilhada. Folksonomias possuem ambiguidade e uma organização plana onde o único relacionamento é correlação das anotações, não existindo um significado conceitual nas relações. Ontologias são inequívocas (não ambíguas), exigem uma noção clara e independente do contexto do que significa um conceito, uma relação e uma instância e são semanticamente ricas, mas são rígidas e possuem o problema da falta de consenso do usuário sobre a visão do domínio.

Dotsika (2009) afirma que ontologias e folksonomias desempenham um papel semelhante em sistemas de classificação de conteúdo *Web*, apesar de serem concebidos de diferentes maneiras.

Para integrar folksonomias e ontologias há diversas abordagens que consistem basicamente na representação de folksonomias em ontologias e na transformação de folksonomias em ontologias ou geração de ontologias a partir de folksonomias. Dotsika (2009) discute várias metodologias existentes para a integração de ontologias e folksonomias e conclui que eles ainda estão em fase experimental, falta automação e não abordam as questões de qualidade. Neste trabalho identificaram-se abordagens que procuram desenvolver ontologias a partir de folksonomias e a representação de ontologias em folksonomias.



### 2.4.3 Aprendizagem de Ontologias a partir de Folksonomias

Há vários trabalhos que procuram gerar algum tipo de estrutura a partir das folksonomias. A Figura 27 classifica estes trabalhos de acordo com o tipo de estrutura que gera e que métodos utilizam para gerar tal estrutura. Na Figura 27 as setas indicam os métodos que utilizam e os retângulos representam o tipo de estrutura gerada.

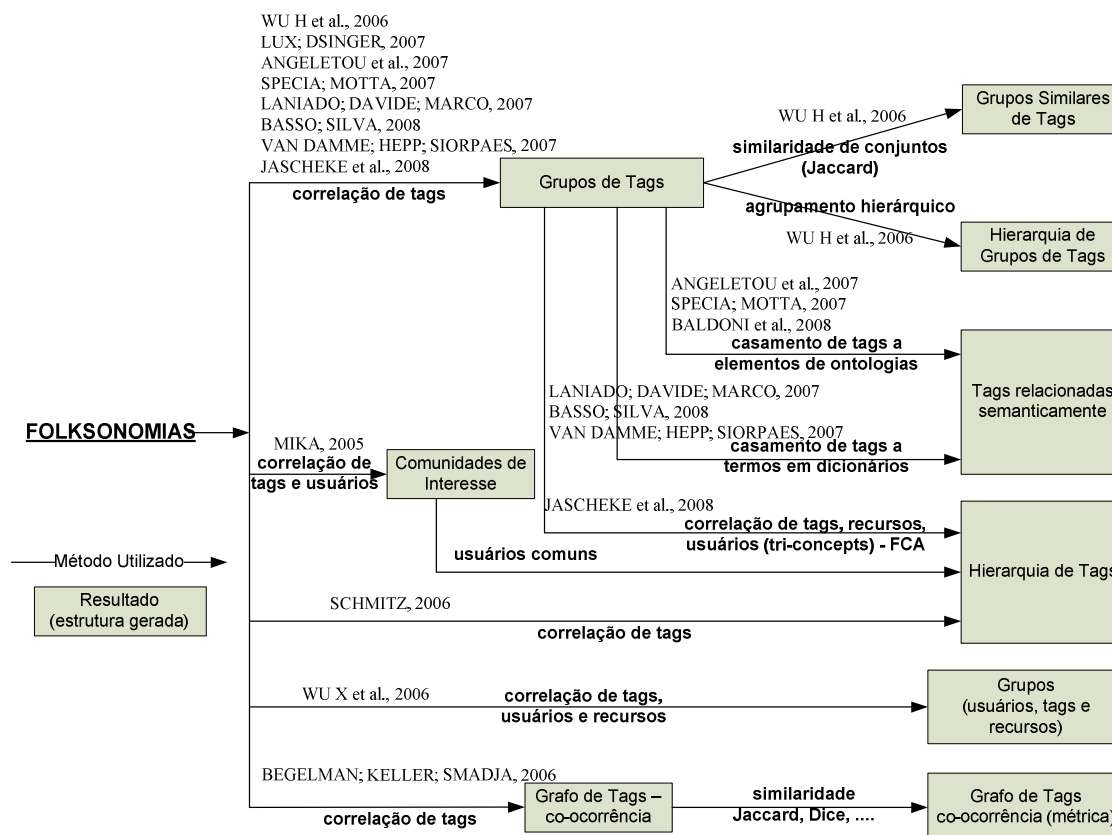


Figura 27 – Classificação dos Trabalhos de Aprendizagem de Ontologias a partir de Folksonomias.

Praticamente todos os trabalhos partem da **correlação de tags**, ou seja, verificam as *tags* que co-ocorrem nas anotações dos recursos, para então, utilizar outro método, entre eles, **similaridade de conjuntos**, **agrupamento hierárquico**, **casamento de tags a elementos em ontologias e dicionários**, **correlação de tags, recursos e usuários (tri-concepts e FCA)** e **comunidades de interesse**. O objetivo dos trabalhos é gerar alguma estrutura ou identificar algum tipo de relação existente entre as *tags* que se aproxime de uma ontologia e que também auxilie na navegação no espaço ou nuvem de *tags*. São as seguintes relações geradas pelos trabalhos: **grupos de tags**, **usuários e recursos**; **hierarquia de tags e de grupos de tags**;

relações **semânticas entre as tags baseado em alguma ontologia ou dicionário; grafos de tags**. A seguir uma breve descrição dos trabalhos.

Wu, Zubair e Maly (2006) criam grupos (*clusters*) de *tags* a partir de um conjunto de recursos anotados. Primeiro, verifica a correlação de *tags* na anotação dos recursos. Com base nesta correlação constrói grupos utilizando o agrupamento (*clusterização*) hierárquico. Neste tipo de agrupamento, cada grupo de *tags* subsequente é descendente do anterior, formando uma estrutura hierárquica, ou seja, grupos possuem subgrupos organizados em árvore. Após o agrupamento hierárquico, verificam-se os grupos mais similares utilizando a medida estatística de *Jaccard* que compara a similaridade e a dissimilaridade de amostras de grupos, sendo definido como o tamanho da interseção entre duas amostras *A* e *B* dividido pelo tamanho da união dessas duas amostras.

Schmitz (2006), a partir da correlação de *tags* utilizadas em anotações de certos documentos, identifica uma hierarquia de *tags*. A hierarquia é criada pelo seguinte princípio: uma *tag x* subsume uma *tag y* se os documentos o qual *y* ocorre são um subconjunto de documentos o qual *x* ocorre. Com isso, uma *tag* subsume outra porque é mais frequente nas anotações.

Begelman, Keller e Smadja (2006), para cada anotação, criam um grafo que corresponde a conexão das *tags* ao recurso anotado. Após, para cada par de *tags* (utiliza as *tags* de todos os conjuntos), verifica o número de vezes em que as *tags* aparecem juntas nas anotações. Cada *tag* é conectada a outra *tag* pelo número de co-ocorrências. Este número também conecta os diversos grafos, que é substituído por um valor de similaridade calculado por alguma medida como *Jaccard* ou *Dice*.

Mika (2005) constrói comunidades de interesse que incluem usuários que utilizam a mesma *tag* nas anotações. Por exemplo, uma comunidade de interesse pode ser representada por usuários que utilizam a *tag* “hotel” em suas anotações. Outra comunidade de interesse pode ser criada por usuários que utilizam a *tag* “acomodação”. Depois de agrupar usuários, que chama de comunidades de interesse, busca usuários comuns a pares de comunidades. Continuando o exemplo, Mika (2005) observa que as comunidades de interesse de “hotel” e “acomodação” possuem usuários em comum. Logo, as considera semanticamente relacionadas. Além disso, se o grupo de usuários da comunidade de interesse de “hotel” é um subconjunto dos usuários da comunidade de interesse “acomodação”, considera “acomodação” um termo mais geral que “hotel”. Com base neste tipo de relação constrói o que chama de ontologia leve (MCGUINNESS, 2002) pela identificação de relações taxonômicas.

Wu, Zhang e Yu (2006) apresentam um modelo probabilístico com o objetivo de gerar grupos de categorias que incluem: as *tags* estão relacionadas se co-ocorrem na anotação de um mesmo recurso; usuários possuem interesses similares se suas anotações compartilham as mesmas *tags*; recursos são relacionados se são etiquetados por usuários com interesses similares; *tags* são semanticamente relacionadas se são utilizadas por usuários com interesses similares; recursos relacionados são etiquetados várias vezes por *tags* semanticamente relacionadas.

Lux e Dsinger (2007) utilizam a medida de *edit-distance* para definir a *tag* original em cada anotação. Após, para cada *tag* original agrupa a mesma às demais *tags* pela co-ocorrência.

Os trabalhos (MIKA, 2005; WU; ZHANG; YU, 2006; BEGELMAN; KELLER; SMADJA, 2006; SCHMITZ, 2006; WU; ZUBAIR; MALY, 2006; LUX; DSINGER, 2007) descritos acima procuram identificar mais a correlação de *tags* do que identificar a semântica destas relações. A relação entre as *tags* não está explícita, ou seja, sabe-se que um par de *tags* está relacionado, mas fica difícil determinar automaticamente que tipo de relação é essa (ex: equivalência, generalização).

Jäschke et al. (2008) propôs o uso de técnicas de FCA, a fim de descobrir os grupos de usuários que compartilham a mesma conceitualização nos mesmos recursos. Para fazer isso, constroem triplas de conjuntos (recursos, usuários, *tags*) chamados *tri-concepts* onde cada usuário tem anotado cada recurso com todas as *tags*.

Outros trabalhos recentes (ANGELETOU et al., 2007; SPECIA; MOTTA, 2007) para gerar ontologias de folksonomias vão além dos trabalhos citados que identificam apenas a correlação implícita entre as *tags*. Specia e Motta (2007) e Angeletou et al. (2007) focam em identificar o sentido das *tags* e associá-las a elementos de outras ontologias disponíveis na *Web Semântica*.

Specia e Motta (2007) utilizam recursos e ontologias existentes para mapear *tags* em conceitos, propriedades ou instâncias. Integram folksonomias com ontologias disponíveis na *Web*. Para cada par de *tags* (todas as combinações de pares de *tags* existentes no grupo – o grupo foi criado com base na co-ocorrência) relacionadas no grupo, verifica quais as ontologias podem ser recuperadas em um sistema de busca. É uma busca sintática – *case matching* – *tag* com conceito, propriedade e instância. Caso encontre ontologias semelhantes, logo as *tags* estão relacionadas, caso contrário descarta. Quando há ontologias semelhantes, investiga os possíveis relacionamentos existentes (taxonômico, propriedades, disjunção) entre as *tags* com base nas ontologias recuperadas (e a conexão sintática entre os elementos).

Basicamente é um sistema de busca por ontologias baseado nas *tags*. O problema é que uma *tag* pode retornar várias ontologias ocorrendo conflitos nos relacionamentos que não são tratados.

Angeletou et al. (2007) propõem um método para enriquecer as folksonomias por explorar ontologias. Angeletou et al. (2007) também procura identificar a semântica entre as *tags*. É semelhante a Specia e Motta (2007). A partir de *tags* correlacionados por alguma abordagem conecta as *tags* a elementos em alguma ontologia utilizando um mapeamento de strings (SABOU; AQUIN; MOTTA, 2006). Basicamente, Angeletou et al. (2007) procurou implementar a proposta de Specia e Motta (2007).

Os trabalhos de Specia e Motta (2007), Angeletou et al. (2007) possuem algumas limitações. O casamento sintático (apenas baseado em *string*) pode não desprezar alguns elementos na ontologia. Além disso, a dificuldade em encontrar ontologias disponíveis e como em algumas vezes diferentes ontologias são recuperadas, ocorrem conflitos porque as ontologias podem ser contraditórias.

Outros trabalhos (LANIADO; EYNARD; COLOMBETTI, 2007; SILVA, 2009; DAMME; HEPP; SIORPAES, 2007) procuram criar relações entre as *tags* pela utilização de dicionários. Laniado, Eynard e Colombetti (2007) constroem uma ferramenta para auxílio a navegação de *tags*. A partir das *tags*, constrói hierarquias ou taxonomias. Estas hierarquias são construídas a partir da relação de hiperonímia da *WordNet*. Silva (2008) utiliza as relações de hiponímia e meronímia da *WordNet* para construir uma ontologia entre as *tags* que co-ocorrem nas anotações. Damme, Hepp e Siorpaes (2007), além de dicionários utilizam ontologias disponíveis para criar os relacionamentos entre as *tags*. Estes trabalhos (LANIADO; EYNARD; COLOMBETTI, 2007; SILVA, 2009; DAMME; HEPP; SIORPAES, 2007) são restritas (tamanho) ao dicionário. Relações específicas para o domínio podem não ser recuperadas devido a definição geral dos dicionários.

Baldoni et al. (2008) adiciona semântica a um sistema de anotações ARSMETEO (um sistema *Web* para compartilhar trabalhos relacionados a arte) relacionando as *tags* a uma ontologia do domínio *OntoEmotions*. O método é baseado em medir a relação existente entre as *tags* presentes no ARSMETEO e os conceitos na ontologia *OntoEmotions*. O trabalho de Baldoni et al. (2008) é restrita ao domínio e relaciona *tags* a ontologias simplesmente baseado em resultados de pesquisa.

#### 2.4.4 Ontologias para Representação das Folksonomias

Um problema nos sistemas baseados em folksonomia atuais é que não há interoperabilidade (KNERR, 2006), ou seja, os sistemas utilizam diferentes representações para o processo de anotações. Todos eles exigem que seus usuários criem e mantenham um vocabulário de termos utilizados para a categorização de seus recursos (sua personomia). Assim, um usuário que utilize vários sistemas baseados em folksonomia deverá manter, obrigatoriamente, um vocabulário em cada sistema, independentemente desses termos serem os mesmos ou completamente diferentes (SILVA, 2009). Visando criar formas padronizadas para a representação de anotações entre aplicações, alguns autores definiram modelos ontológicos.

Em Gruber (2005) e, Wu, Zhang e Yu (2006) são apresentadas as premissas básicas para representar os dados de anotações em algum modelo. Há pouca diferença entre os modelos de Gruber e Wu, uma vez que os dados mínimos necessários para essa representação são o usuário, o recurso categorizado e as *tags* utilizadas (os três componentes do processo de anotação). Wu, Zhang e Yu (2006) adicionam ainda a data em que a categorização ocorreu e Gruber (2005) adiciona dados sobre qual sistema a anotação ocorreu.

Newman (2005), Knerr (2006), Gruber (2008), Kim et al. (2007) e Echarte et al. (2007) focam em definir as premissas de Gruber (2005) e de Wu, Zhang e Yu (2006) em RDF e/ou OWL. Esse tipo de ontologia para a representação de folksonomias trás diversas vantagens sobre as folksonomias e ontologias puras como (i) a utilização de repositórios de *tags* para um conjunto de aplicações baseadas em folksonomias, (ii) a possibilidade de categorizar objetos utilizando *tags* em um ambiente entre aplicações e (iii) a interoperabilidade entre os sistemas baseados em folksonomia por meio de um formato padronizado e extensível. Os dados dessas ontologias podem ser utilizados diretamente pelo usuário e/ou disponibilizados socialmente (KNERR, 2006).

A ontologia SCOT (*Social Semantic Cloud of Tags*) (KIM et al., 2007) descreve a estrutura do processo de anotação. A SCOT é baseada nos conceitos e propriedades do modelo de Newman. A Figura 28 ilustra parcialmente a ontologia SCOT. Por exemplo, cada *tag* é representada como uma instância do conceito do (*scot: Tag*). Além disso, inclui propriedades adicionais para os recursos e o usuário responsável pela anotação.



Figura 28 – Ontologia SCOT

Fonte: Kim et al. (2007).

## 2.5 PROCESSAMENTO DE LINGUAGEM NATURAL

O Processamento de Linguagem Natural (*Natural Language Processing* - NLP) é uma área que se dedica ao estudo, tratamento e compreensão da linguagem humana através da tecnologia computacional. O NLP compreende um conjunto de métodos para analisar textos e gerar frases (compor textos) escritas em alguma linguagem humana. É utilizada por sistemas de geração de linguagem natural que convertem informações armazenadas em computadores em alguma linguagem humana e por sistemas de compreensão de linguagens que convertem escritas em representações mais formais (MANNING; SCHÜTZE, 1999).

Neste trabalho, do NLP, utilizam-se as técnicas de **radicalização**, **lematização**, **remoção de stopwords** e **desambiguação**.

A **radicalização**, também chamada de *stemming*, é um processo para identificar o radical de uma palavra, mais precisamente, reduzir as palavras aos radicais que as compõem de forma a agrupar por similaridade variações ortográficas que poderiam passar como palavras completamente distintas (PLISSON; LAVRAC; MLADENIC, 2004). Por exemplo, as palavras *computes*, *computed* e *computing* possuem em comum o radical *comput*. O radical compreende o elemento estrutural básico da palavra sem o prefixo ou sufixo.

A **lematização** reduz uma palavra a sua forma canônica (normalizada) sem derivações ou flexões. Por exemplo, a palavra *womem* após a lematização é transformada em *woman*. No mesmo exemplo acima da lematização, as palavras *computes*, *computed* e *computing* após a lematização, tornam-se *compute*, a forma infinitiva do verbo.

A **remoção das stopwords** compreende em remover palavras não expressivas dos textos, ou seja, que não enriquecem um texto. Estas palavras incluem artigos, preposições, conjunções e pronomes.

A **desambiguação** do sentido da palavra (*Word Sense Desambiguation* – WSD) identifica qual é o sentido (significado) de uma palavra, em qualquer sentença, baseado no seu contexto de uso, quando a mesma tem vários sentidos distintos (polissemia) (MANNING; SCHÜTZE, 1999). Por exemplo, a palavra “jaguar” pode estar se referindo a marca de carros ou ao animal. Neste trabalho utilizou-se a métrica de relação semântica chamada *Lesk* adaptada (BANERJEE; PEDERSEN, 2002) para a *WordNet*. O princípio da métrica *Lesk* adaptada é que quanto mais palavras houver em comum entre as descrições (*glosses*) de dois *synsets*, mais relacionados eles serão. O resultado da métrica *Lesk* adaptada é dado pela soma dos quadrados dos tamanhos das sobreposições (*length (overlap)*), conforme descrito na equação:

$$Sim_{Lesk} = \sum_i^{\#overlaps} length(overlap_i)^2$$

Conforme equação, a sobreposição de uma palavra traz o resultado 1, a sobreposição de duas palavras não consecutivas traz o resultado 2, a sobreposição de duas palavras consecutivas traz o resultado 4, a sobreposição de 3 palavras consecutivas traz o resultado 9 e assim por diante.

A seguir um exemplo de comparação dos *synsets* das palavras motocicleta (*motorcycle*) e automóvel (*automobile*). Os *synsets* de motocicleta e automóvel possuem as seguintes descrições recuperadas da *WordNet*:

- motocicleta: “*motor vehicle with two wheels and a strong frame*”;
- automóvel: “*motor vehicle with four wheels; usually propelled by an internal combustion engine*”.

Observa-se nas descrições que ao ignorar as *stopwords* (*with, and, a, by na*) ocorrem três sobreposições (*motor, vehicle, wheels*), sendo duas delas consecutivas (*motor, vehicle*). O resultado da métrica *Lesk* para as duas descrições é cinco ( $2^2 + 1 = 5$ ).

Um exemplo completo de desambiguação utilizando a métrica *Lesk* é representado na Figura 29. No exemplo da Figura 29 foram comparados entre si os *synsets* das palavras *Java*, *Programming Language* e *Prolog*.

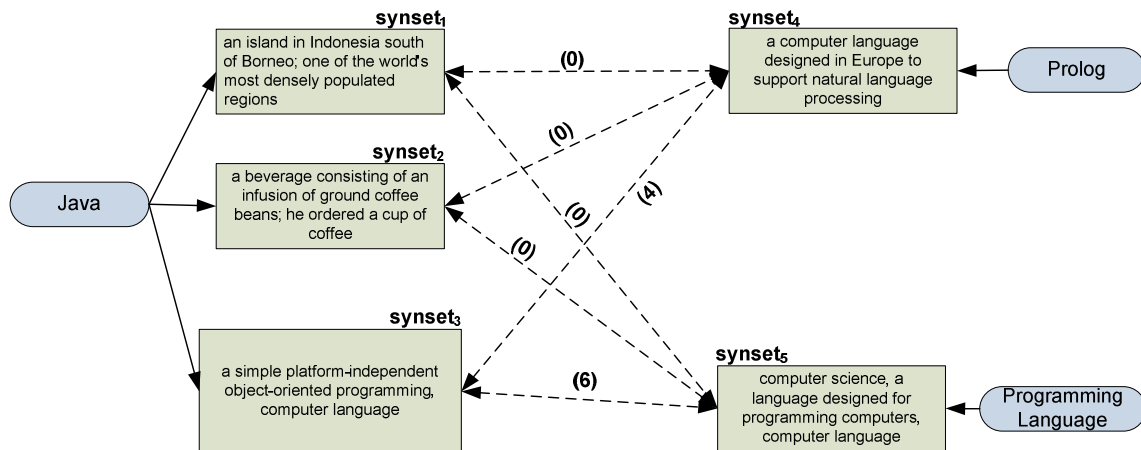


Figura 29 – Relação entre os *Synsets* das palavras *Java*, *Programming Language* e *Prolog*.

A única palavra ambígua do exemplo é *Java*, cujos sentidos estão representados por *synset<sub>1</sub>* (*an island in Indonesia south of Borneo; one of the world's most densely populated regions*), *synset<sub>2</sub>* (*a beverage consisting of an infusion of ground coffee beans; he ordered a cup of coffee*) e *synset<sub>3</sub>* (*a simple platform-independent object-oriented programming, computer language*). O sentido da palavra *Prolog* é descrito pelo *synset<sub>4</sub>* (*a computer language designed in Europe to support natural language processing*) e o sentido da palavra *Programming Language* pelo *synset<sub>5</sub>* (*computer science, a language designed for programming computers, computer language*).

Comparando-se as descrições dos *synsets*, não há sobreposição de palavras entre *synset<sub>1</sub>*, *synset<sub>2</sub>*, *synset<sub>4</sub>* e *synset<sub>5</sub>*. Na comparação do *synset<sub>3</sub>* com o *synset<sub>4</sub>*, ocorre a sobreposição das palavras *computer* e *language*, sendo consecutivas. Nesta comparação o valor da métrica *Lesk* é 4 ( $2^2 = 4$ ). Na comparação do *synset<sub>3</sub>* com o *synset<sub>5</sub>*, as palavras *programming*, *computer* e *language* são sobreposições, sendo *computer* e *language* consecutivas. Nesta comparação o valor da métrica *Lesk* é 6 ( $1+1+2^2=6$ ). Assim, o *synset<sub>3</sub>* é o escolhido como o sentido da palavra ambígua *Java*, ou seja, o *synset<sub>3</sub>* da palavra *Java* é mais



fortemente relacionado com os sentidos das palavras co-ocorrentes *Prolog* e *Programming Language*.

Além das técnicas apresentadas, outra possibilidade é utilizar ontologias linguísticas para obter termos e relações entre os mesmos.

### 2.5.1 Ontologias Linguísticas

O objetivo deste tipo de ontologia é descrever a semântica das unidades gramaticais (substantivo, verbo, adjetivo e advérbio) ao invés de um modelo específico de algum domínio. São exemplos de ontologias linguísticas: *WordNet* (MILLER et al., 1990; MILLER, 1995), *EuroWordNet* (VOSSEN, 1998), GUM (*Generalized Upper Model*) (BATEMAN; FABRIS; MAGNINI, 1995), *Mikrokosmos* (MAHESH; NIRENBURG, 1995; MAHESH, 1996) e SENSUS (SWARTOUT et al., 1997). Estas ontologias são utilizadas no processamento de linguagem natural, no desenvolvimento de tradutores, entre outras aplicações.

A ontologia linguística *WordNet*, também conhecida como banco de dados léxico para a língua inglesa, organiza a informação por meio do significado das palavras. Na *WordNet* as palavras são agrupadas em conjuntos de sinônimos cognitivos chamados de *synsets*, cada um expressando um conceito distinto. Por exemplo, o conjunto  $\langle world, human\ race, humanity, humankind, human\ beings, humans, mankind, man \rangle$  é um *synset* consistindo de palavras e expressões definidas como todos os habitantes da terra (*all inhabitants of the earth*) (KIKAS; TREUMUTH, 2007). Até o início do ano de 2009 a *WordNet* continha mais de 155 mil palavras distintas, distribuídas em mais de 117 mil *synsets* (WORDNET, 2009).

Os *synsets* são conectados de acordo com o significado a outros *synsets* por meio de vários tipos de relações (FELLBAUM, 1998; KIKAS; TREUMUTH, 2007):

- Sinonímia (sinônimos): é a relação que se estabelece entre dois *synsets* ou mais que apresentam significados iguais ou semelhantes em certo contexto. Por exemplo, o *synset*<sub>1</sub> que possui o conjunto de sinônimos  $\langle car, auto, automobile \rangle$  tem uma relação de sinonímia com *synset*<sub>2</sub> que possui o conjunto de sinônimos  $\langle car, railcar, railway\ car, railroad\ car \rangle$ .
- Hiperonímia (*hypernym*): define o *synset* geral de um *synset*. O *synset*<sub>1</sub> é hiperônimo do *synset*<sub>2</sub> se o significado do *synset*<sub>1</sub> subsume o significado do

$synset_2$ . Por exemplo, o *synset* de *furniture* é hiperônimo do *synset* de *chair*. A relação inversa é chamada de hiponímia.

- Hiponímia (*hyponym*): define os *synsets* mais específicos de um *synset*. O  $synset_1$  é hipônimo do  $synset_2$  se o significado do  $synset_1$  é subsumido pelo significado do  $synset_2$ . Por exemplo, os *synsets* de *bed* e *chair* são hipônimos do *synset* de *furniture*.
- Holonímia (*holonym*): define o todo de um *synset*. O  $synset_1$  é parte do  $synset_2$ , ou seja, o  $synset_1$  é holônimo do  $synset_2$  se o  $synset_1$  for a parte principal ou o todo que possui como parte o  $synset_2$ . Por exemplo, o *synset* de *wheel* é holônimo do *synset* de *vehicle*.
- Meronímia (*meronym*): define as partes de um *synset*. O  $synset_1$  é parte do  $synset_2$ , ou seja, o  $synset_1$  é merônimo do  $synset_2$  se  $synset_2$  for parte do  $synset_1$ . Por exemplo, o *synset* de *person* é merônimo de *group*.

A Figura 30 ilustra um exemplo das relações na *WordNet* resultante da pesquisa pela palavra *bike*. A palavra *bike* possui dois *synsets*,  $s_1$   $\langle bike, motorcycle \rangle$  e  $s_2$   $\langle bike, bicycle, wheel, cycle \rangle$  relacionados por sinonímia. Outras relações encontradas para os dois *synsets* são hiperonímia  $\langle wheeled vehicle \rangle$  e meronímia  $\langle mudguard, splash guard \rangle$ . Além das relações, cada *synset* possui uma descrição textual, chamada *gloss*, do significado da palavra.

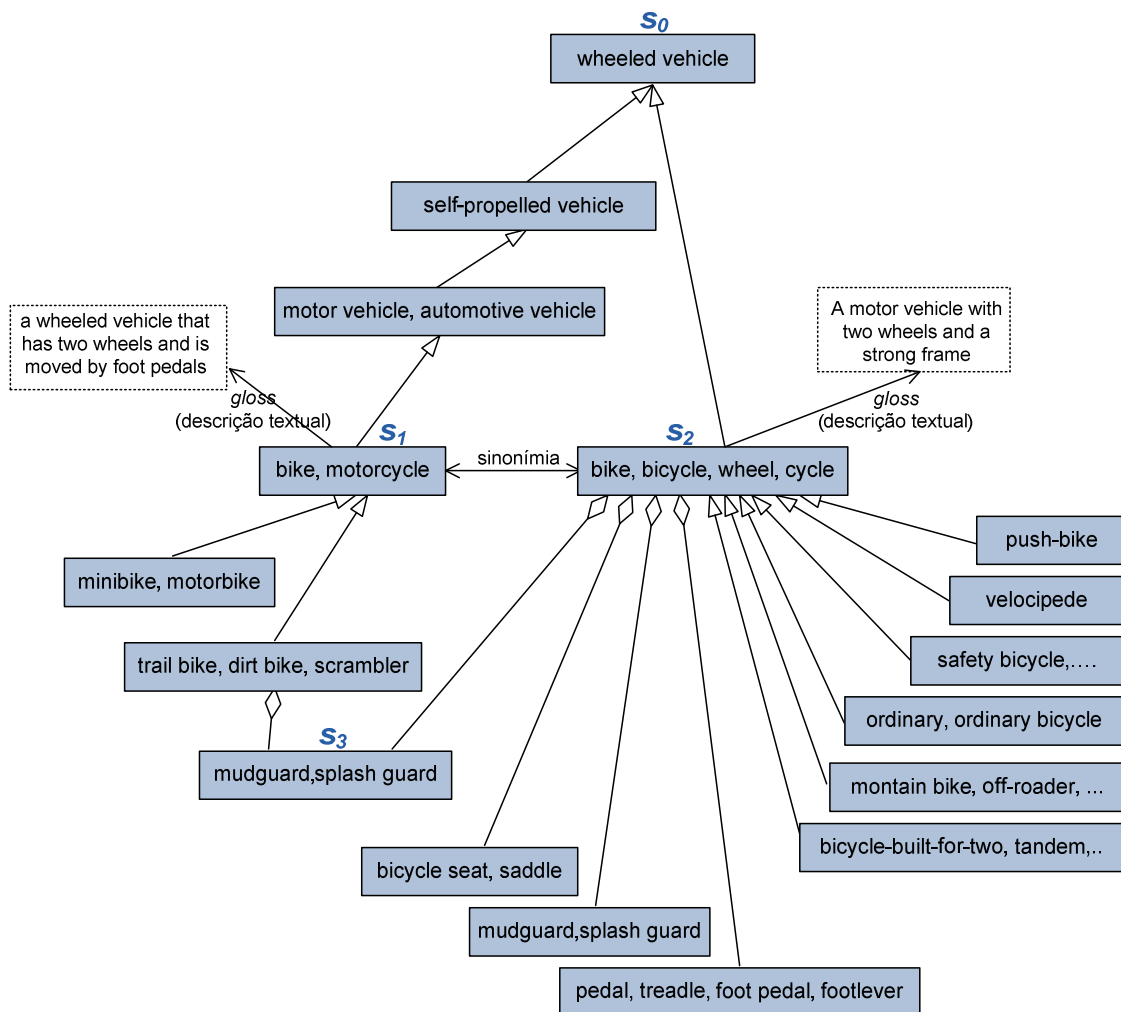


Figura 30 – Exemplo de Relações na *WordNet* resultante da Pesquisa pela Palavra *Bike*.

## 2.6 MEDIDAS DE SIMILARIDADE

Nesta seção são definidas as medidas de similaridade utilizadas no método de alinhamento de ontologias denominado POAM (*Partial Ontology Alignment Method*) (FREDDO et al., 2006; FREDDO; TACLA, 2009) que foi adaptado para ser utilizado neste trabalho. O POAM original faz o alinhamento de conceitos de ontologias diferentes comparando-se as propriedades dos mesmos. O POAM original foi aplicado na interoperabilidade entre serviços *Web* (FREDDO et al., 2008; FREDDO; TACLA, 2010). Neste trabalho, o POAM alinha instâncias ( $I$ ) com base nos valores das propriedades ( $R_I$ ).

Uma função de similaridade recebe como parâmetros duas entidades e retorna como resultado um valor entre 0 e 1. O valor 0 indica que não há nenhuma relação entre duas entidades e 1 indica que as duas entidades são exatamente idênticas. Para os demais resultados, quanto mais próximo de 1, mais as entidades são próximas uma das outras e quanto mais próximo de 0, mais as entidades se distanciam. Define-se similaridade:

**Definição 7 (similaridade):** similaridade é uma característica em comum entre duas entidades  $e$  ( $e_1, e_2$ ). Uma entidade pode ser um termo (*string*) ou um conjunto de termos. Uma similaridade  $sim$  é uma função:

$sim: e_1 \times e_2 \rightarrow [0,1]$ , sendo:

- $sim(e_1, e_2) \geq 0$ : similaridade positiva;
- $sim(e_1, e_2) = 1$ : similaridade máxima;
- $sim(e_1, e_2) = 0$ : não há similaridade.

A partir da Definição 2, são definidas as seis medidas de similaridades ( $sim_1, \dots, sim_6$ ) utilizadas neste trabalho.

**Definição 8 ( $sim_1$  - Levenshtein):** utiliza a medida de *Levenshtein* para calcular o número de operações (inserir, remover, substituir) necessárias para transformar um termo  $i$  em outro  $j$ .

$$sim_1(termo_i, termo_j) = \frac{1}{1 + Levenshtein(termo_i, termo_j)} \in [0 \dots 1]$$

O algoritmo de *Levenshtein* (NAVARRO, 2001) calcula o menor número de inserções, remoções e substituições para igualar dois termos. A Figura 31 ilustra o algoritmo utilizado.

```

levenshtein(termoi, termoj)
início
    //retorna o comprimento da string
    comprimento_termoi ← comprimento(termoi);
    comprimento_termoj ← comprimento(termoj);

    inteiro: tab[0..comprimento_termoi, 0..comprimento_termoj];

    para X de 0 até comprimento_termoi
        tab[X, 0] ← X;
    para Y de 0 até comprimento_termoj
        tab[0, Y] ← Y;
    para X de 1 até comprimento_termoi
        para Y de 1 até comprimento_termoj
            se termoi[X] = termoj[Y]

```

```

        então cost ← 0;
        senão cost ← 1;
        tab[X, Y] ← menor(
            tab[X-1, Y ] + 1, // Deletar
            tab[X , Y-1] + 1, // Inserir
            tab[X-1, Y-1] + cost // Substituir
        );
        levenshtein ← tab[comprimento_termoi, comprimento_termoj];
    fim

```

Figura 31 – Algoritmo de *Levenshtein*

Fonte: Navarro (2001).

A seguir um exemplo de utilização do algoritmo. No exemplo, o algoritmo calcula o número de operações necessárias para transformar o termo *computes* ( $termo_i$ ) em *computing* ( $termo_j$ ). A Figura 32 ilustra a tabela *tab* criada para calcular a distância. Conforme valor da célula correspondente a última linha e última coluna da *tab*, a distância máxima encontrada foi três, ou seja, são necessárias três operações para transformar *computes* em *computing*.

		Y									
		c o m p u t i n g									
X	c	0	1	2	3	4	5	6	7	8	9
	o	1	0	1	2	3	4	5	6	7	8
	m	2	1	0	1	2	3	4	5	6	7
	p	3	2	1	0	1	2	3	4	5	6
	u	4	3	2	1	0	1	2	3	4	5
	t	5	4	3	2	1	0	1	2	3	4
	e	6	5	4	3	2	1	0	1	2	3
	s	7	6	5	4	3	2	1	1	2	3
	s	8	7	6	5	4	3	2	2	2	3

Figura 32 – Exemplo do Cálculo da Distância de *Levenshtein* entre os Termos *computes* e *computing*.

**Definição 9** (*sim<sub>2</sub> - Hierarquia WordNet*): calcula a similaridade entre dois termos  $i$  e  $j$  na hierarquia *WordNet*. A similaridade é baseada na soma das distâncias (número de termos) entre os termos e o antepassado (termo que os subsume) mais comum (*termoComum*) na hiperonímia entre *synsets*.

$$sim_2(termo_i, termo_j) = \frac{1}{1 - dist(termo_i, termo_j)} \in [0 \dots 1]$$

$$dist(termo_i, termo_j) = distância(termo_i, termoComum) + distância(termo_j, termoComum)$$

Para exemplificar o cálculo, utiliza-se a taxonomia ilustrada na Figura 33. O exemplo consiste em calcular a distância do termo *dog* para os termos *cat* e *fish*. Para o cálculo da distância entre os termos *dog* e *cat*, o antepassado mais comum recuperado é *carnivore*. A distância de *dog* para *carnivore* é de dois termos. O mesmo valor também é obtido na distância de *cat* para *carnivore*. Logo, a soma das distâncias de *dog* para *cat* é de quatro termos. Agora, para a distância entre *dog* e *fish*, o termo comum é *vertebrate*. A distância de *dog* para *vertebrate* é de quatro termos. A distância de *fish* para *vertebrate* é de dois termos. Logo, a soma das distâncias de *dog* para *fish* é de seis termos. Assim, a similaridade entre *dog* e *cat* é maior comparada à similaridade entre *dog* e *fish*.

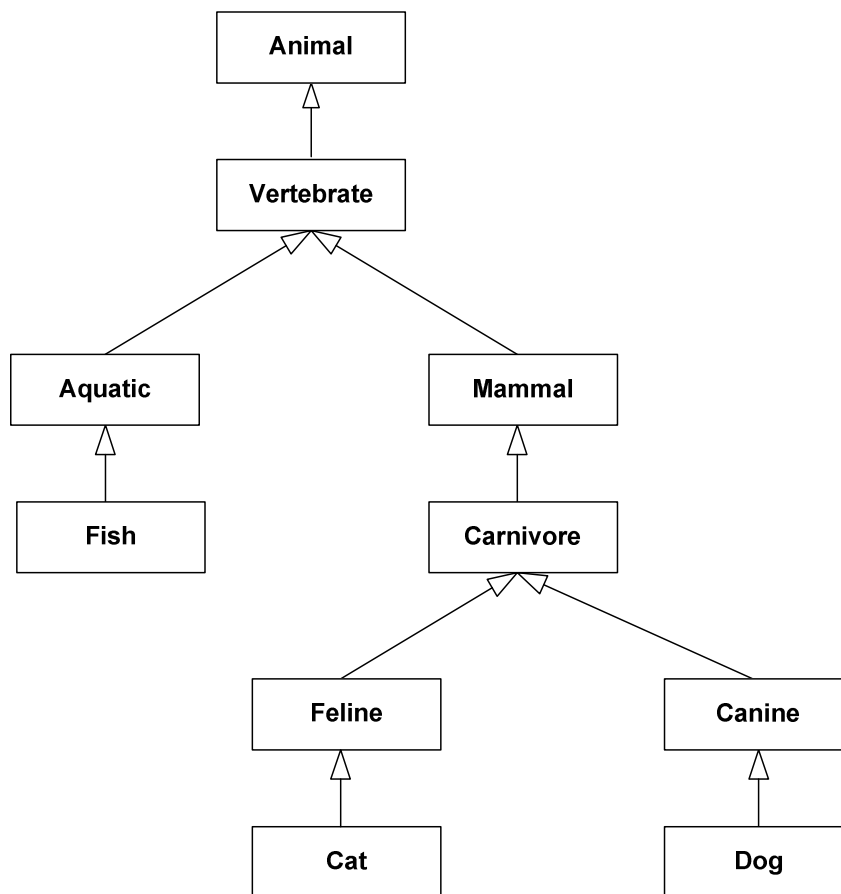


Figura 33 – Exemplo de uma Taxonomia.

**Definição 10** ( $sim_3$  - *Igualdade de Conjuntos*): calcula a similaridade entre os elementos de dois conjuntos  $i$  e  $j$  através dos elementos comuns.

$$sim_3(\text{conjunto}_i, \text{conjunto}_j) = \frac{(\text{conjunto}_i \cap \text{conjunto}_j)}{(\text{conjunto}_i \cup \text{conjunto}_j)} \in [0 \dots 1]$$

Para exemplificar o cálculo utilizaram-se três conjuntos de termos (*conjunto*<sub>1</sub>, *conjunto*<sub>2</sub>, *conjunto*<sub>3</sub>) ilustrados na Figura 34. A similaridade entre o *conjunto*<sub>1</sub> e *conjunto*<sub>2</sub> é calculada pela divisão dos 4 elementos (*dog*, *cat*, *feline*, *canine*) resultantes da interseção, pelos 5 elementos (*dog*, *cat*, *feline*, *canine*, *mammal*) resultantes da união. O resultado é 0,8 de similaridade entre o *conjunto*<sub>1</sub> e *conjunto*<sub>2</sub>. A similaridade entre o *conjunto*<sub>1</sub> e *conjunto*<sub>3</sub> é calculada pela divisão dos 2 elementos (*dog*, *cat*) resultantes da interseção, pelos 8 elementos (*dog*, *cat*, *feline*, *canine*, *mammal*, *fish*, *animal*, *vertebrate*) resultantes da união. O resultado é 0,25 de similaridade entre o *conjunto*<sub>1</sub> e *conjunto*<sub>3</sub>. Logo, a similaridade é maior entre o *conjunto*<sub>1</sub> e *conjunto*<sub>2</sub>.

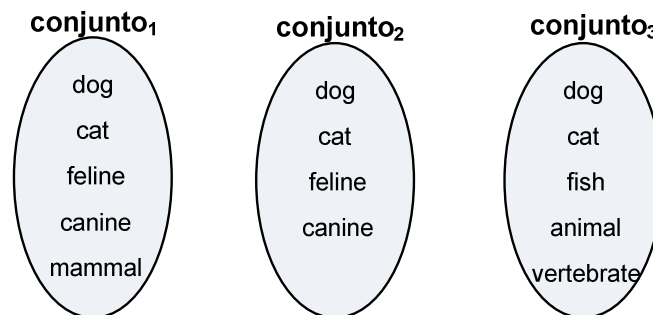


Figura 34 – Exemplo de Conjuntos de Termos.

**Definição 11** (*sim*<sub>4</sub> – *Sintática*): verifica a igualdade sintática entre dois termos *i* e *j*. Se forem iguais retorna 1, caso contrário retorna 0.

$$sim_4(\text{termo}_i, \text{termo}_j) = \text{igualdade}(\text{termo}_i, \text{termo}_j) \in [0,1]$$

**Definição 12** (*sim*<sub>5</sub> – *Descrições Textuais*): calcula a similaridade entre descrições textuais após a lematização de todas as palavras e remoção das *stopwords*. Para o cálculo são criados dois vetores de textos *i* e *j* referente a descrições textuais. A similaridade entre os dois vetores é calculada pelo número de palavras comuns dividido pelo número total de palavras sem contar as repetições.

$$sim_5(texto_i, texto_j) = \frac{(texto_i \cap texto_j)}{(texto_i \cup texto_j)} \in [0 \dots 1]$$

Para exemplificar a similaridade utilizou-se a Figura 35 onde há três descrições textuais que passaram por um processo de lematização e remoção das *stopwords*. A similaridade foi calculada da descrição textual 1 ( $texto_1$ ) para as descrições textuais 2 ( $texto_2$ ) e 3 ( $texto_3$ ). Na similaridade entre as descrições textuais 1 e 2, há 6 termos em comum e 18 termos não repetidos. Logo, a similaridade é de 0,33 (6/18). Na similaridade entre as descrições textuais 1 e 3, há 2 termos em comum e 23 termos não repetidos. Logo, a similaridade é de 0,08. Portanto, a descrição textual 1 é mais similar a descrição textual 2.

**Descrição textual 1**

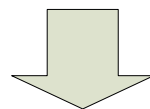
A building where travelers can pay for lodging and meals (restaurants) and other services (childcare, meeting rooms)

**Descrição textual 2**

An establishment that provides paid lodging, a number of additional guest services such as restaurants, a swimming pool or childcare for travelers.

**Descrição textual 3**

Provide accommodation where guests can rent a bed, in a dormitory and share a bathroom, lounge and sometimes a kitchen, overnight lodging for travelers



Após a lematização e remoção de stopwords

**Descrição textual 1 (11 termos)**

build traveler can pay lodge meal restaurant service childcare meet room

**Descrição textual 2 (13 termos)**

establishment provide **pay lodge** number addition guest **service restaurant** swim pool **childcare traveler**

**Descrição textual 3 (14 termos)**

provide accommodation guest **can** rent bed dormitory share bathroom lounge kitchen overnight lodge **traveler**

Figura 35 – Exemplos de Descrições Textuais após Lematização e Remoção das *Stopwords*.

**Definição 13** ( $sim_6$  – *Semantic Cotopy*): calcula a similaridade semântica entre dois conceitos em uma taxonomia. A  $sim_6$  calcula a sobreposição entre as duas taxonomias partindo-se de



dois conceitos  $i$  e  $j$ . A  $SC$  (*semantic cotopy*) (MAEDCHE; STAB, 2002) forma conjuntos de conceitos a partir das relações hierárquicas dos conceitos  $i$  e  $j$ .

$$sim_6 = \frac{|SC(\text{conceito}_i, \text{taxonomia}_i) \cap SC(\text{conceito}_j, \text{taxonomia}_j)|}{|SC(\text{conceito}_i, \text{taxonomia}_i) \cup SC(\text{conceito}_j, \text{taxonomia}_j)|}, \text{ sendo:}$$

- $\text{taxonomia } i \text{ e } j$  construída pelos superconceitos e subconceitos dos conceitos  $i$  e  $j$ .

A  $SC$  compara nodos pelas relações de hierarquia.  $SC$  é definida:

$$SC(C_i, H_c) = \{C_j \in C | H_c(C_i, C_j) \vee H_c(C_j, C_i)\}, \text{ sendo:}$$

- $H_c(C_i, C_j)$  significa que  $C_i$  é um subconceito de  $C_j$ ;
- $H_c(C_j, C_i)$  significa que  $C_j$  é um subconceito de  $C_i$ .

Para exemplificar o cálculo, utilizaram-se as taxonomias ilustradas na Figura 36. No exemplo, calcula-se a  $SC$  dos conceitos 1 ( $\text{conceito}_1$ ) e 15 ( $\text{conceito}_{15}$ ). A similaridade entre as duas taxonomias é de 0,42. Quanto mais próximo de 1 mais similar as taxonomias em que os conceitos fazem parte.

$$\begin{aligned} M_6 &= \frac{SC(\text{conceito}_1, \text{taxonomia}_1) \cap SC(\text{conceito}_{15}, \text{taxonomia}_{15})}{SC(\text{conceito}_1, \text{taxonomia}_1) \cup SC(\text{conceito}_{15}, \text{taxonomia}_{15})} \\ &= \frac{(\text{conceito}_7, \text{conceito}_2, \text{conceito}_8, \text{conceito}_{10}) \cap (\text{conceito}_7, \text{conceito}_8, \text{conceito}_{10})}{(\text{conceito}_7, \text{conceito}_2, \text{conceito}_8, \text{conceito}_{10}) \cup (\text{conceito}_7, \text{conceito}_8, \text{conceito}_{10})} \\ &= \frac{(\text{conceito}_7, \text{conceito}_8, \text{conceito}_{10})}{(\text{conceito}_7, \text{conceito}_2, \text{conceito}_8, \text{conceito}_{10})} = \frac{3}{4} = 0,75 \end{aligned}$$

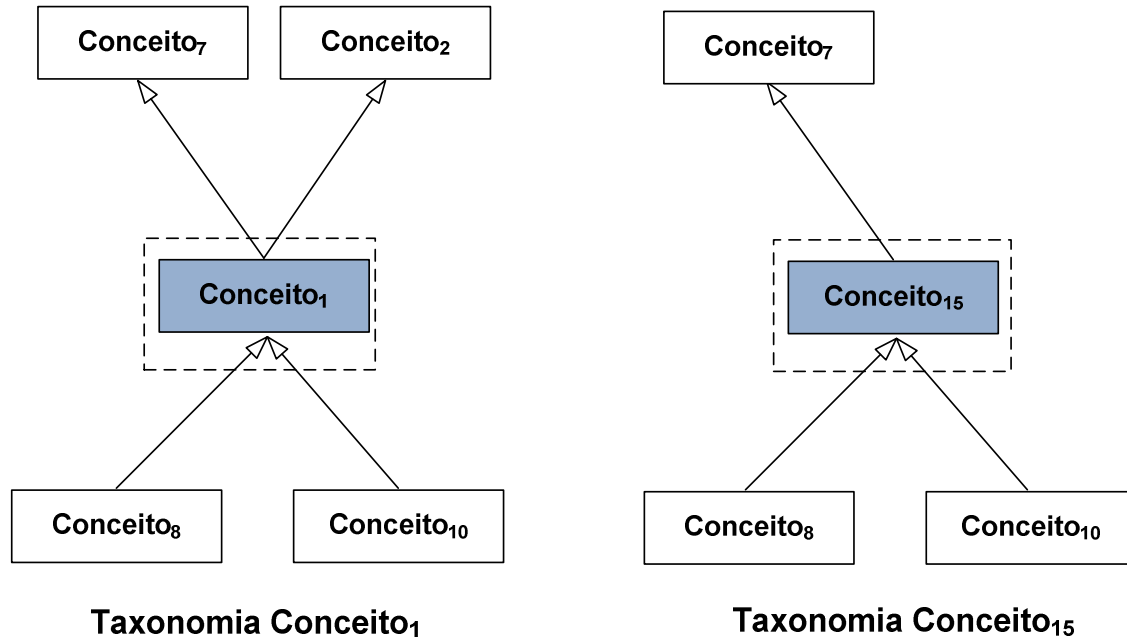


Figura 36 – Exemplos de Taxonomias.

A taxonomia de tags é produzida por FCA com base na correlação de tags utilizadas no processo de anotação.

### 2.6.1 Análise Formal de Conceitos

A análise formal de conceitos (FCA – *Formal Concept Analysis*) (WILLE, 1982) é um modelo matemático baseado na teoria do reticulado e na teoria dos conjuntos ordenados. A FCA relaciona objetos a atributos a partir de uma tabela de contexto. Esta tabela contém os objetos  $G$ , atributos  $M$  e a relação  $I$  entre os mesmos. Define-se contexto formal (*Formal Context*).

**Definição 14 (Formal Context)** (WILLE, 1982) – Um contexto formal é uma tripla  $\mathbb{K} = \langle G, M, I \rangle$ , sendo:

- $G$ : conjunto  $G$  de objetos.
- $M$ : conjunto  $M$  de atributos.
- $I$ : relação binária  $I \subseteq G \times M$ ,  $(g, m) \in I$  (objeto  $g$  tem o atributo  $m$ ).

Os contextos formais delimitam o conjunto de objetos, conjunto de atributos e as relações entre atributos e objetos. Um contexto formal é representado por uma tabela cujas

linhas representam objetos, colunas representam atributos e a interseção entre linhas e colunas representam uma relação binária ou de incidência. Esta relação determina se um objeto  $g \in G$  e possui um atributo  $m \in M$ . A Tabela 2 exemplifica uma tabela de contexto formal.

Tabela 2 – Exemplo de uma Tabela de Contexto.

		<b>Atributos</b>				
		$m_1$	$m_2$	$m_3$	$m_4$	$m_5$
<b>O b j e t o s</b>	$g_1$		x	x	x	
	$g_2$	x	x	x	x	
	$g_3$	x	x	x	x	
	$g_4$	x			x	
	$g_5$		x	x	x	x
	$g_6$				x	x
	$g_7$	x	x	x	x	x

A partir da tabela de contexto formal identificam-se os conceitos formais. Um conceito formal é representado pelo o conjunto de atributos  $M$  e pelos objetos  $G$  que compartilham tais atributos. Um conceito formal é uma relação máxima de objetos que compartilham os mesmos atributos. Por exemplo, os objetos  $g_2, g_3$  com os atributos  $m_1, m_2, m_3, m_4$  representam um conceito formal. Define-se FCA:

**Definição 15 (FCA – Formal Concept Analysis)** (WILLE, 1982) – Um par  $(A, B)$  é um conceito formal de  $\langle G, M, I \rangle$ , se e somente se  $A \subseteq G$  (um subconjunto de objetos),  $B \subseteq M$  (um subconjunto de atributos),  $A' = B$  e  $A = B'$ , ou seja,  $(A, B)$  é um conceito formal se o conjunto de todos os atributos compartilhado pelos objetos de  $A$  é idêntico a  $B$ , ou,  $A$  também o conjunto de todos os objetos que possuem todos os atributos em  $B$ .  $A'$  é a intenção do conceito e  $B'$  a extensão do conceito. Define-se  $A'$  (conjunto de atributos comuns aos objetos de  $A$ ) e  $B'$  (conjunto de objetos que possuem os atributos de  $B$  em comum):

$$A' = \{m \in M | \forall g \in A, (g, m) \in I\}$$

$$B' = \{g \in G | \forall m \in B, (g, m) \in I\}$$

Utilizando a Tabela 2, o par de objetos  $g_2, g_3$  é um conceito formal se, e somente se,  $A=B'$  e  $B=A'$ . Assim por exemplo, o par  $(\{g_2, g_3\}, \{m_1, m_2, m_3, m_4\})$  é um conceito formal

derivado do contexto formal da Tabela 2. A intenção ou  $A'$  de  $(\{g_2, g_3\})$  é igual a  $\{m_1, m_2, m_3, m_4\}$  e a extensão de  $B'$  ou  $\{m_1, m_2, m_3, m_4\}$  é  $\{g_2, g_3\}$ .

O conjunto de conceitos formais associados é um reticulado conceitual. O reticulado conceitual pode ser representado através de um diagrama de linha ou por uma hierarquia criando-se a relação subconceito/superconceito. Essa hierarquia determina que para dois conceitos formais  $(X, Y)$ ,  $X$  subsume  $Y$  se o conjunto de atributos  $X$  está contido em  $Y$ .

Para exemplificar a criação da hierarquia, utilizou-se a Tabela 2. A Figura 37 ilustra uma hierarquia de conceitos formais. Na Figura 37 cada conceito formal é representado por um retângulo que contém os atributos e objetos. Por exemplo, na Figura 37 o conceito formal  $(\{g_1\}, \{m_2, m_3, m_4\})$  subsume o conceito formal  $(\{g_5\}, \{m_2, m_3, m_4, m_5\})$  porque este último herda os atributos  $\{m_2, m_3, m_4\}$  do mesmo, conforme tabela de contexto.

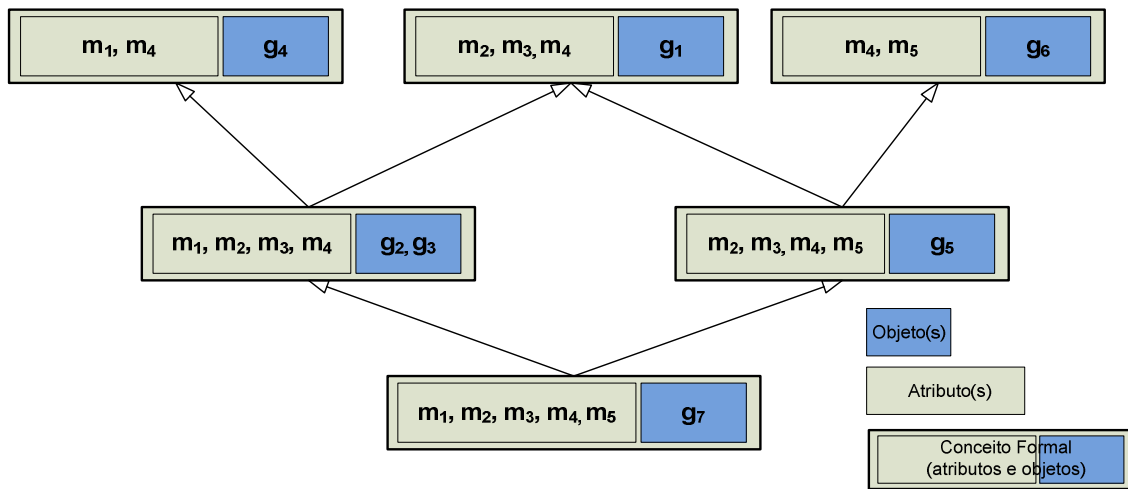


Figura 37 – Hierarquia de Conceitos Formais (Atributos e Objetos) obtida pela FCA.

### 2.6.2 Cálculo da Dissimilaridade

As similaridades entre duas entidades  $e$  ( $e_1, e_2$ ) formam um conjunto de vetores ( $v_1, v_2$ ), ou seja, um vetor para cada entidade  $e$ . Cada elemento (linha) do vetor contém a aplicação de uma das medidas de similaridades ( $sim_1, \dots, sim_6$ ) descritas acima. A partir dos vetores ( $v_1, v_2$ ) calcula-se a distância entre duas entidades. Quanto mais próximo de 0 mais similares as duas entidades. Define-se distância:

**Definição 16 (distância):** a medida é uma função  $dist: \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0,1]$ , sendo:

- $\forall v_1, v_2 \in \mathbb{R}^n, dist(v_1, v_2) \geq 0$ : vetores com alguma similaridade;
- $\forall v_1, v_2 \in \mathbb{R}^n, dist(v_1, v_2) = 1$ : vetores diferentes (não há semelhanças);
- $\forall v_1, v_2 \in \mathbb{R}^n, dist(v_1, v_2) = 0$ : vetores com similaridade total (não há distância).

A função de distância pode ser substituída pela distância euclidiana definida a seguir:

**Definição 17 (distância euclidiana):** a distância euclidiana entre os vetores  $v1 (v1_1, \dots, v1_j)$  e  $v2 (v2_1, \dots, v2_j)$  é definida como:

$$\sqrt{(v1_1 - v2_1)^2 + (v1_2 - v2_2)^2 + \dots + (v1_j - v2_j)^2} = \sqrt{\sum_{i=1}^j (v1_i - v2_i)^2} ,$$

sendo:

- $v1_i$ : o valor máximo de similaridade que pode-se obter;
- $v2_i$ : o valor obtido das medidas de similaridade;
- $j$ : o número de medidas de similaridades utilizadas.



### 3. MÉTODO PROPOSTO: FOLKONCEPT

Este capítulo descreve o método de suporte à atividade de conceitualização a partir de folksonomias denominado Folkconcept (*Folksonomy + Conceptualization*). A Figura 38 ilustra a abordagem apresentada neste trabalho. Parte-se de uma **folksonomia inicial** e de uma especificação inicial do modelo conceitual denominada **ontologia raiz** de agora em diante.

A ontologia raiz é uma especificação conceitual desenvolvida inicialmente pelos projetistas a partir de uma folksonomia inicial resultante do processo de anotação realizado primordialmente pelos especialistas do domínio (outros atores também podem participar). Salienta-se que o termo ontologia raiz não é utilizado como sinônimo de ontologia de nível superior. Raiz apenas indica que a ontologia é a primeira versão da especificação conceitual gerada na atividade de conceitualização.

Para integrar a aquisição de conhecimentos à conceitualização e registrar esta relação, utiliza-se uma metapropriedade denominada *referenceTag*. Os elementos da ontologia raiz estão conectados a *tags* da folksonomia inicial pela metapropriedade *referenceTag*. Assim, há uma relação bidirecional de rastreabilidade entre elemento da ontologia e a *tag* da folksonomia que lhe deu origem. A *tag*, por sua vez, é descrita com base numa ontologia e apresenta várias propriedades, tais como sinônimos e URL (*Uniform Resource Locator*) do recurso anotado.

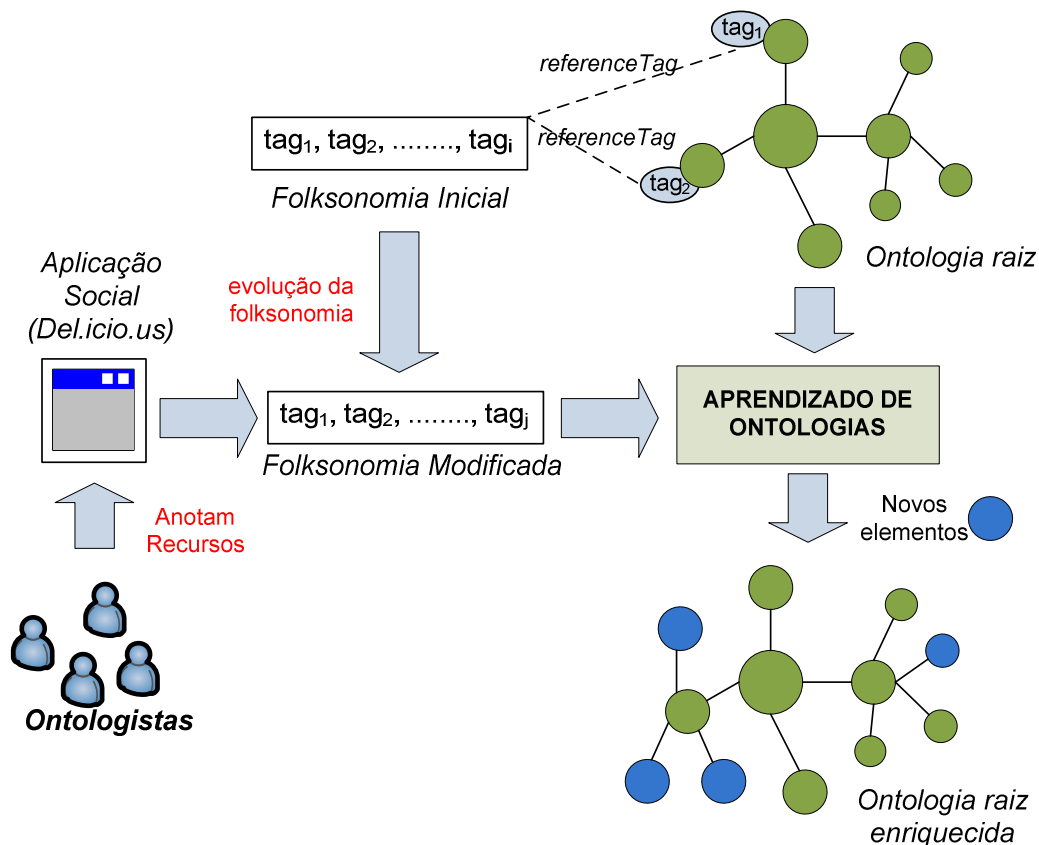


Figura 38 – Fluxo do Método Desenvolvido para o Desenvolvimento de uma Ontologia.

Os atores ao utilizarem uma aplicação social (*e.g. Del.icio.us*) para marcar termos importantes do domínio modificam a folksonomia inicial. Esta folksonomia modificada e a ontologia raiz são entradas para o aprendizado de ontologias. Este último possui como objetivo gerar novos elementos na ontologia raiz e nas versões posteriores. Logo, os conhecimentos adquiridos da folksonomia produzem alterações na ontologia, mais precisamente, enriquecem a ontologia raiz.

A ontologia raiz e as novas versões da ontologia são artefatos resultantes de iterações sucessivas da atividade de conceitualização. A cada novo ciclo, os conhecimentos adquiridos da folksonomia são conceitualizados de forma coerente utilizando um processo de avaliação na ontologia a ser desenvolvida. A cada novo ciclo uma nova versão da ontologia raiz é criada.

A Figura 39 destaca as atividades abordadas neste trabalho. Formalização e implementação são outras atividades que dão sequência ao processo de desenvolvimento, mas estão fora do escopo desta tese.



Como uma ontologia pode ser desenvolvida por sucessivas iterações e por sucessivos incrementos, as atividades podem ser executadas mais de uma vez. A cada iteração novos conhecimentos são adquiridos e conseqüentemente novos elementos são adicionados na ontologia. As seções seguintes detalham as atividades de aquisição de conhecimentos e de conceitualização.

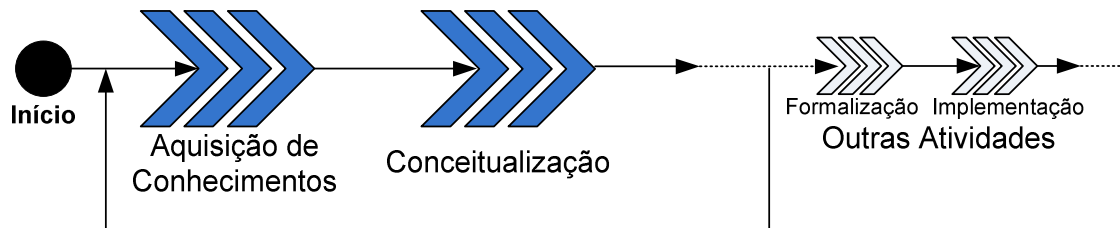


Figura 39 – Atividades de Aquisição de Conhecimentos e Conceitualização abordadas neste Trabalho para o Desenvolvimento de uma Ontologia.

### 3.1 AQUISIÇÃO DE CONHECIMENTOS

No Folkconcept, a atividade de aquisição de conhecimentos no método é feita de maneira semiautomática, a partir da anotação de documentos eletrônicos utilizados pelos especialistas do domínio. A aquisição de conhecimentos proposta possui as subatividades que incluem anotação de páginas *Web* e construção automática das anotações (Figura 33).

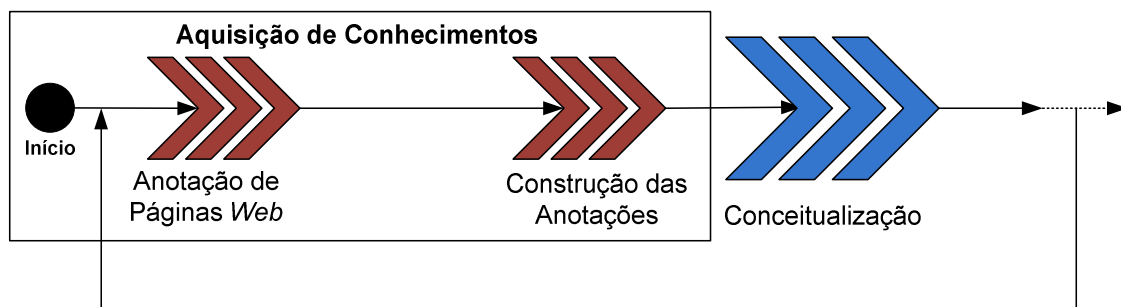


Figura 40 – Subatividades na Aquisição de Conhecimentos.

### 3.1.1 Anotação de Páginas Web

Neste trabalho, as anotações de páginas *Web* foram feitas por meio da aplicação *De.li.cio.us* (<http://delicious.com>). O *De.li.cio.us* é um gerenciador de *bookmarks*, o qual permite a um usuário categorizar e anotar conteúdos disponíveis na *Web* (recursos), facilitando assim a recuperação destes conteúdos pela comunidade de usuários.

O objetivo das anotações é fazer com que os especialistas do domínio explicitem conhecimento que julgam relevante para o domínio em questão e que estas anotações sejam utilizadas na conceitualização. Eventualmente, os outros atores do processo, projetistas e usuários, podem realizar anotações se possuírem conhecimento para tal, porém, de maneira primordial são os especialistas do domínio que as realizam. As anotações evoluem à medida que a comunidade de especialistas explicita seus conhecimentos.

A anotação consiste em atribuir livremente uma ou mais *tags* ao recurso representado por uma URL, descrever e incluir um título a este recurso. Define-se recurso:

**Definição 18 (Recurso)** - Um recurso  $r \in R$  (conjunto de recursos) é definido como:

$r = \langle url_r, descrição_r, título_r, tipo\_recurso_r \rangle$ , sendo:

- $url_r$ : endereço *Web* do recurso  $r$ ;
- $descrição_r$ : descrição textual do recurso  $r$ ;
- $título_r$ : resumo da descrição textual, geralmente título do endereço *Web* do recurso  $r$ .
- $tipo\_recurso_r$ : tipo do documento anotado (texto da *Web*, documento *pdf*, documento *txt*, entre outros).

Um recurso  $r \in R$  está inserido em uma ou mais anotações. Posiciona-se recurso na definição de anotação:

**Definição 19 (Anotação)** - Uma anotação  $a \in A$  (conjunto de anotações) é definida como:

$a = \langle an_a, r_a, \langle tagOriginal_1, \dots, tagOriginal_n \rangle_a, d_a \rangle$ , sendo:

- $an_a$ : anotador ou criador da anotação  $a$ ;
- $r_a$ : recurso  $r$  anotado em  $a$ ;

- $tagOriginal_1, \dots, tagOriginal_n$ : conjunto de *tags* utilizadas na anotação  $a$ , exatamente a *tag* original (escrita original, por exemplo, com espaços, hífen) atribuída pelo anotador à anotação  $a$ ;
- $n$ : número de *tags* utilizadas na anotação  $a$ ;
- $d_a$ : data da realização da anotação  $a$ .

As folksonomias são produzidas a partir do conjunto de anotações. Pela definição original de folksonomia (HOTH0 et al., 2006), posicionam-se as definições de recurso (Definição 18) e anotação (Definição 19), redefinindo-se folksonomias:

**Definição 20 (Folksonomia)** - Uma folksonomia  $\mathbb{F}$  é uma quádrupla  $\mathbb{F} = \langle An, T, R, D, Z \rangle$ , sendo:

- $An$ : conjunto finito de anotadores  $an$ ;
- $T$ : conjunto finito de *tags*  $tagOriginal_1, \dots, tagOriginal_n$ ;
- $R$ : conjunto finito de recursos  $r$  anotados;
- $D$ : conjunto finito de datas  $d$  das anotações;
- $Z$ : relação entre os elementos, i.e.,  $Z \subseteq An \times T \times R \times D$  (anotadores  $an \in An$  utilizam uma ou mais *tags*  $tagOriginal \in T$  para anotar recursos  $r \in R$  em uma data  $d \in D$ ).

Na Definição 20, a folksonomia é o resultado de um conjunto de anotações de vários anotadores. Uma folksonomia pode ser obtida para apenas um anotador, ou seja, do conjunto de anotações de um anotador. Neste caso, denomina-se folksonomia individual ou personomia (JÄSCHKE et al., 2008). O conjunto de personomias forma uma folksonomia (JÄSCHKE et al., 2008; GOLDER; HUBERMAN, 2005). A partir da definição original de personomia (JÄSCHKE et al., 2008; GOLDER; HUBERMAN, 2005), posicionam-se as definições de recurso (Definição 18) e anotação (Definição 19), redefinindo-se personomias:

**Definição 21 (Personomia)** - Uma personomia  $\mathbb{P}$  é uma quádrupla  $\mathbb{P} = \langle an, T, R, D, Z \rangle$ , sendo:

- $an$ : um anotador  $an$ ;
- $T$ : conjunto finito de *tags*  $tagOriginal_1, \dots, tagOriginal_n$  utilizadas pelo anotador  $an$  em suas anotações  $A$ ;
- $R$ : conjunto finito de recursos  $r$  anotados em  $A$  pelo anotador  $an$ ;

- $D$ : conjunto finito de datas  $d$  das anotações  $A$  de  $an$ ;
- $Z$ : relação entre os elementos, i.e.,  $Y \subseteq an \times T \times R \times D$  (um anotador  $an \in An$  utiliza uma ou mais *tags*  $tagOriginal \in T$  para anotar recursos  $r \in R$  em uma data  $d \in D$ ).

### 3.1.2 Construção das Anotações

A subatividade de construção das anotações inclui a recuperação, o enriquecimento e a representação das anotações em uma ontologia de representação de anotações (Figura 41).



Figura 41 – Construção das Anotações na Aquisição de Conhecimentos.

O algoritmo descrito na Figura 42 descreve a construção das anotações. O conjunto de anotações  $a$  ( $a_1, \dots, a_n$ ) é recuperado para um determinado anotador  $an \in An$  no *De.li.cio.us*. Cada anotação  $a$  é enriquecida, produzindo uma anotação enriquecida  $ae$ . Toda anotação enriquecida  $ae$  é representada em uma ontologia de representação de anotações.

```

construção_das_anotações ( $an$ )
início
  // recupera o número de anotações do anotador  $an$ 
  número_anotações  $\leftarrow$  recupera_número_anotações_anotador( $an$ );

  //para cada anotação  $a$  do anotador  $an$ 
  Para índice de 1 até número_anotações faça

    // recupera a anotação  $a$  pelo índice
     $a[\text{índice}] \leftarrow$  recupera_anotação( $\text{índice}$ );

    // enriquece a anotação  $a$ 
     $ae[\text{índice}] \leftarrow$  enriquece_anotação( $a[\text{índice}]$ );

    // representa a anotação enriquecida  $ae$ 
    representa_anotação( $ae[\text{índice}]$ );

  fimPara
fim

```

Figura 42 – Algoritmo para a Construção das Anotações na Aquisição de Conhecimentos.

A recuperação da anotação (*recupera\_anoação*) consiste em recuperar, conforme Definição 19, os recursos *r* anotados pelo anotador *an*, as *tags* originais  $tagOriginal_1, \dots, tagOriginal_n$  atribuídas aos recursos e as datas *d* da anotação dos recursos. Um recurso *r* (Definição 18) possui uma *descrição*, um *título*, uma URL e um tipo de recurso. O algoritmo visualizado na Figura 43 descreve a recuperação de uma anotação.

```

recupera_anoação(índice)
início

    descriçãoa ← recupera_descrição_textual(índice);
    títuloa ← recupera_título(índice);
    urla ← recupera_url(índice);
    da ← recupera_data(índice);

    // recupera tags da anotação a indexada por índice
    para n de 1 até numero_tags_anoação faça
        tag[n]a ← recupera_tag(índice);

    retorna a;

fim

```

Figura 43 – Algoritmo para Recuperação de uma Anotação.

O enriquecimento da anotação consiste no processamento léxico e sintático dos elementos da anotação (Definição 19), na inclusão de novos elementos na anotação pela relação com outras fontes de informação e pelo cálculo da correlação de *tags* no processo de anotação. O enriquecimento e inclusão destes novos elementos produzem uma anotação enriquecida *ae*. Define-se anotação enriquecida *ae*:

**Definição 22 (Anotação Enriquecida)** - Uma anotação  $ae \in AE$  (conjunto de anotações enriquecidas) é definida como:

$$AE_{ae} = \langle \begin{array}{l} an_{ae}, \Gamma_{ae}, \langle tagOriginal_1, \dots, tagOriginal_n \rangle_{ae}, d_{ae}, \\ \langle tagLema_1, \dots, tagLema_n \rangle_{ae}, \\ \langle tagRadical_1, \dots, tagRadical_n \rangle_{ae}, \\ \langle tagVariaçõesEscrita_1, \dots, tagVariaçõesEscrita_n \rangle_{ae}, \\ \langle tagClasseGramatical_1, \dots, tagClasseGramatical_n \rangle_{ae}, \\ \langle tagDescTextWordNet_1, \dots, tagDescTextWordNet_n \rangle_{ae}, \\ \langle tagMeronym_1, \dots, tagMeronym_p \rangle_{ae}, \\ \langle tagHypernym_1, \dots, tagHypernym_q \rangle_{ae}, \\ \langle tagHyponym_1, \dots, tagHyponym_r \rangle_{ae}, \\ \langle tagSinônimos_1, \dots, tagSinônimos_s \rangle_{ae}, \\ \langle tagAcrônimos_1, \dots, tagAcrônimos_t \rangle_{ae}, \\ \langle tagCorrelação_1, \dots, tagCorrelação_u \rangle_{ae} \end{array} \rangle, \text{ sendo:}$$

- *an*: anotador responsável pela anotação *a*.

- *r*: recurso anotado em *a* que corresponde aos elementos originais URL e *tipo\_recurso* e ao processamento textual da *descrição* e do *título*. Este processamento inclui a retirada das *stopwords* e lematização das palavras.
- *data*: data da anotação *a*.
- *tagOriginal*: *tags* originais atribuídas na anotação *a*.
- *tagLema*: lematização da *tag* original.
- *tagRadical*: radical da *tag* original.
- *tagVariaçõesEscrita*: possibilidades de escrita de uma *tag* composta que subdivide-se em *tag* escrita com hífen (*tag\_com\_hífen*), *tag* com sublinha (*tag\_com\_sublinha*) e *tag* sem espaços (*tag\_sem\_espacos*).
- *tagClasseGramatical*: classificação da *tag* como um substantivo, adjetivo, advérbio ou verbo;
- *tagDescTextWordNet*: descrição textual da *tag* original obtida na *WordNet*;
- *tagMeronym*, *tagHypernym*, *tagHyponym*: relações entre as *tags* obtidas na *WordNet*;
- *tagSinônimos*: conjunto de sinônimos das *tags* recuperado na *WordNet*;
- *tagAcrônimos*: siglas encontradas no dicionário de acrônimos (<http://www.all-acronyms.com/>);
- *tagCorrelação*: correlação de *tags* calculadas pela FCA (WILLE, 1982) que subdivide-se em *tags* acima da hierarquia (*tagCorrelação\_Super\_Tags*), *tags* abaixo da hierarquia (*tagCorrelação\_Sub\_Tags*) e *tags* na mesma hierarquia (*tagCorrelação\_Tags\_Irmãs*).

Na Definição 22 de *ae*, *n* é o número ou índice das *tags* utilizadas na anotação. Portanto, os novos elementos que possuem índice *n* têm uma relação de um para um com cada *tag*. Por exemplo, a *tag* original (*tagOriginal*) *computing* possui exatamente um radical (*tagRadical*) *comput*. Os outros índices (*p,q,r,s,t,u*) indicam que pode existir uma relação de um para vários, ou seja, uma *tag* pode ter, por exemplo, vários acrônimos.

Basicamente, a anotação enriquecida *ae* consiste no processamento léxico e sintático (remoção das *stopwords* e lematização das palavras) da *descrição* e do *título*, a inclusão de novas relações (*tagLema*, *tagRadical*, *tagVariaçõesEscrita*, *tagClasseGramatical*, *tagDescTextWordNet*, *tagMeronym*, *tagHypernym*, *tagHyponym*, *tagSinônimos*, *tagAcrônimos*) e o cálculo da correlação entre as *tags* (*tagCorrelação*). O algoritmo visualizado na Figura 44 descreve o processo de enriquecimento de uma anotação.

```

enriquece_annotação(a)
início

    // processamento textual da descrição e do título da anotação a
    descrição_processadaae ← proc_lex_sintático(descriçãoa, "lema", "stopwords");
    título_processadoae ← proc_lex_sintático(títuloa, "lema", "stopwords");

    // para cada tag da anotação a faça
    para n de 1 até numero_tags_annotação faça

        // identifica o lema da tag original
        tagLema[n]ae ← proc_lex_sintático(tagOriginal[n]a, "lema");

        // identifica o radical da tag original
        tagRadical[n]ae ← proc_lex_sintático(tagOriginal[n]a, "radical");

        // identifica variações de escrita da tag original
        tagVariaçõesEscrita[n]ae ← proc_lex_sintático(tagOriginal[n]a, "var_es");

        // recupera acrônimos da tag original
        tagAcrônimos[]ae ← descobrir_acrônimos(tagOriginal[n]a);

        // calcula a desambiguação das tags com todas as tags da anotação
        tagDes[n] ← desambiguação(tagOriginal[n]a, tagOriginal[]a);

        // recupera relações e outras informações da WordNet
        tagHyponyms[]ae ← relaçãoWordNet(tagDes[n], "hyponyms");
        tagMeronyms[]ae ← relaçãoWordNet(tagDes[n], "meronyms");
        tagSinônimos[]ae ← relaçãoWordNet(tagDes[n], "sinônimos");
        tagDescTextWordNet[n]ae ← relaçãoWordNet(tagDes[n], "descrição");
        tagClasseGramatical[n]ae ← relaçãoWordNet(tagDes[n], "classe");

        // calcula a correlação de tags pela FCA
        tagCorrelação[]ae ← cálculo_correlaçãoTags(tag[n]a);

    retorna ae;
fim

```

Figura 44 – Algoritmo para Enriquecimento da Anotação.

No algoritmo (Figura 44), a função *proc\_lex\_sintático()* remove *stopwords*, lematiza as palavras de uma descrição textual, resolve diferentes formas de escrita e encontra o radical das palavras. Utiliza-se uma parte específica do algoritmo KEA<sup>8</sup> (*Keyphrase Extraction Algorithm*) (WITTEN, 1999) para identificar o radical de uma palavra. O algoritmo KEA está implementado em linguagem Java e é distribuído como software livre. Na lematização utiliza-se a API (*Application Programming Interface*) *morphAdorner*<sup>9</sup>.

Este processamento é utilizado na escrita original das *tags*, no título e na descrição textual das anotações. Na descrição textual e no título apenas elimina as *stopwords* e lematiza todas as palavras do texto. Na escrita original das *tags*, recupera o radical, lematiza e resolve

<sup>8</sup> <http://www.nzdl.org/Kea/>

<sup>9</sup> <http://morphadorner.northwestern.edu/>

diferentes formas de escrita das *tags* compostas que incluem *tags* escritas sem espaços, com hífen ou sublinha. Na anotação de páginas *Web*, como não há regras de sintaxe para a escrita das *tags* e os usuários não conseguem construir *tags* compostas, os mesmos utilizam caracteres especiais (e.g “\_”, “-“) para a construção das *tags*, ou constroem *tags* compostas sem espaçamento.

A função *descobrir\_acrônimos()*, para cada *tag*, identifica quando houver um ou mais acrônimos e abreviaturas e siglas recuperadas do dicionário de acrônimos (<http://www.all-acronyms.com/>).

Quando uma *tag* tem mais de um significado, a função *desambiguação()* resolve a ambiguidade utilizando todas as *tags* atribuídas na anotação. A *tag* resolvida ou desambiguada (*tagDes*) é utilizada para recuperar outras informações na *WordNet* pela função *relaçãoWordNet()*. As seguintes informações são obtidas na *WordNet*: *meronym*, *hypernym*, *hyponym*, sinônimos, descrição textual da *tag* e classe gramatical.

O objetivo da função *cálculo\_correlaçãoTags(tag)* é correlacionar as anotações para construir uma hierarquia de *tags* para uma *tag* utilizando a FCA. Na correlação verificam-se quais as anotações apresentam quais *tags*. A Tabela 13 (i.e tabela de contexto) é utilizada para exemplificar o processo de correlação. Nela estão relacionadas as anotações com as respectivas *tags*. A Figura 45 ilustra a hierarquia de *tags* construída a partir da Tabela 13 utilizando-se da FCA.

Tabela 3 – Tabela de Contexto ou Correlação de *Tags* nas Anotações.

	<i>a</i> <sub>1</sub>	<i>a</i> <sub>2</sub>	<i>a</i> <sub>3</sub>	<i>a</i> <sub>4</sub>	<i>a</i> <sub>5</sub>	<i>a</i> <sub>6</sub>
<i>tag</i> <sub>1</sub>		x	x	x		
<i>tag</i> <sub>2</sub>	x	x	x	x		
<i>tag</i> <sub>3</sub>	x	x	x	x		
<i>tag</i> <sub>4</sub>	x			x		
<i>tag</i> <sub>5</sub>		x	x	x	x	
<i>tag</i> <sub>6</sub>				x	x	
<i>tag</i> <sub>7</sub>	x	x	x	x	x	
<i>tag</i> <sub>8</sub>				x	x	
<i>tag</i> <sub>9</sub>				x		x

Com base nas anotações, as *tags* podem ser posicionadas em uma relação hierárquica. Por exemplo, na Figura 45, a *tag*<sub>4</sub> possui as *tags* *tag*<sub>2</sub>, *tag*<sub>3</sub> e *tag*<sub>7</sub> acima de sua hierarquia, pois as mesmas estão presentes nas anotações *a*<sub>1</sub> e *a*<sub>4</sub>. A *tag*<sub>9</sub> é uma *tag* abaixo da



hierarquia da  $tag_4$ , pois está presente apenas na anotação  $a_1$ . Além disso, extraem-se dessa correlação as  $tags$  que são utilizadas nas mesmas anotações. Estas são consideradas irmãs. Por exemplo, as  $tags$   $tag_2$  e  $tag_3$  são utilizadas nas mesmas anotações, nesse caso  $a_1, a_2, a_3, a_4$ . Com base neste cálculo da correlação, a anotação é enriquecida com um novo elemento  $tagCorrelação$  que subdivide-se em:  $tagCorrelação\_Super\_Tags$  ( $tags$  acima da hierarquia),  $tagCorrelação\_Sub\_Tags$  ( $tags$  abaixo da hierarquia) e  $tagCorrelação\_Tags\_Irmãs$  ( $tags$  não relacionadas pela hierarquia, mas utilizadas em anotações semelhantes).

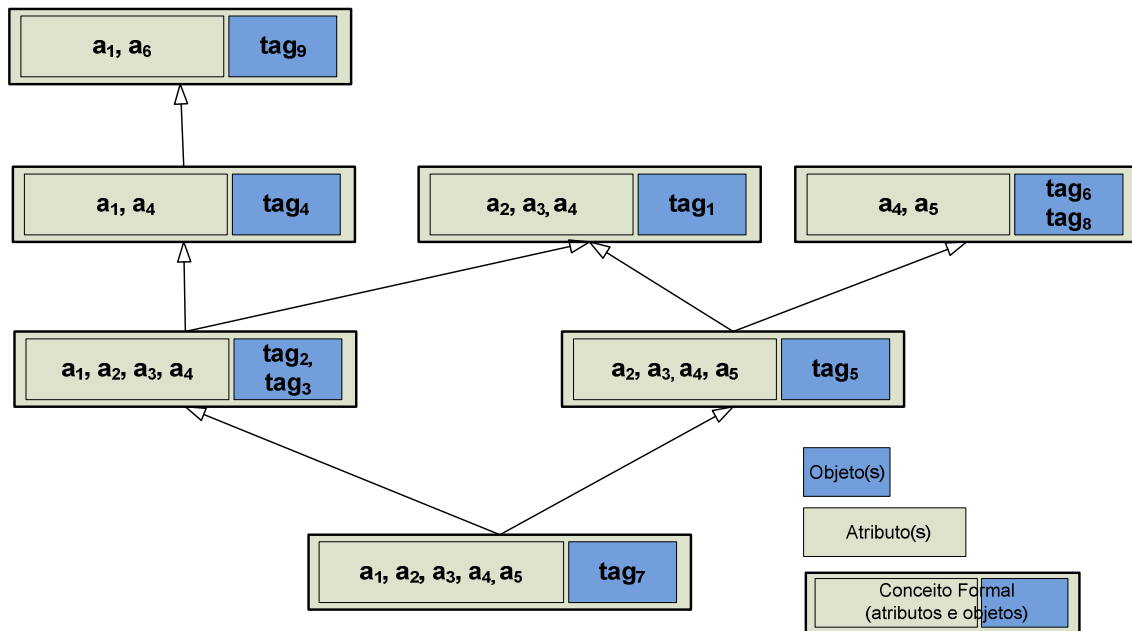


Figura 45 – Hierarquia obtida pela Correlação de *Tags* das Anotações.

Seguindo o algoritmo da Figura 42, após a recuperação e enriquecimento da anotação, toda anotação enriquecida  $ae$  deve ser representada ( $representa\_anotação()$ ) por uma ontologia. Para a representação das anotações em uma ontologia há diversos trabalhos (KNER, 2006; GRUBER, 2005; KIM et al., 2007). Todas as abordagens seguem as observações de Gruber (2005), ou seja, toda anotação deve incluir o recurso anotado, as  $tags$  utilizadas e o anotador (*i.e* usuário) que efetuou a anotação. A Figura 46 apresenta a ontologia construída neste trabalho para representar as anotações geradas por um ou mais projetistas. Do modelo original do Gruber (2005), surgiu a necessidade de criar outras relações e de descrever melhor os elementos. Isto deve-se ao fato do enriquecimento da anotação pelos processos de busca de outras informações no *WordNet* (meronímia, hiperonímia, hiponímia,

sinônimos), pelo cálculo da correlação de tags (utilizando-se do FCA) e também pelo processamento léxico e sintático (lematização, radicalização, remoção de stopwords). Além disso, no modelo original do Gruber, um recurso era especificado apenas por um identificador, sendo que, nesta tese, um recurso possui, além do identificador, um título, uma descrição textual e informações sobre o tipo do recurso anotado.

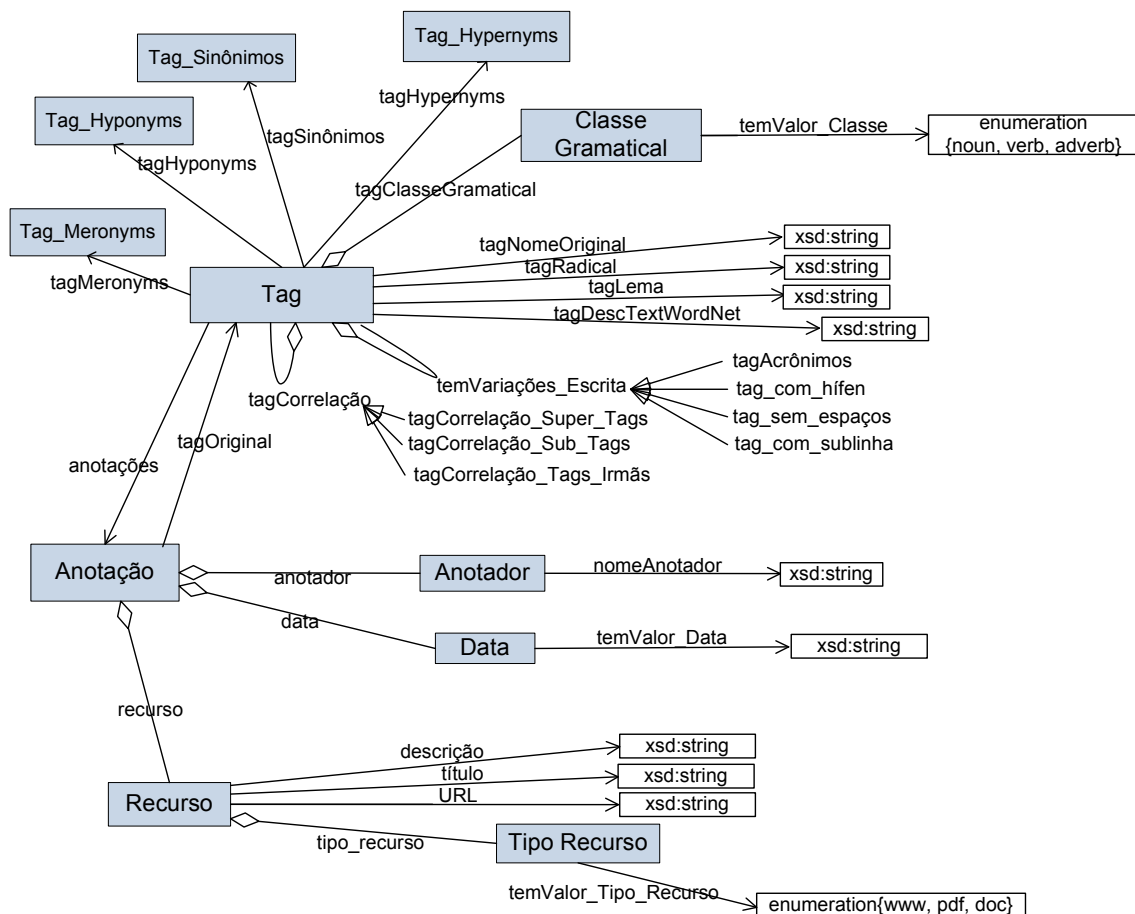


Figura 46 – Ontologia para Representação das Anotações.

O Quadro 8 descreve os conceitos e o Quadro 9 descreve as propriedades da ontologia de representação das anotações.

Quadro 8 - Descrição dos Conceitos da Ontologia para Representação das Anotações.

Conceito	Descrição
<i>Tag</i>	Palavras chave atribuídas pelo anotador às anotações.
<i>Anotação</i>	Categorização/Etiquetagem de um recurso.
<i>Recurso</i>	Objeto de categorização de um anotador, mais precisamente uma URL anotada.
<i>Tipo Recurso</i>	Tipo de documento apontado pela URL. Pode ser um documento <i>pdf</i> , <i>doc</i> ou uma página <i>Web</i> .

<i>Anotador</i>	Anotador responsável pela anotação.
<i>Data</i>	Data da anotação do recurso.
<i>Classe Gramatical</i>	Classe gramatical da <i>tag</i> utilizada na anotação. Pode ser um substantivo, verbo, advérbio ou adjetivo.

Quadro 9 - Descrição das Propriedades da Ontologia para Representação das Anotações.

Propriedade	Domínio	Imagem	Descrição
<i>tagNomeOriginal</i>	<i>Tag</i>	<i>xsd:string</i>	Rótulo textual de uma <i>tag</i> , da forma em que foi escrita originalmente pelo anotador.
<i>tagVariações_Escrita</i>	<i>Tag</i>	<i>Tag</i>	Alternativas de escrita que representam a mesma <i>tag</i> . Uma <i>tag</i> pode ser escrita como um acrônimo ou por palavras compostas. <i>Tags</i> compostas são escritas sem espaços entre as palavras e com a utilização de símbolos, que incluem, hífen ("-"), e sublinha ("_"). Por exemplo, um anotador anota um recurso com a <i>tag</i> composta "SemanticWeb". Esta <i>tag</i> pode ter outras variações de escrita como "Semantic Web" e "Semanti-Web".
<i>tagAcrônimos</i>	<i>Tag</i>	<i>Tag</i>	Subpropriedade da propriedade <i>temVariações_Escrita</i> . Representa uma sigla ou abreviatura de uma <i>tag</i> . Extraída de um dicionário de acrônimos.
<i>tag_com_hifen</i>	<i>Tag</i>	<i>Tag</i>	Subpropriedade da propriedade <i>temVariações_Escrita</i> . Variação com hífen de uma <i>tag</i> composta escrita originalmente. Ex, as <i>tags</i> "SemanticWeb" e "Semantic Web" são <i>tags</i> que podem ser escritas por hífen (e.g "Semanti-Web").
<i>tag_com_sublinha</i>	<i>Tag</i>	<i>Tag</i>	Subpropriedade da propriedade <i>temVariações_Escrita</i> . Variação com sublinha de uma <i>tag</i> composta escrita originalmente. Ex, as <i>tags</i> "SemanticWeb" e "Semantic-Web" são <i>tags</i> que podem ser escritas por uma sublinha (e.g "Semantic Web").
<i>tag_sem_espacos</i>	<i>Tag</i>	<i>Tag</i>	Subpropriedade da propriedade <i>temVariações_Escrita</i> . Variação sem espaços de uma <i>tag</i> composta escrita originalmente. Ex, as <i>tags</i> "Semantic-Web" e "Semantic Web" são <i>tags</i> que podem ser escritas sem a utilização dos símbolos hífen e sublinha (e.g "SemanticWeb").
<i>tagSinônimos (*)</i>	<i>Tag</i>	<i>Tag</i>	Sinônimos de uma <i>tag</i> extraídos da <i>WordNet</i> .
<i>tagClasseGramatical (*)</i>	<i>Tag</i>	<i>Classe Gramatical</i>	Classificação gramatical de uma <i>tag</i> como substantivo, adjetivo, verbo ou advérbio. Esta classificação é recuperada da <i>WordNet</i> .
<i>tagRadical (**)</i>	<i>Tag</i>	<i>xsd:string</i>	Elemento estrutural básico da <i>tag</i> (radical). Exprime o significado da <i>tag</i> mesmo sem o prefixo ou sufixo. O processo para encontrar o radical é conhecido como <i>stemming</i> . Por exemplo, as <i>tags</i> "computes", "computing" e "computed" são transformadas em "compute".
<i>tagLema (**)</i>	<i>Tag</i>	<i>xsd:string</i>	Substituição do sufixo da <i>tag</i> , ou seja, reduzir a <i>tag</i> a sua forma canônica (normalizada). Uma <i>tag</i> sem derivações ou flexões. Por exemplo, a <i>tag</i> "women" após a lematização é transformada em "woman".
<i>tagDescTextWordNet(*)</i>	<i>Tag</i>	<i>xsd:string</i>	Descrição textual da <i>tag</i> extraída da <i>WordNet</i> .
<i>tagMeronyms (*)</i>	<i>Tag</i>	<i>Tag</i>	Partes de um objeto representado pela <i>tag</i> . Por exemplo, alguns <i>meronyms</i> de "computer" são "central processing unit" e "computer circuit". Esta relação é recuperada da <i>WordNet</i> .
<i>temHypernyms (*)</i>	<i>Tag</i>	<i>Tag</i>	Palavras mais genéricas da <i>tag</i> . Por exemplo, um <i>hypernym</i> de "computer" é "device". Esta relação é recuperada da <i>WordNet</i> .
<i>tagHyponyms (*)</i>	<i>Tag</i>	<i>Tag</i>	Quais os objetos que a <i>tag</i> representa, mais precisamente, o inverso de <i>meronyms</i> . Por exemplo, um <i>hyponym</i> de "computer circuit" é "computer". Esta relação é recuperada da <i>WordNet</i> .
<i>tagCorrelação (***)</i>	<i>Tag</i>	<i>Tag</i>	Co-ocorrência de <i>tags</i> utilizadas nas mesmas anotações. Calculada pela FCA. Possui subpropriedades <i>tagCorrelação_tags_irmãs</i> , <i>tagCorrelação_super_tags</i> e <i>tagCorrelação_sub_tags</i> .
<i>tagCorrelação_tags_irmãs (***)</i>	<i>Tag</i>	<i>Tag</i>	Co-ocorrência de <i>tags</i> utilizadas nas mesmas

			anotações.
<i>tagCorrelação_super_tags</i> (***)	<i>Tag</i>	<i>Tag</i>	<i>Tags</i> que co-ocorrem em anotações. Relaciona uma <i>tag</i> a outra com relação a herança.
<i>tagCorrelação_sub_tags</i> (***)	<i>Tag</i>	<i>Tag</i>	<i>Tags</i> que co-ocorrem em anotações. Relaciona uma <i>tag</i> a outra com relação a herança.
<i>temValor_Classe</i>	<i>Classe Gramatical</i>	<i>xsd:string</i>	Possibilidades de classe gramatical: substantivo, adjetivo, verbo, advérbio.
<i>tagOriginal</i>	<i>Anotação</i>	<i>Tag</i>	Relaciona uma anotação com as <i>tags</i> utilizadas para categorizar um recurso. Mostra a co-ocorrência entre as <i>tags</i> na anotação de recursos. Propriedade inversa de <i>temAnotações</i> .
<i>anotações</i>	<i>Tag</i>	<i>Anotação</i>	Relaciona uma <i>tag</i> com uma ou mais anotações. Propriedade inversa de <i>temTags_Anotadas</i> .
<i>anotador</i>	<i>Anotação</i>	<i>Anotador</i>	Relaciona uma anotação ao anotador que categorizou um recurso.
<i>nomeUsuário</i>	<i>Usuário</i>	<i>xsd:string</i>	Nome do anotador que produz as anotações.
<i>descrição</i>	<i>Anotação</i>	<i>xsd:string</i>	Descrição textual da anotação. Esta descrição é feita pelo anotador.
<i>Título</i>	<i>Anotação</i>	<i>Data</i>	Identifica a data de criação da anotação pelo anotador.
<i>temValor_Data</i>	<i>Data</i>	<i>xsd:date</i>	Valor da data no padrão ("dd/mm/AA").
<i>recurso</i>	<i>Anotação</i>	<i>Recurso</i>	Identifica o recurso anotado pelo usuário. O recurso é representado por uma <i>URL</i> .
<i>Título</i>	<i>Anotação</i>	<i>xsd:string</i>	Pequena descrição textual da anotação. O título é descrito pelo anotador.
<i>URL</i>	<i>Recurso</i>	<i>xsd:string</i>	Possui o valor da <i>URL</i> .
<i>tipo_recurso</i>	<i>Recurso</i>	<i>xsd:string</i>	Identifica o tipo de documento que a <i>URL</i> aponta.
<i>temValor_Tipo_Recurso</i>	<i>Recurso</i>	<i>xsd:string</i>	Possui o valor do tipo de documento que inclui pdf, doc ou página <i>Web</i> .

\* as propriedades possuem valores recuperados da *WordNet*. A recuperação destas informações é realizada posteriormente a um processo de desambiguação de *tags*.

\*\* para a identificação do radical e do lema, utiliza-se o algoritmo KEA (*Keyphrase Extraction Algorithm*) (WITTEN et al., 1999).

\*\*\* o calculo FCA (WILLE, 1982) é realizado com base na co-ocorrência das *tags* nas anotações.

Após adquirir o conhecimento nas atividades de anotação e construção das anotações, estas devem ser articuladas na ontologia raiz. Esta articulação é realizada na atividade de conceitualização da ontologia que inclui o alinhamento de tags e a interação com o projetista.

### 3.2 CONCEITUALIZAÇÃO

No Folkconcept, a atividade de conceitualização sugere elementos ontológicos (*i.e.* conceitos, propriedades e instâncias) para o projetista a partir do conhecimento que foi capturado na atividade de aquisição de conhecimentos. O conhecimento explicitado pelo projetista na construção de suas anotações produz alterações na ontologia.

A atividade de conceitualização proposta infere novos elementos a partir de folksonomias e procura inseri-los na ontologia em desenvolvimento de maneira semiautomática (*i.e.* com a intervenção do projetista) e consistente (*i.e.* explorar as

consequências ontológicas da natureza ontológica dos elementos). A Figura 47 apresenta as subatividades da conceitualização: alinhamento de *tags* e interação com o projetista.

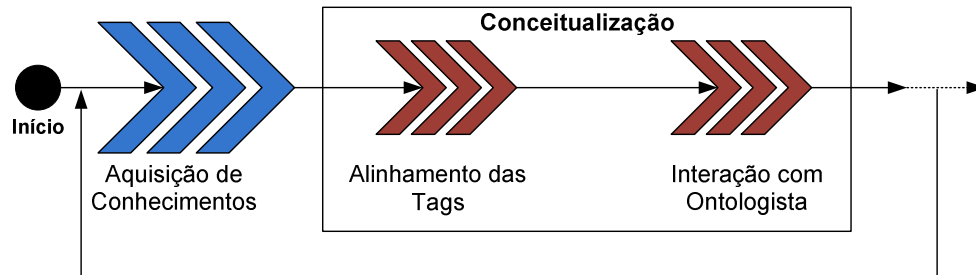


Figura 47 – Subatividades da Atividade de Conceitualização.

O alinhamento das *tags* alinha as *tags* da folksonomia antiga com as *tags* da nova folksonomia. Como as *tags* da folksonomia antiga estão associadas a elementos da ontologia raiz, o alinhamento acaba por posicionar os novos elementos na ontologia em desenvolvimento. O alinhamento é feito pelo método POAM (FREDDO et al., 2009; FREDDO; TACLA, 2009) alterado em relação às publicações originais para esta aplicação. No POAM original os conceitos são alinhados com base nas definições e conteúdo das propriedades. Nesta tese, utilizou-se o POAM apenas para alinhar *tags* com base no valor de suas propriedades.

A subatividade de interação com o projetista consiste na intervenção do mesmo no processo de identificação e classificação ontológica dos novos elementos na ontologia raiz. A Figura 48 posiciona as subatividades de alinhamento e interação com o projetista no processo de conceitualização.

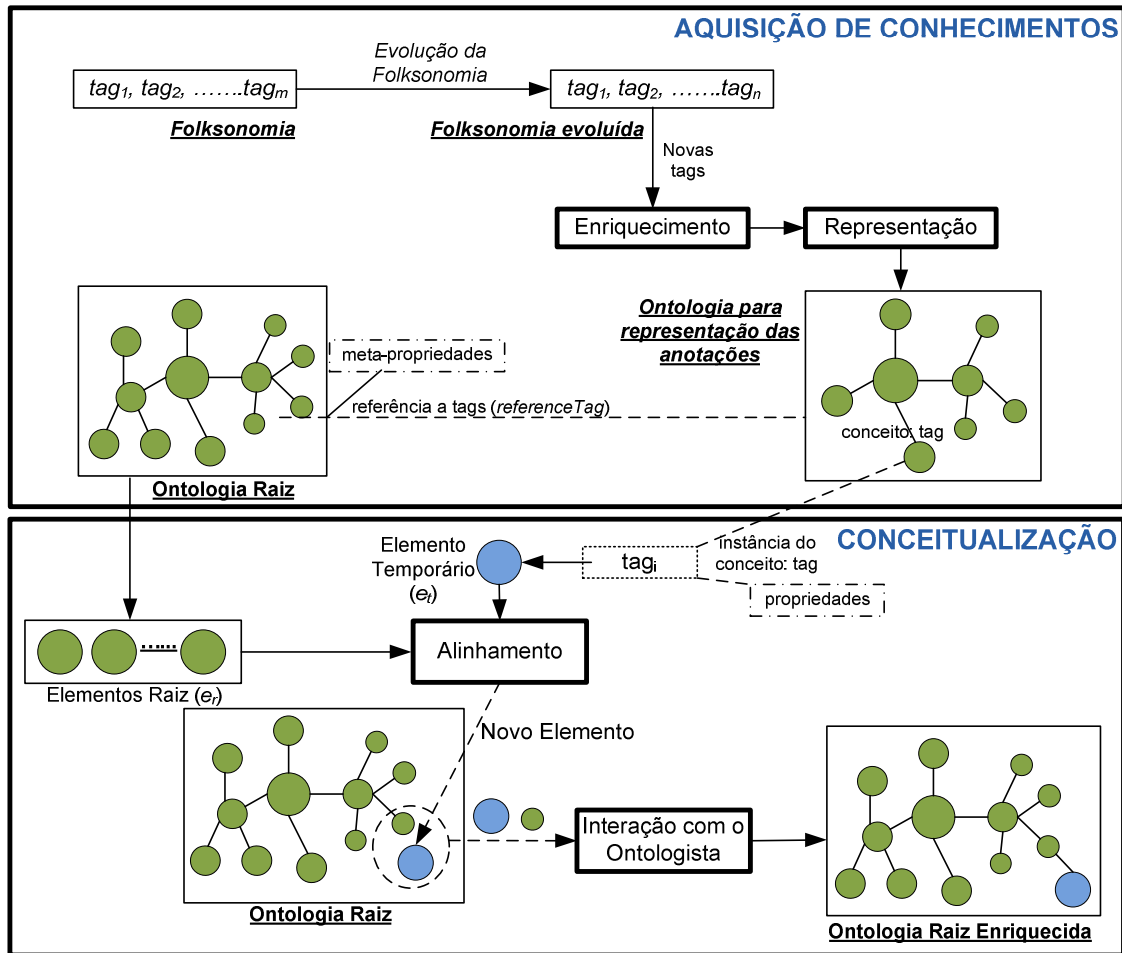


Figura 48 – Fluxo da Atividade de Conceitualização que inclui Alinhamento e Interação com o Projetista.

Conforme ilustrado na Figura 48, quando os especialistas realizam novas anotações, a folksonomia evolui. As novas *tags* inseridas pelos projetistas nestas anotações são recuperadas. Este processo consiste em recuperar as anotações, enriquecê-las e representá-las em uma ontologia de representação de anotações. Nesta ontologia, cada nova *tag* torna-se uma instância do conceito *tag* que, juntamente com os valores das suas propriedades, torna-se um elemento temporário  $e_t$  (Definição 24). Um  $e_t$  é alinhado com todos os elementos raiz  $e_r$  (Definição 23) da ontologia raiz.

Um  $e_r$  pode ser um conceito (elemento raiz conceito –  $e_{rc}$ ) ou uma instância (elemento raiz instância –  $e_{ri}$ ). O alinhamento das *tags* é realizado com base nos valores das propriedades do  $e_t$  e do  $e_r$ . Aos valores das propriedades, são aplicadas diversas medidas de similaridade (descritas na seção 2.3) que resultam um valor único de similaridade entre um  $e_t$  e todos os  $e_r$  existentes. Para cada  $e_t$  haverá um  $e_r$  mais similar. O alinhamento apenas indica

uma proximidade, o mesmo não define o tipo do elemento e nem qual a relação (taxonômica, propriedade, instância) do  $e_t$  com o  $e_r$ . Por exemplo, se o  $e_t$  é um conceito filho, irmão ou pai de algum dos  $e_r$  ou se o  $e_t$  é uma propriedade de algum  $e_r$ . As classificações do elemento e da relação são realizadas na interação com o projetista pela aplicação de um questionário eletrônico. No final deste questionário são sugeridas as alterações para o enriquecimento da ontologia raiz. Com vários questionamentos, o projetista é induzido a enriquecer a ontologia raiz.

As respostas do projetista ao questionário devem estar de acordo com o preconizado pela metodologia *OntoClean*, ou seja, não podem violar as restrições impostas pelas metapropriedades já associadas ao  $e_r$  na ontologia. Por exemplo, as metapropriedades rigidez, semirrigidez e antirrigidez atribuídas às primitivas de modelagem de uma ontologia impõem restrições na estrutura taxonômica da mesma. A análise destas restrições auxilia na avaliação e validação das escolhas de modelagem feitas durante o desenvolvimento. Outras formas de rigidez não utilizadas nesta tese, mas podem ser obtidas em Welty e Andersen (2005).

O questionário indaga o projetista guiando-o a escolher uma das metapropriedades (+R, -R, +I, -I, +O, -O, +D e -D) da *OntoClean* para o  $e_t$ . Para facilitar a utilização da *OntoClean*, o questionário apresenta exemplos pré-definidos advindos do domínio estudantil e inspirados dos exemplos dados por Guarino e Welty (2004). A seguir são apresentados detalhes do alinhamento de *tags* e da interação com o projetista.

**Definição 23 (Elemento Raiz):** um elemento raiz  $e_r$  é definido na ontologia raiz como um conceito ou instância rastreável a pelo menos uma *tag* da *folskonomia*:

$e_r \equiv (e_{rc} \sqcup e_{ri}) \sqcap (\exists referenceTag.tag)$ , sendo:

- $e_{rc}$  (elemento raiz conceito): um conceito na ontologia raiz;
- $e_{ri}$  (elemento raiz instância): uma instância na ontologia raiz;
- *referenceTag*: propriedade que faz referência a ontologia de representação de anotações;
- *tag*: conceito *tag* na ontologia de representação de anotações.

**Definição 24 (Elemento Temporário):** um elemento temporário  $e_t$  é definido como:

$e_t \equiv (e_{tc} \sqcup e_{tp} \sqcup e_{ti})$ , sendo:

- $e_{tc}$  (elemento temporário conceito): um conceito na ontologia raiz;
- $e_{tp}$  (elemento temporário propriedade): uma propriedade na ontologia raiz;

- $e_{ii}$  (elemento temporário instância): uma instância na ontologia raiz.

### 3.2.1 Alinhamento das *Tags*

O objetivo do alinhamento é identificar qual é o elemento raiz ( $e_r$ ) da ontologia raiz mais similar ao elemento temporário ( $e_t$ ). Esta similaridade é calculada pelas métricas de similaridade apresentadas na seção 2.8.

As medidas de similaridade para o cálculo do alinhamento são aplicadas para os valores das seguintes propriedades do conceito *tag*: *tagOriginal* (nesse caso recupera-se o valor da propriedade *nomeOriginal*), *tagLema*, *tagRadical*, *tagAcrônimos*, *tagVariaçõesEscrita* (que inclui *tag\_com\_hífen*, *tag\_sem\_espacos*, *tag\_com\_sublinha*), *tagSinônimos*, *anotações*, *tagHypernyms*, *tagMeronyms*, *tagHyponyms*, *tagClasseGramatical*, *anotador*, *tagDescTextWordNet*, *tagCorrelação\_Tags* (que inclui *tagCorrelação\_Sub\_Tags*, *tagCorrelação\_Super\_Tags*, *tagCorrelação\_Tags\_Irmãs*). Estas propriedades possuem como domínio o conceito *tag*. Com base na propriedade *anotações*, que relaciona uma *tag* a uma ou várias anotações, o método de alinhamento também recupera os valores das seguintes propriedades atribuídas a um recurso: *descrição* e *título*.

O pseudo-código do algoritmo de alinhamento é apresentado na Figura 49. O algoritmo de alinhamento utiliza diversas medidas de similaridade ( $sim_1, \dots, sim_6$ ) que produzem várias comparações ( $c_1, \dots, c_{21}$ ). As medidas ( $sim_1, \dots, sim_6$ ) retornam valores entre 0 e 1. Quanto mais próximo de um, maior a similaridade dos valores. As comparações correspondem ao valor obtido das medidas subtraído de um (quanto falta para obter o valor máximo de similaridade) e elevado a potência de dois. Esta operação é realizada para calcular a distância euclidiana entre os valores das distintas métricas de dimensão. A Figura 50 apresenta um refinamento do mesmo algoritmo.

```

alinhamento ( $e_t$ )
início
  Recupera os valores das propriedades do elemento temporário ( $e_t$ );
  Para cada elemento raiz ( $e_r$ ) referenciado faça {
    Recupera os valores das propriedades do elemento raiz ( $e_r$ );
    Executa as comparações ( $c_1 \dots c_{21}$ ) com as medidas de similaridade ( $sim_1 \dots sim_6$ );
    Calcula a distância euclidiana entre o elemento raiz ( $e_r$ ) e o temporário ( $e_t$ );
  }
  Retorna o elemento raiz ( $e_r$ ) com menor valor de distância;
Fim

```

Figura 49 – Pseudo-código do Algoritmo para o Alinhamento de *Tags*.



```

alinhamento(et)
início

//recupera o valor das propriedades para o elemento temporário (et)
prop_et[] ← recupera_valor_propriedades(et);

// comparar o elemento temporário (et) com cada elemento raiz (er) referenciado
para i de 1 até número_de_elementos_raiz faça
    se (er[i] referencia tags) então

        // recupera o valor das propriedades para o elemento raiz(er)
        prop_er[i] ← recuperar_valor_propriedades(er)

        // calcula o alinhamento
        c1 ← sim1 (prop_et.tagLema, prop_er[i].tagLema);
        c2 ← sim1 (prop_et.tagRadical, prop_er[i].tagRadical);
        c3 ← sim1 (prop_et.tagAcrônimos, prop_er[i].tagAcrônimos);
        c4 ← sim1 (prop_et.tag_com_hífen, prop_er[i].tag_com_hífen);
        c5 ← sim1 (prop_et.tag_sem_espacos, prop_er[i].tag_sem_espacos);
        c6 ← sim1 (prop_et.tag_com_sublinha, prop_er[i].tag_com_sublinha);

        c7 ← sim2 (prop_et.tagLema, prop_er[i].tagLema);
        c8 ← sim2 (prop_et.tagSinônimos, prop_er[i].tagSinônimos);

        c9 ← sim3 (prop_et.tagSinônimos, prop_er[i].tagSinônimos);
        c10 ← sim3 (prop_et.tagOriginal, prop_er[i].tagOriginal);
        c11 ← sim3 (prop_et.anotações, prop_er[i].anotações);
        c12 ← sim3 (prop_et.tagHyponyms, prop_er[i].tagHyponyms);
        c13 ← sim3 (prop_et.tagMeronyms, prop_er[i].tagMeronyms);
        c14 ← sim3 (prop_et.tagHypernyms, prop_er[i].tagHypernyms);

        c15 ← sim4 (prop_et.tagClasseGramatical, prop_er[i].tagClasseGramatical);
        c16 ← sim4 (prop_et.nomeGrupo, prop_er[i].nomeGrupo);
        c17 ← sim4 (prop_et.nomeUsuário, prop_er[i].nomeUsuário);

        c18 ← sim5 (prop_et.descricao, prop_er[i].descricao);
        c19 ← sim5 (prop_et.título, prop_er[i].título);
        c20 ← sim5 (prop_et.tagDescTextWordNet, prop_er[i].tagDescTextWordNet);
        c21 ← sim6 (prop_et.tagCorrelaçãoi, prop_er[i].tagCorrelaçãoj);
        distância[i] ←  $\sqrt{(1 - c_1)^2 + (1 - c_2)^2 + \dots + (1 - c_{21})^2}$ ;
    fim se
fim para

// recupera o índice com menor distância
índice ← Recupera_menor_distância(distância[]);

// retorna elemento raiz (er) mais similar
Retorna (er[índice]);

Fim

```

Figura 50 – Refinamento do Algoritmo para o Alinhamento das Tags.

No algoritmo da Figura 50, a função *recupera\_valor\_propriedades()*, recupera o valor das propriedades (*tagLema*, *tagRadical*, *tagAcrônimos*, *tag\_com\_hífen*, *tag\_sem\_espacos*, *tag\_com\_sublinha*, *tagLema*, *tagOriginal*, *anotações*, *tagMeronyms*, *tagHyponyms*, *tagHypernyms*, *tagClasseGramatical*, *nomeGrupo*, *nome\_Usuário*, *descricao*, *título*, *tagDescTextWordNet* e *tagCorrelação*) de um *e<sub>r</sub>* ou *e<sub>t</sub>*. Os valores das propriedades são armazenadas na variável *prop\_e<sub>r</sub>* para o *e<sub>r</sub>* e *prop\_e<sub>t</sub>* para o *e<sub>t</sub>*. A variável

*número\_de\_elementos\_raiz* indica o número de conceitos, propriedades e instâncias que possuem referência a uma ou mais *tags* na ontologia de representação de *tags*.

Recuperados os valores das propriedades para o  $e_r$  e para o  $e_t$ , comparam-se os valores utilizando as medidas de similaridade ( $sim_1, \dots, sim_6$ ). São geradas 21 comparações ( $c_1, \dots, c_{21}$ ).

A  $sim_1$  (Definição 8 - distância de *Levensthein*) gera seis comparações ( $c_1, \dots, c_6$ ) sendo aplicada nas seguintes propriedades: *tagLema* ( $c_1$ ), *tagRadical* ( $c_2$ ), *tagAcrônimos* ( $c_3$ ), *tag\_com\_hífen* ( $c_4$ ), *tag\_sem\_espacos* ( $c_5$ ), *tag\_com\_sublinha* ( $c_6$ ). As propriedades *tagAcrônimos*, *tag\_com\_hífen*, *tag\_sem\_espacos* e *tag\_com\_sublinha* fazem relação entre instâncias do mesmo conceito, nesse caso o conceito *tag*. Logo, recupera-se da instância do conceito *tag*, o valor da propriedade *tagNomeOriginal*. Quando as propriedades citadas acima possuem mais de um valor, comparam-se todos os elementos e retorna-se o mais similar.

A  $sim_2$  (Definição 9 – hierarquia na *WordNet*) produz duas comparações ( $c_7, c_8$ ) ao ser aplicada nas propriedades *tagLema* ( $c_7$ ) e *tagSinônimos* ( $c_8$ ). No caso da propriedade *tagSinônimos*, que também faz referência ao conceito *tag*, comparam-se todos os elementos (se houver mais de uma relação) e recupera-se o menor valor. A Figura 51 ilustra o algoritmo utilizado para a medida  $sim_2$ . No algoritmo, recupera-se na hierarquia da *WordNet* os significados (*i.e senses*) com os respectivos posicionamentos na taxonomia. Após, calcula-se a distância dos significados com relação ao termo comum mais próximo.

```

dist(termoi, termoj)
início

    // recupera os senses para os dois termos da ontologia de tags
    sense_termoi ← termoi.hasSense;
    sense_termoj ← termoj.hasSense;

    // recupera no WordNet a hierarquia taxonômica entre os dois senses
    taxonomia ← taxonomiaWordNet(sense_termoi , sense_termoj);

    // recupera na taxonomia o termo comum mais próximo aos dois senses
    termoComum ← termoComumTaxonomia(taxonomia, sense_termoi , sense_termoj);

    // calcula as distâncias
    di ← distância(termoComum, sense_termoi);
    dj ← distância(termoComum, sense_termoj);

    dist ← (di + dj);

fim

```

Figura 51 – Algoritmo para o Cálculo da  $sim_2$ .

A  $sim_3$  (Definição 10 – similaridade de conjuntos) produz seis comparações ( $c_9, \dots, c_{14}$ ) sendo aplicada nas seguintes propriedades: *tagSinônimos* ( $c_9$ ), *tagOriginal* ( $c_{10}$ ), *anotações*

( $c_{11}$ ), *tagHyponyms* ( $c_{12}$ ), *tagHypernyms* ( $c_{13}$ ), *tagMeronyms* ( $c_{14}$ ). Estas propriedades possuem um ou mais valores que constituem conjuntos. As propriedades *tagSinônimos*, *tagOriginal*, *tagHyponyms*, *tagHypernyms* e *tagMeronyms* fazem relação com instâncias do conceito *tag*. Para estas propriedades recupera-se o valor da propriedade *tagLema*. Para a propriedade *anotações* verifica-se o conjunto de anotações que utilizam as *tags*. A propriedade *tagOriginal* faz referência as *tags* que co-ocorrem na anotação de recursos.

A *sim<sub>4</sub>* (Definição 11 – igualdade de termos) gera três comparações ( $c_{15}, \dots, c_{17}$ ) ao ser aplicada nas propriedades seguintes: *tagClasseGramatical* ( $c_{15}$ ), *anotador (nomeUsuário)* ( $c_{16}$ ), *anotador (nomeGrupo)* ( $c_{17}$ ).

A *sim<sub>5</sub>* (Definição 12 – similaridade entre descrições textuais) aplicada nas seguintes propriedades do conceito *tag* e gera três comparações ( $c_{18}, \dots, c_{20}$ ): *descricao* ( $c_{18}$ ), *título* ( $c_{19}$ ) e *tagDescTextWordNet* ( $c_{20}$ ). Para as propriedades *tagDescTextWordNet* e *tagTítulo* pode haver mais de uma relação. Nesse caso são recuperadas todas as relações. Para a propriedade *tagDescTextWordNet* há apenas uma descrição textual que foi recuperada após o processo de desambiguação no enriquecimento da ontologia.

A *sim<sub>6</sub>* (Definição 13 - FCA) calcula a similaridade semântica entre dois conceitos em uma taxonomia. Neste caso, os conceitos são substituídos pelas *tags*. A taxonomia para as *tags* é construída pelas propriedades *tagCorrelação\_Sub\_Tags*, *tagCorrelação\_Super\_Tags*, *tagCorrelação\_Tags\_Irmãs* a partir da ontologia de representação de *tags*. Estas propriedades foram valoradas durante o cálculo da correlação no enriquecimento da ontologia de *tags* baseado nas anotações.

O resultado da subatividade de alinhamento é uma relação que apresenta para cada  $e_t$  um  $e_r$  mais similar (Figura 52). Esta relação é a entrada para a subatividade interação com o projetista, onde o  $e_t$  é alinhado ao  $e_r$  na ontologia raiz.

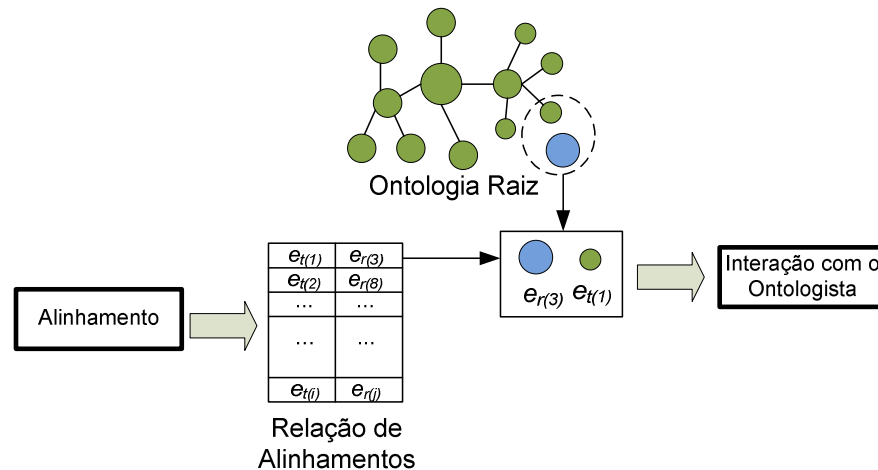


Figura 52 – Resultado da Subatividade de Alinhamento e Dados de Entrada para a Subatividade Interação com o Projetista.

### 3.2.1.1 Complexidade do Algoritmo de Alinhamento

No algoritmo ilustrado na Figura 50, se considerado apenas o laço principal, o mesmo apresenta complexidade linear  $O(|n|)$ , sendo  $n$  o número de elementos raiz ( $e_r$ ) a serem comparados com o elemento temporário ( $e_t$ ). No total, são realizadas 21 comparações ( $c_1, \dots, c_{21}$ ) que utilizam seis medidas de similaridade ( $sim_1, \dots, sim_6$ ). Para cada medida de similaridade há diferentes entradas que podem alterar a complexidade do algoritmo principal de alinhamento. Segue a descrição da complexidade de cada medida:

- $sim_1$  (*Levensthein*): utiliza-se uma matriz de  $j$  linhas e  $k$  colunas, onde  $j$  e  $k$  são respectivamente a quantidade de caracteres do primeiro e segundo termo. A complexidade do algoritmo de *Levensthein* é  $O(|j * k|)$ . As comparações ( $c_1, \dots, c_6$ ) utilizam a  $sim_1$ , sendo que o tamanho máximo tanto de  $j$  e como de  $k$  é da ordem de 20 caracteres.
- $sim_2$  (*hierarquia WordNet*): recupera-se o primeiro antepassado comum na hierarquia *WordNet* para os dois termos que serão comparados. Logo, para cada termo há um conjunto de termos recuperados da hierarquia taxonômica da *WordNet*. Entre os conjuntos, verifica-se qual o primeiro termo em comum. A complexidade desta medida é  $O(|l * m|)$ , sendo  $l$  e  $m$  respectivamente o número de termos do primeiro e segundo conjuntos. As comparações  $c_7$  e  $c_8$  utilizam a  $sim_2$ , sendo que  $l$  e  $m$  possuem um valor máximo determinado pela profundidade da hierarquia *WordNet*. A complexidade máxima de  $sim_2$  é  $O(|l * m|)$ .

- *sim3* (igualdade de conjuntos): compara-se todos os elementos de dois conjuntos. Semelhante a *sim2*, a complexidade também é definida pelo número de elementos dos dois conjuntos. A complexidade desta medida é  $O(|p * q|)$ , sendo  $p$  e  $q$  respectivamente o número de termos do primeiro e segundo conjuntos. As comparações ( $c_9, \dots, c_{14}$ ) utilizam a *sim2*. Porém, para cada comparação ocorre diferentes valores para  $p$  e  $q$ , sendo:
  - $c_9$ : número máximo de sinônimos obtidos na *WordNet*. A complexidade máxima de *sim2* é  $O(|p * q|)$ .
  - $c_{10}$ : número máximo de *tags* que se correlacionam nas anotações. A complexidade máxima de *sim2* é  $O(|p * q|)$ .
  - $c_{11}$ : número máximo de anotações que utilizam a mesma *tag*. A complexidade máxima de *sim2* é  $O(|p * q|)$ .
  - $c_{12}$ : número máximo de hipônimos da *WordNet* (profundidade máxima da *WordNet* subtraído de um). A complexidade máxima de *sim2* é  $O(|p * q|)$ .
  - $c_{13}$ : número máximo de hiperônimos da *WordNet* (profundidade máxima da *WordNet* subtraído de um). A complexidade máxima de *sim2* é  $O(|p * q|)$ .
  - $c_{14}$ : número máximo de merônimos da *WordNet* (relação todo/parte). A complexidade máxima de *sim2* é  $O(|p * q|)$ .
- *sim4* (igualdade sintática): A *sim4* é utilizada pelas comparações ( $c_{15}, \dots, c_{17}$ ). É apenas uma comparação de igualdade entre dois termos e efetua apenas uma comparação. Logo a complexidade é semelhante ao laço principal do algoritmo de alinhamento, sendo  $O(|r|)$ , onde  $r$  é a quantidade de elementos raiz.
- *sim5* (descrições textuais): utilizam-se os conjuntos de termos que aparecem em duas descrições textuais. Cada conjunto de termos possui um tamanho definido pela quantidade de termos presentes nas respectivas descrições textuais, sendo um título e uma descrição do recurso elaborado pelo projetista e uma descrição obtida da *WordNet*. A complexidade é semelhante a *sim2* e *sim3*. A complexidade desta medida é  $O(|s * t|)$ , sendo  $s$  e  $t$  respectivamente o número de termos do primeiro e segundo conjuntos. As comparações ( $c_{18}, \dots, c_{20}$ ) utilizam a *sim5*. Porém, para cada comparação ocorre diferentes valores para  $s$  e  $t$ , sendo:

- $c_{18}$ : número máximo de caracteres da descrição elaborada pelo projetista, neste caso, determina-se 50 caracteres. A complexidade máxima de  $sim_2$  é  $O(|50 * 50|)$ .
  - $c_{19}$ : número máximo de caracteres da descrição do título (resumo da descrição) elaborado pelo projetista, neste caso, determina-se 10 caracteres. A complexidade máxima de  $sim_2$  é  $O(|10 * 10|)$ .
  - $c_{20}$ : número máximo de caracteres da descrição obtida da WordNet, neste caso, determina-se 50 caracteres. A complexidade máxima de  $sim_2$  é  $O(|50 * 50|)$ .
- $sim_6$  (*semantic cotopy*): utiliza-se a *semantic cotopy* entre duas *tags* ( $tag_i, tag_j$ ) em uma dada hierarquia (calculada previamente pela FCA). Cada *tag* possui um conjunto de conceitos acima (superconceitos) e abaixo (subconceitos) da hierarquia. A divisão da interseção destes destes conjuntos representa a similaridade entre as tags na hierarquia. Similar as medidas  $sim_2, sim_3$  e  $sim_5$ , complexidade da  $sim_6$  é  $O(|v * w|)$ , sendo  $v$  e  $w$  respectivamente o numero de termos dos conjuntos da  $tag_i$  e  $tag_j$ .

A análise geral da complexidade do algoritmo é:

$$O(|j * k|) + O(|l * m|) + O(|p * q|) + O(r) + O(|s * t|) + O(O(|v * w|) + O(|x * z|)).$$

### 3.2.2 Interação com o Projetista

A interação com o projetista inicia-se a partir da relação dos  $e_t$  alinhados com os  $e_r$  na ontologia raiz. O alinhamento indica uma área da ontologia raiz que pode ser enriquecida pela conexão do  $e_r$  com o  $e_t$ . Um  $e_t$  compreende uma *tag* que pode vir a ser um conceito ou uma instância na ontologia raiz. Além disso, um  $e_t$  também pode auxiliar o projetista a fazer outras alterações na ontologia raiz. Por exemplo, em uma similaridade de um  $e_t$  com um conceito ( $e_r$ ), pode-se criar um novo conceito e uma nova relação entre  $e_t$  e o  $e_r$ . Logo, além do conceito também foi criada uma nova relação.

O objetivo da interação com o usuário é identificar o tipo ontológico (*e.g* conceito, instância) do  $e_t$  e que tipo de relação (*e.g* propriedade, taxonômica, instância) existe entre o  $e_t$  e o  $e_r$ .

Um questionário eletrônico é utilizado para fazer com que os projetistas auxiliados pelos especialistas do domínio identifiquem a natureza ontológica dos  $e_i(s)$  pelas metapropriedades da metodologia *OntoClean*. Pela natureza ontológica dos  $e_i(s)$  explora-se as consequências ontológicas da inserção do  $e_t$  na ontologia raiz. As respostas ao questionário eletrônico, que tornam transparente para os atores, a complexidade da metodologia *OntoClean*, classificam, direcionam e validam a inclusão dos novos elementos na ontologia raiz em desenvolvimento.

Com base nas respostas dos projetistas a cada pergunta do questionário e nas consequências ontológicas da *OntoClean*, a inclusão de um novo elemento segue um determinado fluxo. Ao final do fluxo são geradas sugestões de enriquecimento da ontologia raiz que podem ser aceitas ou recusadas.

Ao final da subatividade de interação com o projetista, uma nova ontologia raiz é gerada, sendo que os elementos desta ontologia estão conectados às *tags* fonte pela metapropriedade (*referenceTag*). Isto implementa a rastreabilidade, pois conecta a *tag* ao  $e_r$  na ontologia raiz (um  $e_t$  torna-se um  $e_r$ ). Com isso, uma nova evolução da folksonomia pode produzir uma nova versão da ontologia raiz. A Figura 53 ilustra o fluxo do questionário. O Quadro 10 apresenta resumidamente as questões presentes no questionário eletrônico. O questionário completo é apresentado no Apêndice A.

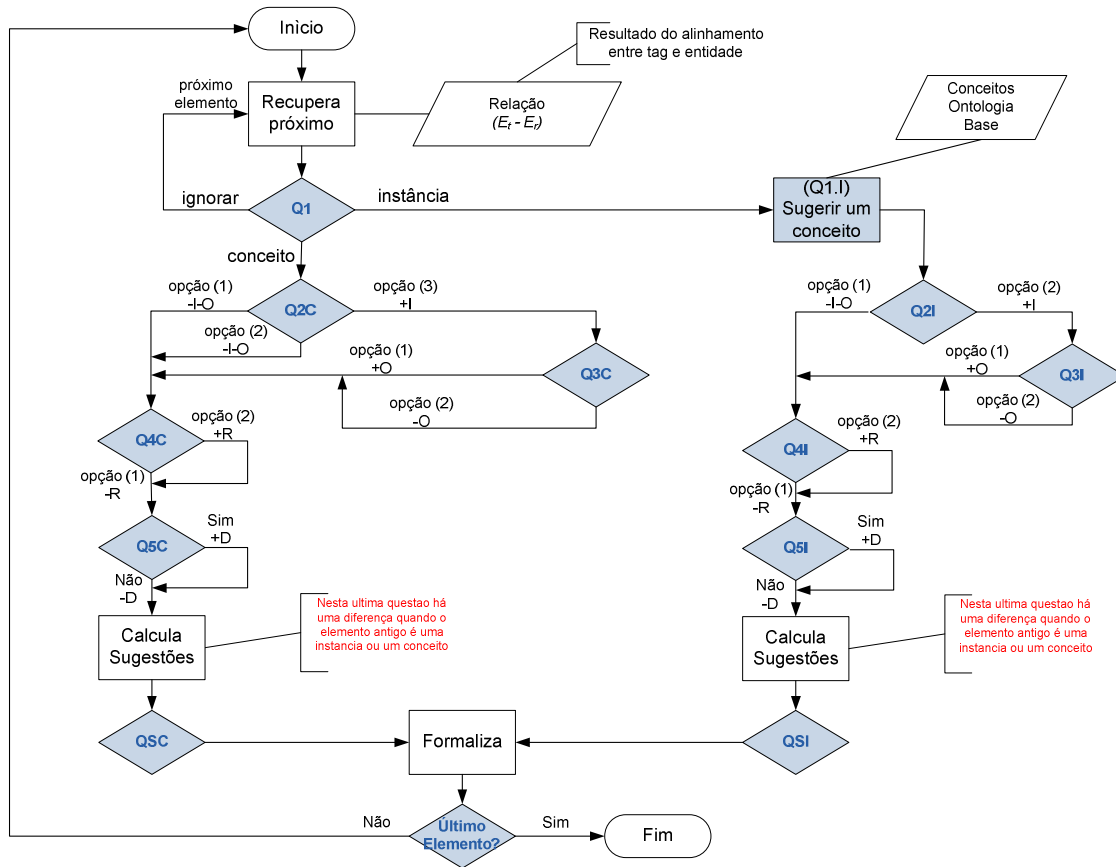


Figura 53 – Fluxo do Questionário na Interação com o Projetista.

Quadro 10 - Resumo das Questões do Questionário Eletrônico.

Identificação da Questão	Objetivo da Questão	Descrição da Questão
Q1	Classificar a <i>tag</i> como um conceito, instância ou ignorar.	A <i>tag</i> é um conceito, uma instância, ou ignore esta <i>tag</i> ?
Q2C	Classifica o conceito como +O (opção 1), -O/-I (opção 2), +I (opção 3).	(1) Conceito, cuja instâncias representam possíveis valores de um atributo. (2) Conceito que não possui instâncias, não possui atributos. Utilizado para organizar uma taxonomia. (3) Conceito que pode ser uma especialização de um conceito mais geral (herda atributos), ou conceito que fornece atributos a outros conceitos abaixo de sua hierarquia.
Q3C	Classifica o conceito como +O (opção 1), -O (opção 2).	(1) Conceito que apenas fornece atributos ou características para outros conceitos abaixo de sua hierarquia. (2) Conceito que herda atributos de outros conceitos e que também fornece atributos a outros conceitos.



<i>Q4C</i>	Classifica o conceito como <i>+R</i> (opção 1), <i>-R</i> (opção 2).	(1) Se as instâncias do conceito podem deixar de ser instâncias deste conceito, mas continuam a ser instâncias de um conceito mais geral. (2) As instâncias do conceito não deixam de ser instâncias deste conceito.
<i>Q5C</i>	Classifica o conceito como <i>+D</i> (ao responder Sim), <i>-D</i> (ao responder Não).	As instâncias do conceito possuem alguma relação de dependência extrínseca com instâncias de outro conceito?
<i>Q3C</i>	Sugestões calculadas com base nas restrições da <i>OntoClean</i> .	(1) (2) Conceitos relacionados por subsunção. (3) Conceitos irmãos (mesma hierarquia). (4) Conceitos relacionados por propriedade. (5) Nenhuma das alternativas.
<i>Q1.I</i>	Sugere um conceito para a <i>tag</i> que foi classificada como instância na questão <i>Q1</i> .	(1) Escolha um conceito da ontologia raiz que possui a instância. (2) Sugira um novo conceito para a instância.
<i>Q2I</i>	Classifica o conceito como <i>-I-O</i> (opção 1), <i>+I</i> (opção 2).	(1) Instâncias que representam possíveis valores de um atributo. (2) Instâncias de algum conceito.
<i>Q3I</i>	Classifica o conceito como <i>+O</i> (opção 1), <i>-O</i> (opção 2).	(1) Instâncias de algum conceito que apenas fornecem atributos ou características para outras instâncias abaixo da sua hierarquia. (2) Instâncias de algum conceito que herdam e também fornecem atributos a outros conceitos.
<i>Q4I</i>	Classifica o conceito como <i>-R</i> (opção 1), <i>+R</i> (opção 2).	A instância possui alguma dependência com outra instância?
<i>Q5I</i>	Classifica o conceito como <i>+D</i> (ao responder Sim), <i>-D</i> (ao responder Não).	(1) Se a instância pode deixar de existir e tornar-se instância de um conceito mais geral. (2) A instância não pode deixar de existir.
<i>Q5I</i>	Sugestões calculadas com base nas restrições da <i>OntoClean</i> .	(1) (2) Conceitos relacionados por subsunção. (3) Conceitos irmãos (mesma hierarquia). (4) Conceitos relacionados por propriedade. (5) Nenhuma das alternativas.

A primeira questão, *Q1*, possui o objetivo de classificar o  $e_t$  como conceito ou instância. Na *Q1* também é possível ignorar a relação do  $e_t$  com  $e_r$  apresentada pelo alinhamento. Neste caso, a relação entre o  $e_t$  e o  $e_r$  criada no alinhamento não induz a nenhum enriquecimento na ontologia raiz. Ao ser ignorado, o  $e_t$  é descartado e recupera-se o próximo elemento da relação  $e_t$  com  $e_r$ .

Se na *Q1* o projetista responder que o  $e_t$  pode tornar-se um conceito na ontologia raiz, uma segunda questão *Q2C* é realizada. Caso o projetista responda “instância” na *Q1*, o fluxo é direcionado para a questão *Q2I*. Há dois fluxos a partir da *Q1*, um para instâncias (a

partir da *Q2I*) e outro para conceitos (a partir da *Q2C*). As respostas a partir das questões *Q2C* e *Q2I* classificam o  $e_t$  com as noções filosóficas da *OntoClean* que incluem rigidez (+*R/-R*), identidade (+*I/-I*, +*O/-O*) e dependência (+*D/-D*).

#### *Fluxo a partir da questão Q2C.*

A *Q2C* classifica o  $e_t$  com a metapropriedade +*I/-I*, mais precisamente, se  $e_t$  é um *sortal* ou *non-sortal*. Na *Q2C* o projetista deve escolher uma das três alternativas da questão. As duas primeiras alternativas descrevem conceitos *non-sortal*. A terceira alternativa descreve um *sortal*. Se o projetista escolher uma das duas primeiras alternativas para o  $e_t$ , o mesmo receberá a metapropriedade -*I*. Pelas restrições *OntoClean*, todo conceito que recebe a classificação -*I* também é -*O*. Ao escolher uma das duas primeiras alternativas, o fluxo é direcionado para a questão *Q4C*. A terceira alternativa da *Q2C* classifica o  $e_t$  como +*I* e direciona o fluxo para a questão *Q3C*.

A *Q3C* possui duas alternativas que classificam o  $e_t$  como +*O/-O*. A primeira alternativa classifica o  $e_t$  como +*O*. A segunda opção classifica o  $e_t$  como -*O*. As duas opções direcionam o questionamento para a questão *Q4C*.

A *Q4C* verifica a rigidez do  $e_t$  classificando o mesmo como +*R/-R*. Independente da resposta, a próxima questão é a *Q5C*.

A *Q5C* verifica a dependência do  $e_t$ . Esta é a última questão do fluxo. Conforme as respostas das questões, o questionário fornece algumas sugestões de enriquecimento e alteração da ontologia raiz ao projetista na questão *Q5C*. As sugestões referem-se à relação existente entre o  $e_t$  e o  $e_r$  e estruturam o conhecimento do domínio, ou seja, conceitualizam as alterações na ontologia raiz.

#### *Fluxo a partir da questão Q2I*

Quando o projetista sugere que o  $e_t$  é uma instância, a questão *Q2I* solicita ao projetista que indique qual seria o conceito de tal instância. Para isto, são selecionados conceitos da ontologia raiz. O projetista também pode sugerir outros conceitos não existentes na ontologia raiz. Logo, além de uma nova instância, surge um novo conceito na ontologia raiz que deve ter alguma relação com um ou mais  $e_r$ .

O fluxo a partir da questão *Q2I* é muito semelhante ao fluxo anterior (a partir da *Q2C*). A diferença está no conteúdo das questões. A partir da *Q2I*, as questões são direcionadas em conjunto para o conceito escolhido e para a instância, nesse caso  $e_t$ . O

objetivo é obter uma classificação *OntoClean* para um provável conceito que possui um  $e_t$  como instância.

A *Q2I* classifica o conceito como *+I/-I*. Porém diferente da *Q2C*, esta questão possui apenas duas alternativas (a segunda alternativa na *Q2C* é para conceitos sem instâncias e que organizam uma taxonomia, logo este não é o caso, porque neste fluxo há um conceito e uma instância). A primeira alternativa da *Q2I* direciona o fluxo para a questão *Q4I* e a segunda alternativa para a questão *Q3I*.

A *Q3I* classifica o  $e_t$  como *+O/-O*. A *Q4I* classifica o  $e_t$  como *+R/-R*. Por último, a questão *Q5I* classifica o  $e_t$  como *+D/-D*. Ao final deste fluxo, conforme respostas do projetista, são geradas alternativas de enriquecimento na questão *QSI*.

As sugestões de enriquecimento da ontologia raiz nas questões *QSC* e *QSI* são calculadas com base na classificação fornecida pelo projetista ao responder o questionário e pela classificação já existente do  $e_t$  na ontologia raiz. Pela combinação das metapropriedades, ou seja, consequências ontológicas da metodologia *OntoClean* (*rigidez, identidade e dependência*) são possíveis as seguintes classificações: *Type, Quasi-Type, mixin, Mat. Role, Phased Sortal, Category, Formal role e Attribution*.

O Quadro 14 do Apêndice B descreve as possibilidades de sugestões (combinação das metapropriedades e consequências ontológicas). No Quadro 14 apresentam-se as possibilidades para dois elementos (A e B), sendo que tanto A quanto B podem ser  $e_t$  ou  $e_i$ , as justificativas e possíveis sugestões fornecidas nas questões *QSC* e *QSI*.

### 3.3 PRIMEIRA VERSÃO DA ONTOLOGIA RAIZ

A sequência das atividades do método Folkconcept, tal como apresentada, requer no uma folksonomia e uma ontologia raiz logo no início. Enquanto a folksonomia é gerada como descrito anteriormente, a elaboração da primeira versão da ontologia raiz deve ser ligeiramente modificada. Recorda-se que os elementos da ontologia devem estar anotados com metapropriedades da *OntoClean* e que a metapropriedade *referenceTag* deve estar instanciada para permitir o rastreamento bidirecional das *tags* aos elementos da ontologia.

Na primeira iteração, a ontologia raiz ainda não existe. Portanto, recomenda-se aos projetistas a criação manual desta com alguns conceitos e relações taxonômicas. As metapropriedades *OntoClean* são então atribuídas a estes elementos com base nas questões apresentadas na seção 3.2.2 pelos projetistas e especialistas do domínio. Ressalta-se que a

ontologia raiz é desenvolvida com base nas primeiras anotações, ou seja, na primeira folksonomia produzida pelos participantes do projeto.

Em relação à metapropriedade *referenceTag*, sugere-se que os projetistas auxiliados pelos especialistas instanciem-na de forma manual. Desta forma, a folksonomia inicial é associada a elementos da ontologia raiz que também foi construída com base nos conhecimentos adquiridos da primeira anotação.

Em caso de reuso de um ontologia não desenvolvida por meio do Folkconcept, esta deve receber as metapropriedades da *OntoClean* também de forma manual. Uma folksonomia deve ser construída para a instanciação da metapropriedade *referenceTag*.

A situação ideal de aplicação do método é aquela em que uma comunidade com um vocabulário já estabilizado decide representá-lo por meio de uma ontologia e, assim, a criação da folksonomia não é induzida pelo método, mas sim utilizada como ponto de partida para o mesmo.

## 4. EXPERIMENTOS

Apresentam-se neste capítulo a metodologia e os resultados dos dois experimentos realizados.

### 4.1 METODOLOGIA DOS EXPERIMENTOS

Nos dois experimentos realizados neste trabalho, parte-se de uma ontologia raiz construída pelo autor deste trabalho. Os elementos desta ontologia raiz inicial são associados a *tags* das cinco primeiras anotações dos atores. A ontologia raiz é um ponto inicial e será enriquecida pelo método. O método foi experimentado individualmente, isto é, somente um ator desenvolve a ontologia, e colaborativamente (com mais de um ator).

Para a realização do experimento, foi escolhido o domínio do turismo, mas particularmente, senso-comum sobre atividades de turismo, já que não havia especialistas de domínio de fato. Justifica-se esta escolha pelo número reduzido de participantes disponíveis, conforme descrito nas seções 4.1.1 e 4.1.2. Idealmente, seriam necessários participantes para desempenhar separadamente os papéis de projetista, de especialista do domínio e de usuário. Com a escolha de um domínio de fácil compreensão, estes papéis puderam ser desempenhados simultaneamente pelos participantes.

Os participantes selecionados foram alunos de mestrado e doutorado, a maior parte deles com graduação em computação, que cursaram a disciplina de Tópicos Especiais em Ontologias com carga horária de 45 horas ofertada pelo CPGEI e ministrada pelo professor Dr. Cesar Augusto Tacla em 2009.

Os experimentos foram realizados por grupos que utilizam o método e por grupos que não o utilizam a fim de avaliar o método por meio do resultado, isto é, das ontologias produzidas. Grupos que utilizam o método são auxiliados com a automação das atividades de aquisição de conhecimentos e de conceitualização da ontologia raiz. Grupos que não utilizam o método são auxiliados computacionalmente apenas no processo de anotação de recursos (*i.e.* qualquer página *Web*), parte integrante da aquisição de conhecimento, e realizam a conceitualização da ontologia raiz manualmente. Por manual, entende-se que os projetistas recebem as *tags* criadas no processo de anotação e procuram identificar manualmente novos elementos na ontologia raiz.

No primeiro experimento (individual), o aprendizado subsequente à aquisição é feito individualmente pelos projetistas. Cada projetista escolhe as *tags* que induzem a um novo elemento na ontologia raiz e responde ao questionário eletrônico.

No segundo experimento (colaborativo), os especialistas do domínio etiquetam sozinhos, mas atualizam a ontologia em grupo presencialmente. Neste experimento, é utilizado o conjunto de *tags* dos especialistas pertencentes ao grupo e a resposta ao questionário eletrônico é feita colaborativamente pelo grupo, sendo que somente um deles interage com o sistema que implementa o método.

#### **4.1.1 Primeiro Experimento**

Neste experimento participaram 18 atores que etiquetaram individualmente 20 sítios no domínio do turismo. Nove atores utilizaram o Folkconcept para atualização da ontologia raiz, e 9 atualizaram a ontologia raiz manualmente, ou seja, sem a utilização do método. Assim foram produzidas 18 novas ontologias a partir das *tags* individuais dos atores. Cada ator utilizou apenas o seu conjunto de *tags*. As ontologias dos dois grupos (com/sem o método) foram comparadas utilizando-se de métricas que verificam o conteúdo e a correteza das relações taxonômicas.

#### **4.1.2 Segundo Experimento**

Neste experimento participaram outros seis atores, divididos em dois grupos de três participantes. Os atores dos dois grupos etiquetaram individualmente 20 sítios. Neste experimento, um grupo enriqueceu simultaneamente a ontologia raiz utilizando o Folkconcept, e o outro grupo, sem a utilização do método. Assim, foram produzidas duas novas ontologias a partir das folksonomias de cada grupo. As duas ontologias desenvolvidas pelos grupos também foram comparadas de forma similar ao primeiro experimento.

### 4.1.3 Aquisição de Conhecimentos nos Experimentos

A aquisição de conhecimento nos dois experimentos foi dividida em duas fases. Na primeira fase, cada especialista do domínio anotou 5 páginas *Web* no domínio do turismo individualmente. Estas anotações foram recuperadas, enriquecidas e representadas na ontologia de representação de *tags*. Esta ontologia de anotações, nesta primeira fase, possui as *tags* de todos os especialistas.

As *tags* desta primeira fase foram associadas a conceitos e instâncias da ontologia raiz inicial que já estava construída. Assim, os elementos da ontologia raiz inicial passaram a ser rastreáveis a partir da folksonomia inicial.

Na segunda fase, cada especialista do domínio anotou individualmente 15 páginas *Web* no mesmo domínio a fim de simular a evolução da folksonomia. Nesta fase, para cada especialista, foi criada uma nova ontologia de anotações a partir da primeira (inicial).

### 4.1.4 Conceitualização nos Experimentos

A conceitualização consiste em criar os novos elementos na ontologia raiz inicial, ou seja, o enriquecimento da ontologia raiz. Nesta atividade, cada projetista recebe a ontologia raiz que deverá ser atualizada a partir das novas *tags* criadas na aquisição de conhecimentos. Como existe uma divisão dos experimentos pela utilização ou não do método, os projetistas que não utilizaram o método conceitualizaram manualmente os novos elementos na ontologia raiz. Os projetistas que utilizaram o método, responderam o questionário eletrônico a partir das relações construídas pelo alinhamento. Tanto os projetistas que utilizaram o método, quanto os que fizeram manualmente o enriquecimento da ontologia raiz, tiveram 60 minutos para realizar a conceitualização.

### 4.1.5 Desenvolvimento da Ontologia Raiz

Segundo Gómez-Pérez (1999) uma ontologia pode ser desenvolvida do “zero” ou a partir de uma ontologia existente. Esta última inclui ontologias de domínio, aplicação ou alguma ontologia de nível superior. Em todos os casos busca-se enriquecer uma ontologia existente pela inclusão de novos elementos ontológicos, novas relações entre os mesmos e

novos axiomas. Nesta tese a ontologia raiz não é considerada uma ontologia de nível superior, mas uma ontologia inicial de domínio. Esta ontologia pode ser desenvolvida pelos projetistas quando da criação da folksonomia e início da atividade de aquisição de conhecimentos. Além disso, uma ontologia pode ser desenvolvida por sucessivas interações, ou seja, são necessárias representações intermediárias até a versão final. Os projetistas desenvolvem a ontologia a medida que adquirem novos conhecimentos. Logo, a cada iteração, uma nova ontologia é desenvolvida, sendo esta, na próxima iteração, a ontologia raiz.

A primeira versão da ontologia seria normalmente desenvolvida pelos projetistas em uma etapa anterior ao experimento realizado. Para simplificar a experimentação, a primeira versão foi feita pelo autor deste trabalho em duas etapas.

Na primeira, verificou-se alguns conceitos existentes no domínio turismo pela análise de algumas ontologias e especificações existentes (OnTour<sup>10</sup>, Mondeca<sup>11</sup>, OTA – *Open Travel Alliance*<sup>12</sup>). Por exemplo, conceitos como itinerário, acomodação e transporte foram criados nesta primeira etapa e também algumas propriedades.

Na segunda etapa, as cinco primeiras anotações de cada especialista foram recuperadas. Destas cinco primeiras anotações foram retiradas as *tags* que poderiam conectar-se aos conceitos e instâncias da ontologia raiz inicial. Procurou-se utilizar todas as *tags* das cinco primeiras anotações. Ao total foram utilizadas 185 *tags*. Para algumas *tags* foram criados novos conceitos e também novas instâncias. Por exemplo, as *tags event* e *china* deram origem respectivamente a um novo conceito (*event*) e uma nova instância (*china*) do conceito *country*. As 185 *tags* foram associadas como *referenceTag* aos conceitos e instâncias da ontologia raiz, ou seja, conectou-se a folksonomia inicial a ontologia raiz inicial. A Figura 54 ilustra parte da ontologia raiz construída inicialmente. São mostradas a taxonomia e algumas instâncias da ontologia raiz. Cada conceito ou instância possui uma propriedade *referenceTag* com o respectivo *label* do elemento na ontologia de representação de anotações. A ontologia raiz possui 98 conceitos, 15 propriedades e 76 instâncias.

---

<sup>10</sup> <http://e-tourism.deri.at/ont/index.html>.

<sup>11</sup> <http://www.mondeca.com>

<sup>12</sup> <http://www.opentravel.org/>



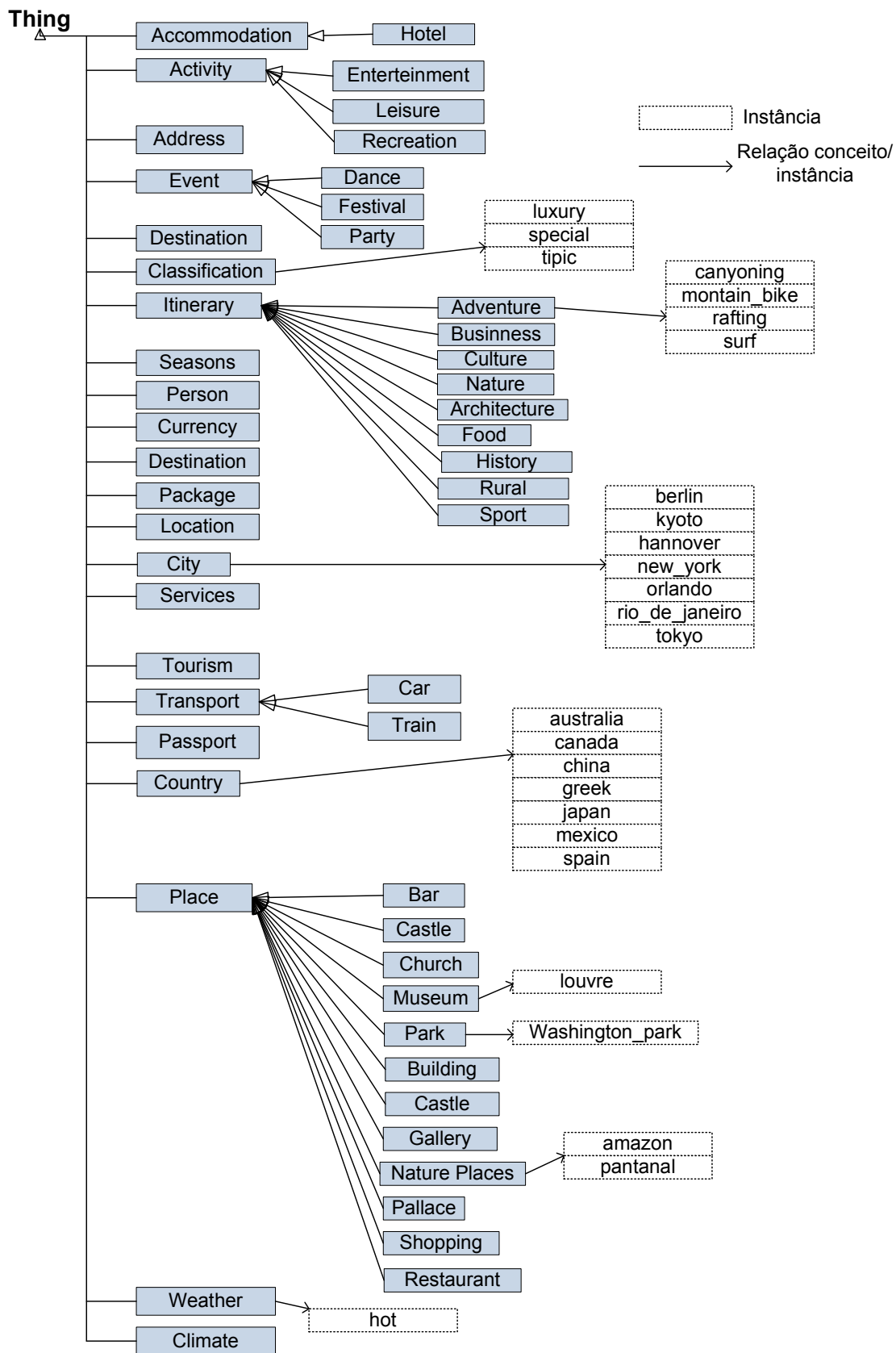


Figura 54 – Ontologia Raiz Inicial.

#### 4.1.6 Avaliação dos Experimentos

As ontologias desenvolvidas nos experimentos foram avaliadas quantitativamente. A Figura 55 ilustra o posicionamento da avaliação das ontologias deste experimento aos trabalhos relacionados na avaliação de ontologias. Avalia-se o conteúdo de uma ontologia pela utilização de métricas quantitativas relacionadas à parte **estrutural** e **funcional** de uma ontologia.

Na avaliação estrutural, foram utilizadas métricas relacionadas ao número de elementos criados, métricas do *OntoQA* e outras que incluem largura, profundidade e emaranhamento descritas na *oQual* (GANGEMI et al., 2006).

Na parte funcional, verificou-se a conceitualização pretendida relacionado a critérios de projeto, ou seja, a correte das relações taxonômicas com base na natureza ontológica dos elementos preconizados na *OntoClean*. Não verificou-se a cobertura e precisão da ontologia porque não há uma ontologia padrão, ou uma ontologia considerada ideal e completa para o domínio de turismo, ou alguma fonte de referência a ser utilizada.

A avaliação de qualidade das ontologias nesta tese é realizada com base em um conjunto de métricas, tanto na análise do conteúdo na parte estrutural quanto no significado pretendido ou critérios de projeto na parte funcional pela aplicação da *OntoClean*. O objetivo da avaliação é verificar se o método auxiliou os grupos a desenvolverem ontologias de melhor qualidade do que as produzidas sem a utilização do método.

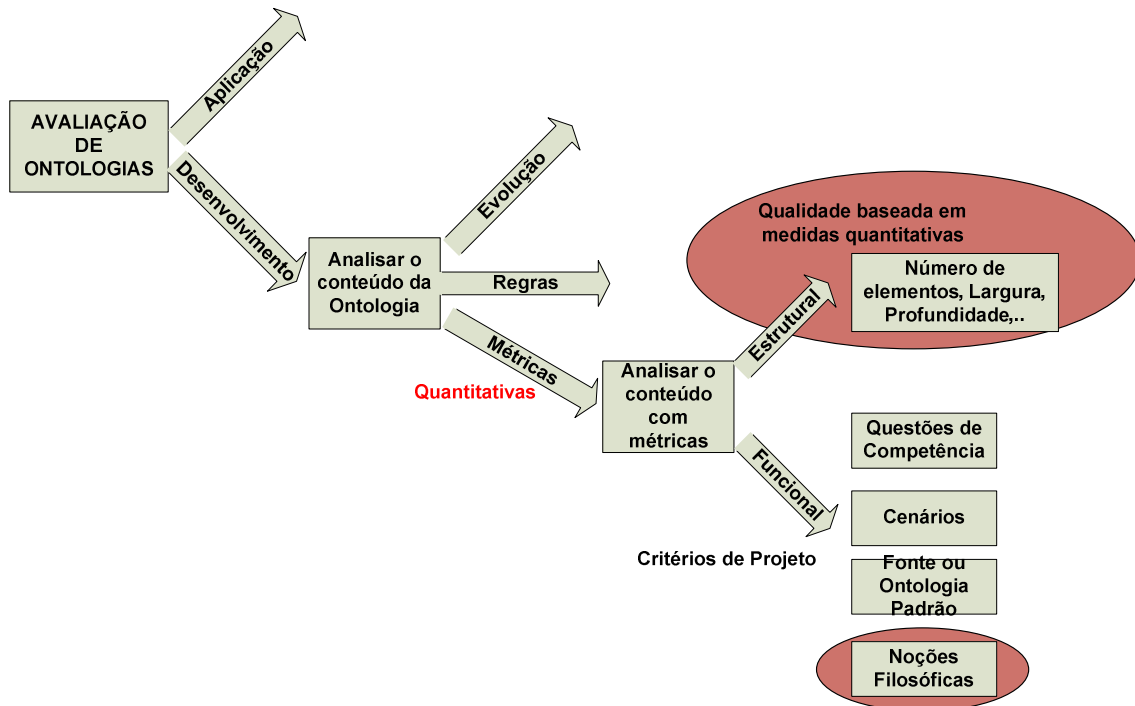


Figura 55 – Posicionamento da Avaliação dos Experimentos na Avaliação de Ontologias.

Outras avaliações realizadas nos experimentos foram acerca do aproveitamento das *tags* criadas no processo de anotação, a precisão do alinhamento e do questionário eletrônico na conceitualização. Enfim, nos dois experimentos realizados procurou-se indícios para verificar a validade das hipóteses desta tese:

- O método desenvolvido, ajuda na capacidade cognitiva dos atores diminuindo o problema de aquisição de conhecimentos tornando mais fácil a conceitualização e explicitação de tal conceitualização?
- O método desenvolvido (aquisição de conhecimentos e conceitualização com a integração de um processo de avaliação) baseado nas folksonomias ajuda na elaboração de ontologias mais descritivas e com formalismo correto baseado nas consequências ontológicas?

## 4.2 RESULTADOS DO EXPERIMENTO INDIVIDUAL

Na Tabela 4 são apresentados os resultados obtidos para os dois grupos de atores, sendo o grupo 1 aquele que utilizou o **Folkconcept** e o grupo 2, o que trabalho **manualmente**.

A primeira coluna mostra o total de *tags* criadas no processo de anotação e as colunas Conceitos, Propriedades e Instâncias, o número de elementos criados na ontologia raiz. O grupo 1 criou mais propriedades e instâncias. O grupo 2 foi superior apenas no número de conceitos.

Tabela 4 – Elementos criados nas Ontologias do Primeiro Experimento.

	Total de Tags	Conceitos	Propriedades	Instâncias	Total de Elementos
<u>grupo 1</u>	359	131	91	105	327
<u>grupo 2</u>	385	156	50	82	288
<b>Total</b>	744	287	141	274	702

A Tabela 5 apresenta o número de conceitos populados, aqueles que foram instanciados, e de relações taxonômicas criadas nos dois grupos. O valor para conceitos populados é maior no grupo 1. Para as relações taxonômicas, o grupo 2 foi superior. As relações taxonômicas foram criadas quando da inclusão de um novo conceito na ontologia raiz, porém em alguns casos, conceitos foram inseridos sem um posicionamento taxonômico (neste trabalho filho direto de *Thing* não foi considerado um posicionamento taxonômico).

Tabela 5 – Número de Conceitos Populados e Relações Taxonômicas das Ontologias do Primeiro Experimento.

	Conceitos Populados	Relações Taxonômicas
<u>grupo 1</u>	62	40
<u>grupo 2</u>	49	78
<b>Total</b>	122	134

A Tabela 6 apresenta o desvio padrão em relação à média de elementos criados na ontologia e ao número de tags por ator número. Percebe-se que para o total de *tags* o desvio padrão é semelhante nos dois grupos, porque a atividade de anotação é a mesma para os dois grupos. Porém, para os demais elementos, o desvio padrão é diferente, visto que o grupo 1 utilizou o método e o grupo 2 fez o processo manualmente.

Tabela 6 – Desvio Padrão do Total de Tags e Elementos Criados pelos Atores no Primeiro Experimento.

	Total de Tags	Conceitos	Propriedades	Instâncias	Conceitos Populados	Relações Taxonômicas
<u>grupo 1</u>	6,13	3,97	2,42	3,35	2,97	3,87
<u>grupo 2</u>	5,11	10,83	9,88	10,19	6,80	9,80

Observando a Tabela 4, pode-se obter o aproveitamento das *tags*, ou seja, se as *tags* induziram a criação de um novo elemento na ontologia raiz. A Tabela 7 descreve o aproveitamento das *tags* apresentando o total de *tags* criadas no processo de anotação, o total de elementos criados na ontologia raiz, as *tags* que foram ignoradas e as *tags* aceitas ou *tags* que induziram novos elementos. O grupo 1 criou 327 elementos de 359 *tags* utilizadas nas anotações. O grupo 2 com 385 *tags* criou 288 elementos. Porém, as 213 *tags* aceitas no grupo 1 geraram 327 elementos, um índice de 1,53 elementos criados para cada *tag* aceita. O grupo 2 criou 288 elementos a partir das 254 *tags* aceitas, um índice de 1,13 elementos criados para cada *tag*. Tanto no grupo 1 quanto no grupo 2, uma *tag* poderia induzir a criação de um ou mais elementos na ontologia raiz. Por exemplo, na criação de um conceito, o projetista também poderia criar uma instância do mesmo.

Tabela 7 – Relação de *Tags*, Elementos, *Tags* Ignoradas, *Tags* Aceitas e Índice Elementos/*Tags* Aceitas no Primeiro Experimento.

	Total de Tags	Total de Elementos Criados	Tags Ignoradas	Tags Aceitas	Elementos Criados/ Tags Aceitas
<u>grupo 1</u>	359	327	146 (40,66%)	213 (59,34%)	1,53 (327/213)
<u>grupo 2</u>	385	288	131 (34,03%)	254 (65,97%)	1,13 (288/254)

Com relação as *tags* ignoradas e aceitas do grupo 1 (ver Tabela 8), das 146 *tags* ignoradas, 107 foram no alinhamento, ou seja, o projetista não aceitou a associação criada entre  $tag(e_t)$  e elemento ( $e_r$ ) da ontologia, e 39 foram ignoradas no final do processo de conceitualização. Observando-se a Tabela 8 pode-se calcular a precisão do alinhamento e da conceitualização (aceite das sugestões na conceitualização). Das 359 *tags* alinhadas, 213 alinhamentos entre  $e_t$  e  $e_r$  foram aceitos pelos projetistas. Isto indica uma precisão de 59,34%. Com relação à precisão das sugestões no questionário eletrônico, os 213 alinhamentos geraram 213 sugestões, sendo que apenas 39 sugestões foram ignoradas. Assim, os projetistas aceitaram 174 sugestões do questionário, uma precisão de 81,69%.

Tabela 8 – Relação de *Tags* Ignoradas, *Tags* Aceitas e Sugestões Aceitas no Grupo 1 do Primeiro Experimento.

	Total de <i>Tags</i>	<i>Tags</i> Ignoradas	<i>Tags</i> Ignoradas (Alinhamento)	<i>Tags</i> Ignoradas (Sugestão)	<i>Tags</i> Aceitas	Sugestões Aceitas
<u>grupo 1</u>	359	146 (40,66%)	107 (73,29%)	39 (26,71%)	213 (59,34%)	174 (81,69%)

Na avaliação pelas métricas da *OntoQA*, utilizou-se o número de elementos criados na ontologia (Tabela 4), número de conceitos populados e relações taxonômicas (Tabela 5). A Tabela 9 apresenta os resultados. Além das métricas na Tabela 9, foi incluído o *score* que corresponde ao somatório de todas as métricas. O grupo 1 foi inferior apenas na medida *schema deepness*. O motivo é que o grupo 1 criou menos conceitos que o grupo 2, logo menos relações taxonômicas. A métrica *average population* e *class utilization* foi influenciada pelo número de instâncias. Porém, observa-se que mesmo com um número inferior de conceitos criados, o grupo 1 criou mais instâncias e populou mais conceitos. A métrica *attribute richness* foi superior no grupo 2 porque o mesmo criou mais propriedades, ou seja, descreveu com maior detalhamento os conceitos. Para exemplificar as métricas, utilizou-se parte das ontologias desenvolvidas pelos projetistas nos grupos.

Tabela 9 – Avaliação das Ontologias do Primeiro Experimento utilizando-se das Métricas *OntoQA* no Primeiro Experimento.

	Métricas					
	<i>Attribute Richness</i>	<i>Relationship Diversity</i>	<i>Schema Deepness</i>	<i>Average Population</i>	<i>Class Utilization</i>	<i>score</i>
<u>grupo 1</u>	0,69	0,66	0,30	0,80	0,47	2,96
<u>grupo 2</u>	0,32	0,34	0,50	0,52	0,31	2,08

A métrica *attribute richness* mensura o número de propriedades por conceitos. Quanto maior o valor desta métrica, maior o número de propriedades por conceito, ou seja, mais rica é a descrição das instâncias. Na Figura 56, há parte de uma ontologia desenvolvida com a utilização do método e, na Figura 57, sem a utilização do método. Na Figura 56 o somatório das propriedades de todos os 12 conceitos é 20. Logo, o *attribute richness* é 1,66 (20/12). Na Figura 57, o somatório das propriedades é 8 em 16 conceitos, sendo o *attribute richness* igual a 0,5 (8/16).

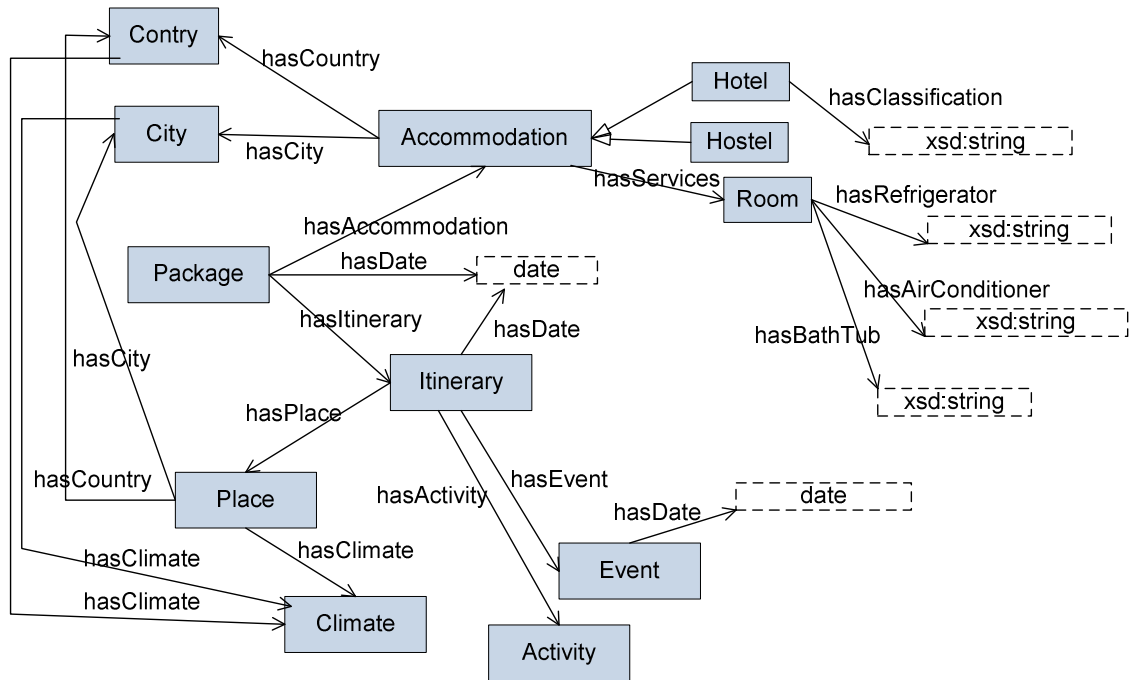


Figura 56 – Ontologia desenvolvida pelo Grupo 1 para exemplificar a Métrica *Attribute Richness*.

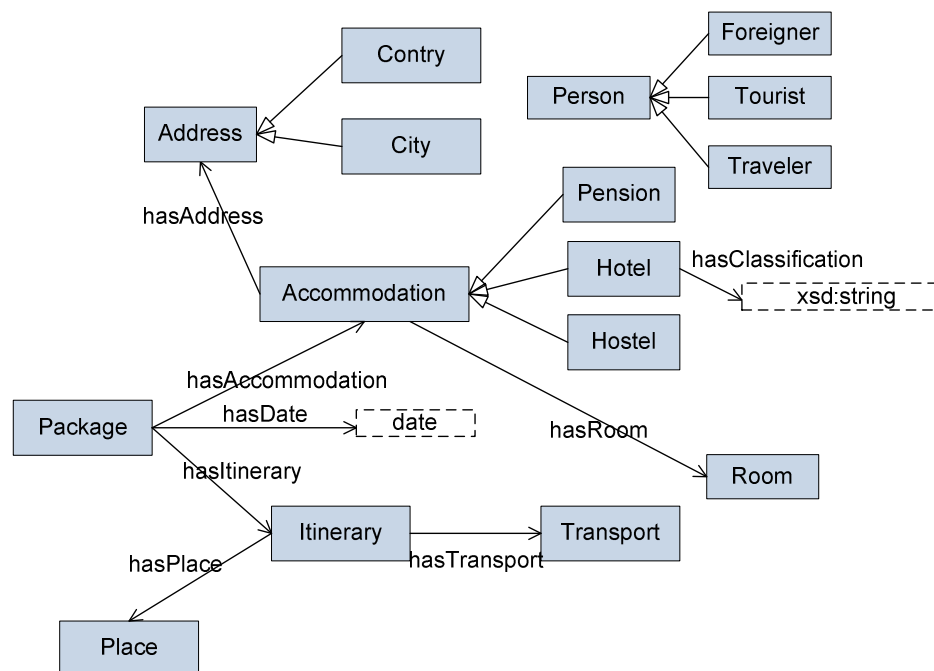


Figura 57 – Ontologia desenvolvida pelo Grupo 2 para exemplificar a Métrica *Attribute Richness*.

A Figura 58 e a Figura 59 exemplificam o cálculo da métrica *relationship diversity* para os grupos F e M respectivamente. Na Figura 58 (grupo 1), há 9 propriedades e 10 relações taxonômicas. O valor da *relationship diversity* desta ontologia é 0,47 ( $9/(9+10)$ ). Na

ontologia da Figura 59 (grupo 2), há 3 propriedades e 17 relações taxonômicas. O *relationship diversity* desta ontologia é 0,15 ( $3/(3+17)$ ). Isto mostra que a ontologia da Figura 58 possui maior diversidade de relações que a da Figura 59. No limite, quando não há propriedades, a única relação entre os conceitos é a taxonômica. Quando isto ocorre tem-se uma ontologia leve.

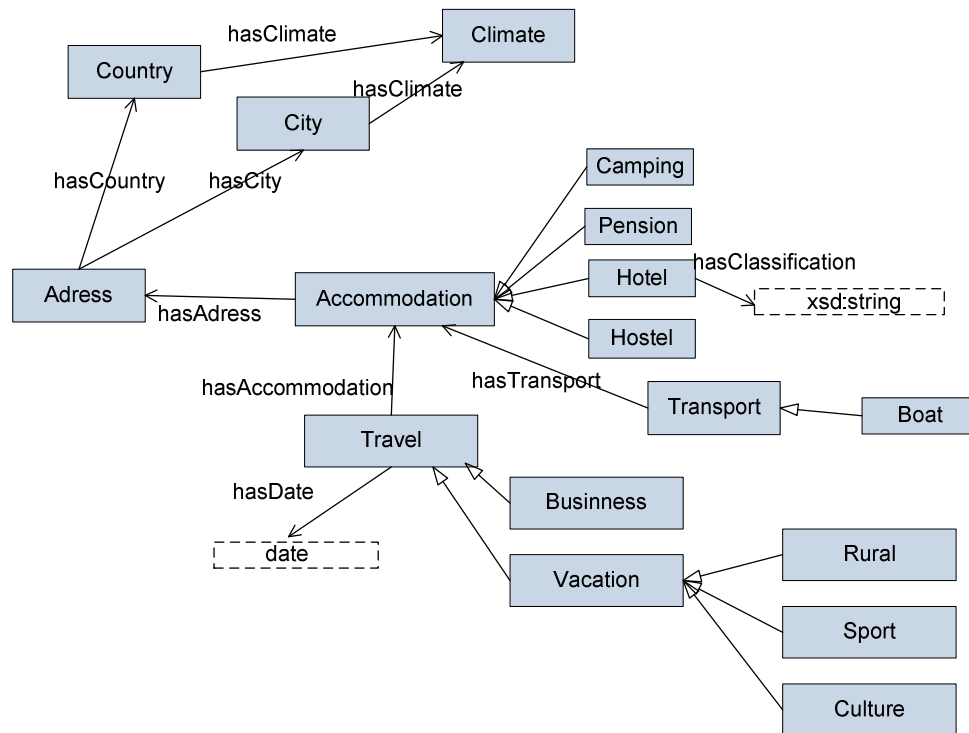


Figura 58 – Ontologia desenvolvida pelo Grupo 1 para exemplificar as Métricas *Relationship Diversity* e *Schema Deepness*.



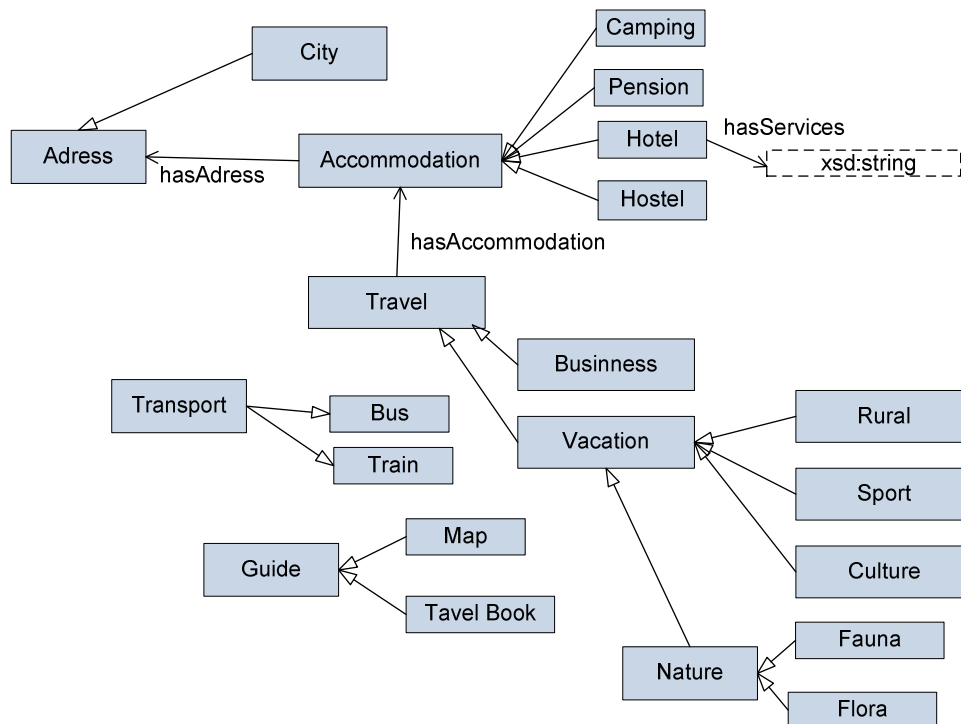


Figura 59 – Ontologia desenvolvida Grupo 2 para exemplificar as Métricas *Relationship Diversity* e *Schema Deepness*.

Para exemplificar a métrica *schema deepness* (número médio de relações taxonômicas/subclasses por conceito), utiliza-se também a Figura 58 e a Figura 59. Para a ontologia da Figura 58 (grupo 1), o valor do *schema deepness* é 0,58 (10/17), ou seja, há 10 relações taxonômicas (sem considerar filhos de *Thing*) em 17 conceitos. Na ontologia da Figura 59 (grupo 2), há 17 relações taxonômicas em 22 conceitos. Logo o *schema deepness* é 0,77. Esta métrica é influenciada pelo número de conceitos e pelas relações taxonômicas entre os mesmos em uma ontologia. Um número maior para *schema deepness* indica que uma ontologia possui conceitos conectados em sua maioria por relações taxonômicas. Um número inferior de *schema deepness* indica que há conceitos dão origem a hierarquias independentes (subárvores conceituais) que são conectadas por meio de propriedades. A métrica *schema deepness* também descreve a distribuição de classes entre os diferentes níveis hierárquicos da ontologia. Valores baixos para *schema deepness* representam que cada classe possui poucas especializações (conhecimento mais geral com poucos níveis de detalhe na hierarquia). Um valor alto descreve uma ontologia com muitas especializações por classes (ontologia que reflete um conhecimento mais detalhado, domínios mais específicos).

A métrica *average population* (número médio de instâncias por conceito) é exemplificada pela Figura 60 (grupo 1) e Figura 61 (grupo 2). Na ontologia da Figura 60 há

22 instâncias e 16 conceitos. Logo o *average population* é de 1,46 (22/15). Na ontologia da Figura 61 o *average population* é de 0,55 (10/18). Percebe-se que a ontologia do grupo 1 (Figura 60) possui menos conceitos, porém com um número maior de instâncias que a ontologia da Figura 61.

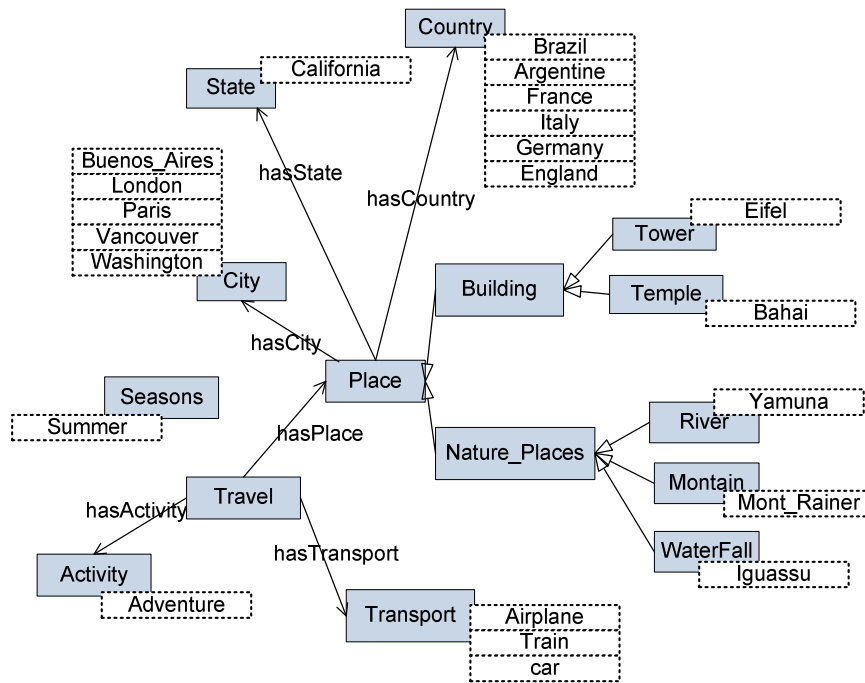


Figura 60 – Ontologia desenvolvida pelo Grupo 1 para exemplificar as Métricas *Average Population* e *Class Utilization*.

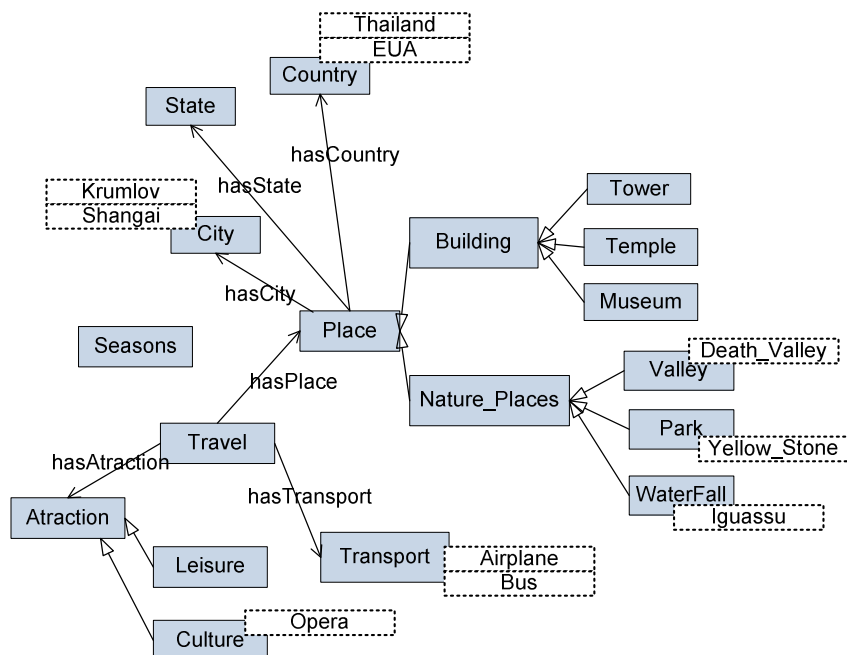


Figura 61 – Ontologia desenvolvida pelo Grupo 2 para exemplificar as Métricas *Average Population* e *Class Utilization*.

A Figura 60 e a Figura 61 exemplificam o cálculo da métrica *class utilization*. Na ontologia da Figura 60, há 11 conceitos populados em um total de 15 conceitos, sendo o valor de *class utilization* igual a 0,73 (11/15). Na ontologia da Figura 61, há 7 conceitos populados em 18 conceitos. Assim, o *class utilization* é 0,38 (7/18). Isto mostra que apesar de um número maior de conceitos na ontologia da Figura 61, a grande maioria não possui instâncias sendo o valor de *class utilization* menor do que para a ontologia do grupo 1.

Além das métricas da *OntoQA*, também utilizou-se métricas descritas pelo trabalho *oQual* (GANGEMI et al., 2006), entre elas, profundidade, largura e emaranhamento (*tangledness*) dos nós. A Tabela 10 descreve a profundidade máxima, média de profundidade, largura máxima e média de largura das ontologias desenvolvidas pelos dois grupos. Percebe-se que as ontologias criadas pelos grupos é semelhante na largura. Na profundidade há uma diferença, sendo que o grupo 1 criou ontologias com maior profundidade.

Tabela 10 – Largura e Profundidade das Ontologias Desenvolvidas pelos Grupos no Primeiro Experimento.

	Profundidade Máxima	Média de Profundidade	Largura Máxima	Média de Largura
<u>grupo 1</u>	<u>6</u>	<u>3,11</u>	<u>4</u>	<u>1,88</u>
<u>grupo 2</u>	<u>3</u>	<u>2,13</u>	<u>4</u>	<u>1,94</u>

No emaranhamento verificou-se que alguns conceitos possuem relações multihierárquicas. Nesse sentido o grupo 1 não criou conceitos multihierárquicos. Nas ontologias do grupo 2 houve 6 relações multihierárquicas. A Figura 62 ilustra uma relação multihierárquica. Na ontologia, o conceito *Farm* é o resultado da especialização dos conceitos *Adventure* e *Place*. Logo, o conceito *Farm* é resultante de um relacionamento multihierárquico, porém *Farm* possui dois significados diferentes (local de turismo, tipo de turismo) por herdar diferentes características dos conceitos *Place* e *Adventure*.

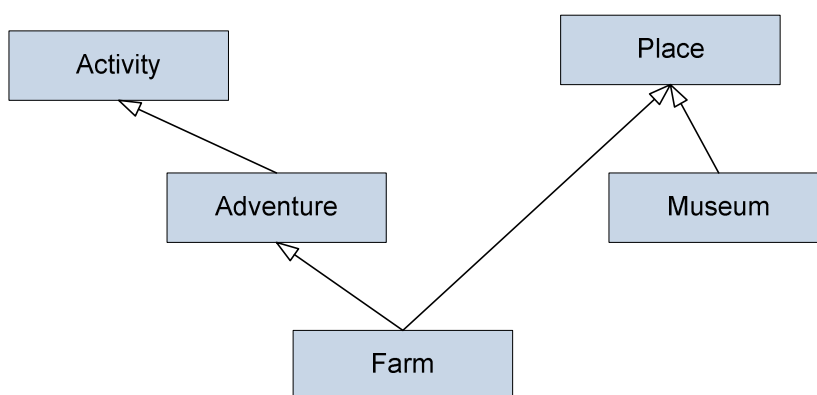


Figura 62 – Exemplo de Relações Multihierárquica.

Nas ontologias desenvolvidas pelos grupos, também verificou-se erros de modelagem relacionados a taxonomia idealizada pela *OntoClean* (ver Figura 16). Na taxonomia, *Sortals* podem ser subsumidos por *Non-Sortals*. Nessa mesma taxonomia não é possível um *Non-Sortal* ser subsumido por um *Sortal* devido ao critério de identificação.

Verificou-se nas ontologias desenvolvidas, erros de relacionamento taxonômico entre conceitos classificados como *Sortal* e *Non-Sortal*. A Tabela 11 descreve o número de erros relacionados a taxonomia idealizada pela *OntoClean*. O grupo 1 criou 9 relacionamentos taxonômicos inconsistentes de 40 relações taxonômicas criadas. O grupo 2 criou 32 relacionamentos taxonômicos inconsistentes de 78 criados.

Tabela 11 – Erros de Modelagem relacionados a Taxonomia Idealizada *OntoClean* no Primeiro Experimento.

	Erros taxonômicos
<u>grupo 1</u>	9 (de 40) (22,5%)
<u>grupo 2</u>	35 (de 78) (44,9%)

#### 4.2.1 Síntese do Primeiro Experimento

Pelos resultados descritos na avaliação quantitativa (Tabela 4 e Tabela 5), percebe-se que o grupo 2 teve um número superior de conceitos e relações taxonômicas criados em relação ao grupo 1. Parece que a quantidade de conceitos criados pelo grupo 2 levou-os a criarem um número maior de relações taxonômicas. O grupo 1 foi inferior nas relações taxonômicas e conceitos, porém criou mais propriedades e mais instâncias. Com isso, os conceitos possuem são mais bem descritos, além de possuírem mais instâncias. Consequentemente o número de conceitos populados também é maior. A Figura 63 ilustra um gráfico comparativo dos elementos criados nas ontologias nos dois grupos.

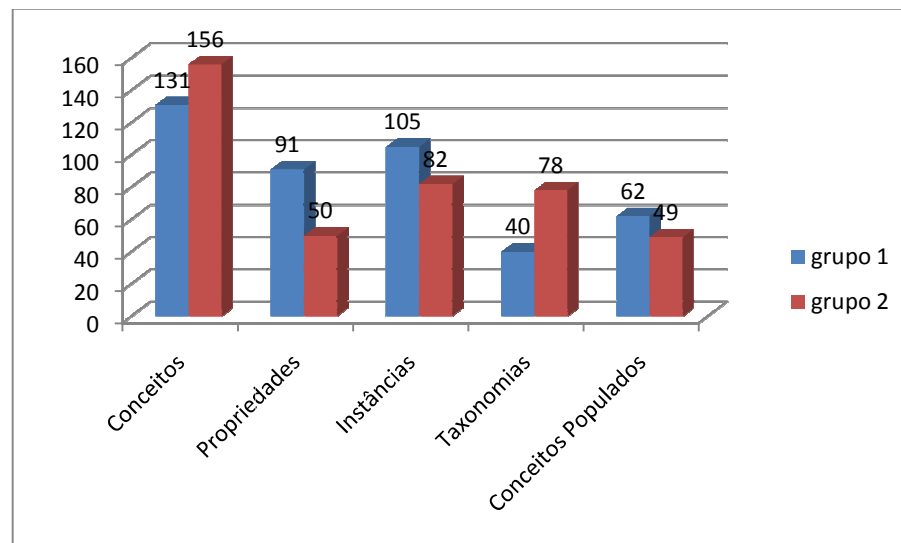


Figura 63 – Gráfico Comparativo da Avaliação Quantitativa do Primeiro Experimento.

A Figura 64 ilustra o desvio padrão das tags criadas e dos elementos criados nas ontologias. Percebe-se que os grupos foram semelhantes no total de *tags*. Isto porque a mesma atividade de anotação foi executada nos dois grupos. Porém, no grupo 1, que utilizou o Folkconcept, todos os projetistas criaram uma quantidade similar de elementos. No grupo 2, a não utilização do sistema dispersou a criação dos elementos, ou seja, os projetistas foram dissimilares no número de elementos criados. O grupo 1 desenvolveu as ontologias de forma mais uniforme. Todos os projetistas identificaram elementos na mesma proporção.

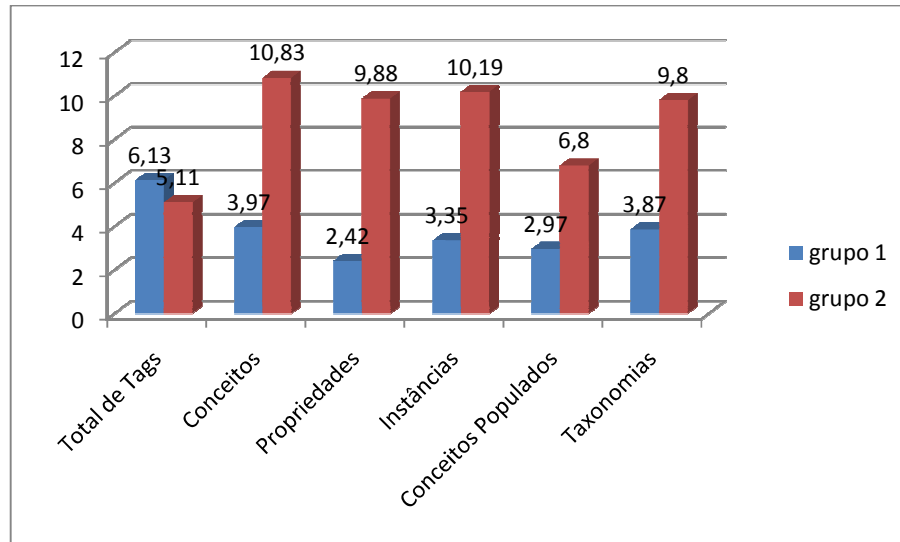


Figura 64 – Gráfico do Desvio Padrão dos Elementos Criados pelos Grupos no Primeiro Experimento.

A Figura 65 ilustra um gráfico relacionado a estrutura da ontologia na largura e profundidade. Percebe-se que o grupo 1 criou ontologias mais profundas ao enriquecer a ontologia raiz. Com isso, os projetistas procuraram estender relações taxonômicas de conceitos já presentes na ontologia raiz. Com relação à largura, as ontologias foram muito semelhantes. Porém, como não foram considerados filhos de *Thing*, percebe-se que o grupo 2 dispersou (maior largura na taxonomia) muitos conceitos a partir de *Thing*.

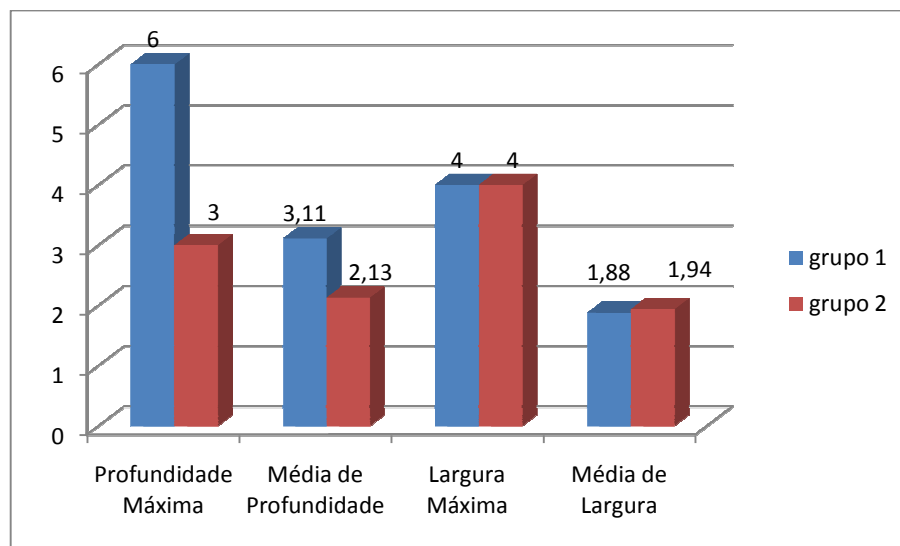


Figura 65 – Largura e Profundidade das Ontologias dos Grupos no Primeiro Experimento.

Com relação ao aproveitamento de *tags* (Tabela 7), o grupo 1 também foi superior ao grupo 2. A Figura 66 ilustra um gráfico comparativo do total de *tags* criadas, ignoradas e aceitas e do total de elementos criados a partir das *tags*. O grupo 1 obteve um índice de 1,53 elementos por *tag*, enquanto que no grupo 2 este índice foi de 1,13. O grupo 1, apesar de ignorar mais *tags*, obteve um aproveitamento melhor das *tags*. Como o grupo 2 não utilizou o Folkconcept, os projetistas deste grupo tinham uma certa liberdade para conectar a *tag* a um elemento na ontologia raiz, ou seja, induzir através da *tag* um novo elemento na ontologia raiz. No grupo 1, esta conexão é feita pelo algoritmo de alinhamento das *tags*. É o método que induz a conexão entre *tag* e elemento na ontologia raiz. Por isso o grupo 1 teve mais *tags* ignoradas.

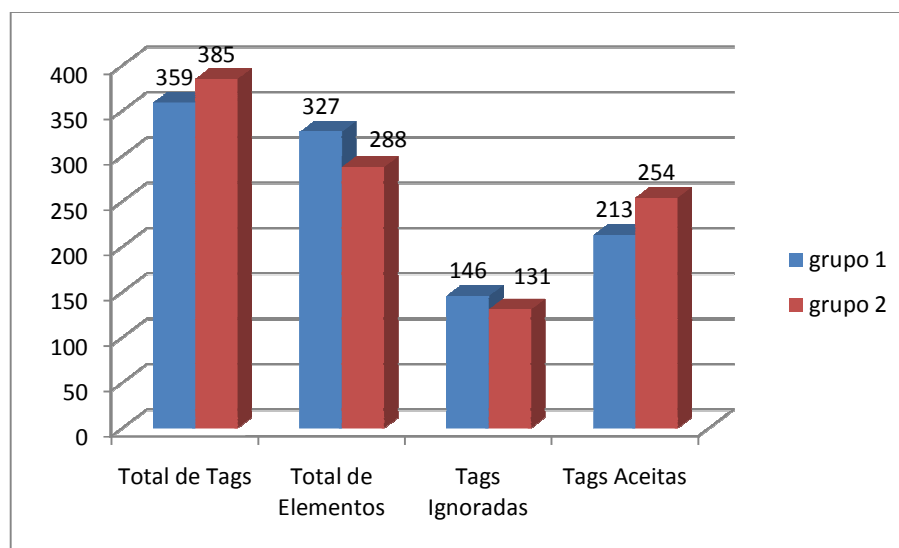


Figura 66 – Gráfico do Aproveitamento de *Tags* no Primeiro Experimento.

Na Figura 67, das *tags* aceitas e ignoradas, foi verificado apenas para o grupo 1 se as *tags* foram ignoradas no alinhamento ou nas sugestões criadas pelo questionário eletrônico na conceitualização. Das 359 *tags*, 146 foram ignoradas pelo alinhamento e as 213 restantes foram aceitas pelos projetistas. Logo, 213 alinhamentos de *tags* foram aceitos gerando uma precisão de 59,34%. As 213 *tags* aceitas geraram 213 sugestões. Destas apenas 39 foram ignoradas pelos projetistas. Logo, o questionário eletrônico teve uma precisão de 81,69%.

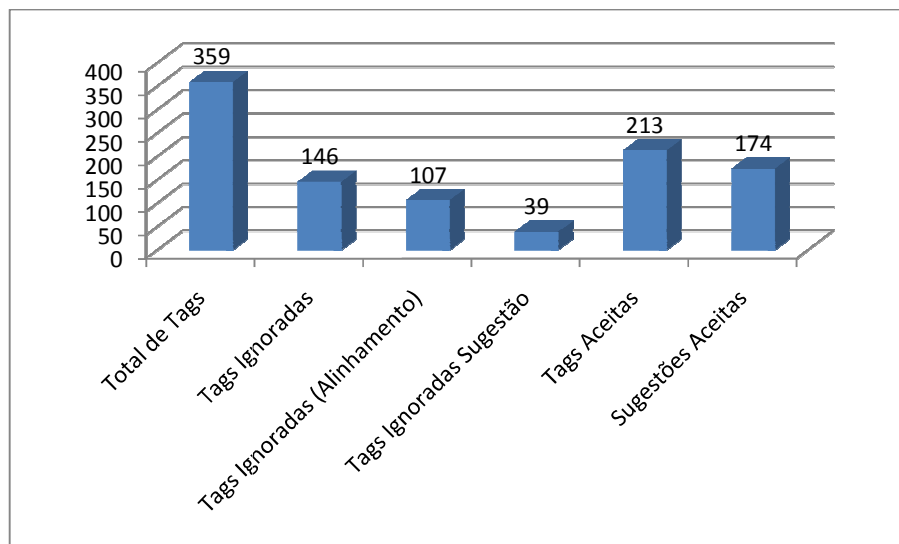


Figura 67 – Gráfico do Aproveitamento de *Tags* para o Grupo 1 no Primeiro Experimento.

Na avaliação com base nas métricas da *OntoQA*, (Tabela 9), o grupo 1 foi superior ao grupo 2 para as métricas *attribute richness*, *relationship diversity*, *average population* e *class utilization*. Apenas para a métrica *schema deepness* o grupo 1 foi inferior. Conclui-se que o grupo 1 gerou ontologias mais descritivas devido a quantidade de atributos, propriedades e instâncias criadas. As ontologias do grupo 1 possuem uma descrição mais detalhada dos conceitos, uma superioridade da quantidade de propriedades em comparação ao número de relações taxonômicas. Além disso, o grupo 1 também populou mais conceitos ao criar instâncias e criou mais instâncias por conceito. Devido ao fato do grupo 1 criar menos conceitos e, conseqüentemente, um número menor de relações taxonômicas, a métrica *schema deepness* foi inferior ao grupo 2. Porém, todos os conceitos criados nas ontologias do grupo 1 possuem alguma relação seja taxonômica ou por propriedade com outro conceito. Isto indica que todos os conceitos estão relacionados, ao contrário do grupo 2, que devido ao grande número de conceitos e baixo número de propriedades, alguns conceitos ficaram isolados na ontologia.



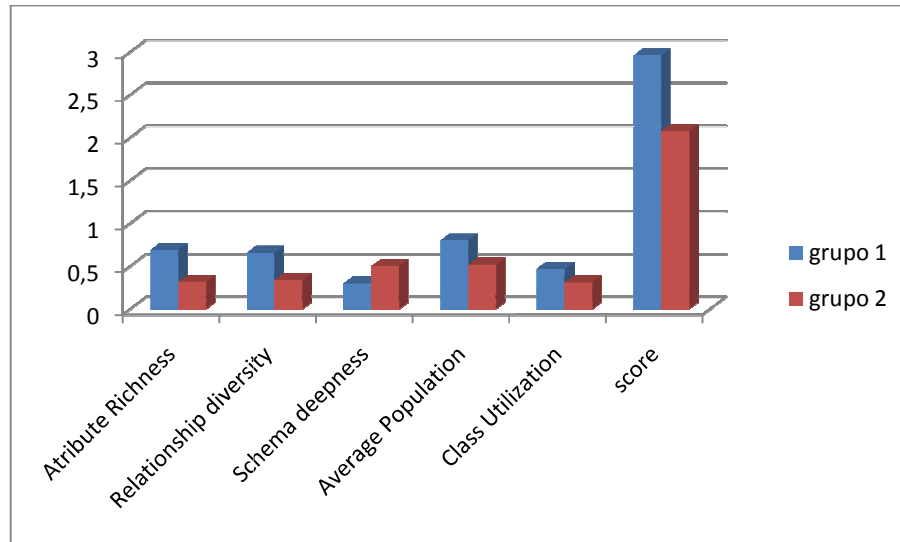


Figura 68 – Gráfico Comparativo da Avaliação Primeiro Experimento utilizando-se das Métricas da *OntoQA*.

A respeito do alinhamento das *tags* com elementos da ontologia raiz e das sugestões do questionário eletrônico, o entendimento dos projetistas pode ser diferente. Percebe-se uma tendência entre os projetistas do grupo 1 em criar mais instâncias e propriedades. Ao contrário do grupo 2 que seguiu a tendência em desenvolver mais conceitos e relações taxonômicas. A Figura 69 ilustra um gráfico comparativo dos erros taxonômicas relacionados à taxonomia idealizada pela *OntoClean*. Estes erros taxonômicos na especificação do modelo conceitual se propagam para as fases subsequentes do processo de desenvolvimento diminuindo a qualidade da ontologia produzida.

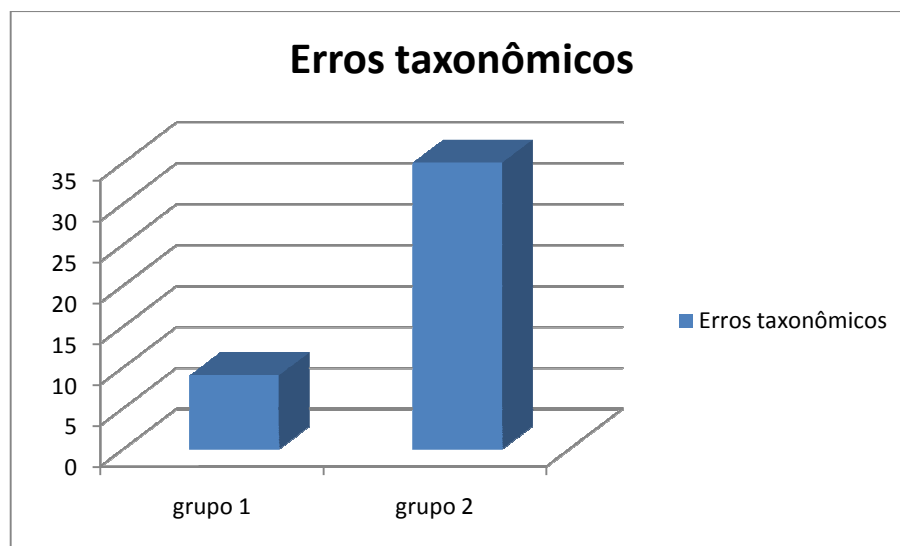


Figura 69 – Erros Taxonômicos no Desenvolvimento das Ontologias pelos Grupos no Primeiro Experimento.

Também foi verificado que o método ajudou aos projetistas do grupo 1 a conhecer a ontologia inicial sem que os mesmos tivessem atenção especial a este fato. Contrariamente, o grupo 2 teve um pouco de dificuldade em interpretar e inserir novos elementos na ontologia raiz. Este fato pode ser significado no reuso de ontologias.

#### 4.3 RESULTADOS DO SEGUNDO EXPERIMENTO

As ontologias desenvolvidas neste experimento também foram avaliadas de forma similar ao primeiro experimento. Este segundo experimento foi realizado por dois grupos de 3 projetistas cada, o que constitui um ambiente colaborativo. Neste caso, foram produzidas duas ontologias. Uma pelo grupo 1 que utilizou o método e outra pelo grupo 2 que desenvolveu a ontologia manualmente a partir das *tags* criadas pelos componentes do grupo.

Neste experimento, além de medir se a metodologia ajuda na aquisição de conhecimentos e conceitualização, pretende-se verificar se o método é um direcionador de discussão no desenvolvimento de uma ontologia.

A Tabela 12 apresenta a avaliação quantitativa do experimento. De acordo com a Tabela 12, o desempenho dos grupos foi muito semelhante com relação ao total de elementos criados. O grupo 1 criou 34 elementos e o grupo 2 criou 36 elementos. A principal diferença dos grupos está no número de conceitos e propriedades criados. O grupo 1 criou mais propriedades, enquanto que o grupo 2 criou mais conceitos.

Tabela 12 – Elementos criados nas Ontologias desenvolvidas no Segundo Experimento.

	Total de Tags	Conceitos	Propriedades	Instâncias	Total de Elementos
<u>grupo 1</u>	34	10	12	12	34
<u>grupo 2</u>	51	22	3	11	36
<b>Total</b>	85	32	15	23	70

Na Tabela 13 são apresentadas as quantidades de conceitos populados e de relações taxonômicas criadas nos grupos. O grupo 1 criou mais instâncias e populou mais conceitos, porém não produziu nenhuma relação taxonômica. O grupo 2, pelo fato de criar mais conceitos, produziu mais relações taxonômicas, porém populou menos conceitos.

Tabela 13 – Número de Conceitos Populados e Relações Taxonômicas do Segundo Experimento.

	Conceitos Populados	Relações Taxonômicas
<u>grupo 1</u>	9	0
<u>grupo 2</u>	5	17
<b>Total</b>	14	17

Com relação ao aproveitamento das *tags* (Tabela 14), apesar do grupo 1 ter utilizado menos *tags*, o mesmo criou um número semelhante de elementos. O grupo 1 aceitou 19 *tags*, ou seja, 19 *tags* induziram a criação de 34 elementos. Um índice de 1,78 elementos criados para cada *tag*. O grupo 2 criou 36 elementos de 19 *tags* aceitas. Um índice de 1,89 elementos criados para cada *tag*. O número de *tags* aceitas é semelhante nos dois grupos. A diferença maior dos grupos está nas *tags* ignoradas.

O grupo 1 ignorou 15 *tags* e o grupo 2 ignorou 32 *tags*. No grupo 1, conforme a Tabela 15, das 15 *tags* ignoradas, os projetistas não aceitaram as sugestões do alinhamento para 13 *tags* e as sugestões do questionário eletrônico para 2 *tags*. As 19 *tags* aceitas no grupo 1 indicam que as 19 sugestões do alinhamento foram aceitas em 34. Logo uma precisão do alinhamento em 55,88%. Com relação à precisão das sugestões no questionário eletrônico, os 19 alinhamentos geraram 19 sugestões, sendo que apenas 2 sugestões foram ignoradas. Assim, os projetistas aceitaram 17 sugestões do questionário, uma precisão de 89,47%.

Tabela 14 – Relação de *Tags*, Elementos, *Tags* Ignoradas, *Tags* Aceitas e Índice de Elementos Criados por *Tags* Aceitas no Segundo Experimento.

	Total de <i>Tags</i>	Total de Elementos Criados	<i>Tags</i> Ignoradas	<i>Tags</i> Aceitas	Elementos Criados/ <i>Tags</i> Aceitas
<u>grupo 1</u>	34	34	15 (44,12%)	19 (55,88%)	1,78 (34/19)
<u>grupo 2</u>	51	36	32 (66,74%)	19 (33,26)	1,89 (36/19)

Tabela 15 – Relação de *Tags* Aceitas, *Tags* Ignoradas e Sugestões Aceitas no Grupo 1 do Segundo Experimento.

	Total de <i>Tags</i>	<i>Tags</i> Ignoradas	<i>Tags</i> Ignoradas (alinhamento)	<i>Tags</i> Ignoradas (sugestão)	<i>Tags</i> Aceitas	Sugestões Aceitas
<b>grupo 1</b>	34	15 (44,12%)	13 (86,67%)	2 (13,33%)	19 (55,88%)	17 (89,47%)

Com base nas Tabelas (Tabela 12 e Tabela 13) calculou-se a qualidade das ontologias desenvolvidas. A Tabela 16 relaciona os valores para cada métrica. Percebe-se que apesar de muita semelhança na quantidade de elementos criados, as medidas do grupo 1 foram superiores. O grupo 1 teve apenas um valor inferior na métrica *schema deepness*. A Figura 71 ilustra um gráfico da qualidade das ontologias desenvolvidas neste segundo experimento.

Tabela 16 – Avaliação das Ontologias do Segundo Experimento com relação as Métricas da *OntoQA*.

	Métricas					
	<i>Attribute Richness</i>	<i>Relationship Diversity</i>	<i>Schema Deepness</i>	<i>Average Population</i>	<i>Class Utilization</i>	<i>score</i>
<b>grupo 1</b>	1,12	1	0	1,2	0,9	4,3
<b>grupo 2</b>	0,13	0,15	0,77	0,5	0,22	1,78

Neste segundo experimento, também foi solicitado o preenchimento de um questionário para o grupo 1 que utilizou o Folkconcept. Para cada novo elemento criado, foi questionado ao grupo sobre o que influenciou a criação do novo elemento na ontologia. Para esta questão havia 4 alternativas de repostas:

- (1) A *tag* etiquetada (relação de *tags*) (a etiquetagem ajudou a adquirir novos conhecimentos).
- (2) Uma nova entidade criada anteriormente (o que já foi criado nesta lista influenciou).
- (3) Conhecimento do domínio anterior a etiquetagem (conhecimentos que você já possuía do domínio anterior a etiquetagem).
- (4) Discussão no grupo.

Nesta questão os projetistas puderam escolher mais de uma alternativa para cada elemento criado. A Tabela 17 ilustra o resultado obtido para cada alternativa. Os projetistas

informaram que a *tag* etiquetada e a discussão no grupo foi a principal motivação para criar um novo elemento. O conhecimento anterior sobre o domínio influenciou muito pouco a criação do novo elemento. A segunda alternativa, um elemento criado anteriormente, não foi levada em consideração para nenhum novo elemento.

Tabela 17 – Questionário sobre a criação de um Novo Elemento no Segundo Experimento grupo 1.

<b>O que influenciou a criação do novo elemento na ontologia raiz?</b>	
<b>Alternativa</b>	<b>Número de escolhas</b>
<b>1</b>	<b>20</b>
<b>2</b>	<b>0</b>
<b>3</b>	<b>4</b>
<b>4</b>	<b>25</b>

#### **4.3.1 Síntese do Segundo Experimento**

Houve uma semelhança dos grupos no número de elementos criados, porém a tendência do grupo 1 foi criar mais propriedades do que conceitos. Nas instâncias, as quantidades criadas foram semelhantes, sendo a diferença maior nos conceitos populados. O grupo 1 populou mais classes do que o grupo 2. Isto deve-se ao alinhamento que cria conexões que devem ser interpretadas pelo projetista. Uma observação importante é que o grupo 1 criou poucos conceitos e praticamente nenhuma relação taxonômica. A Figura 70 ilustra um gráfico comparativo entre grupos 1 e 2 dos elementos criados no segundo experimento.

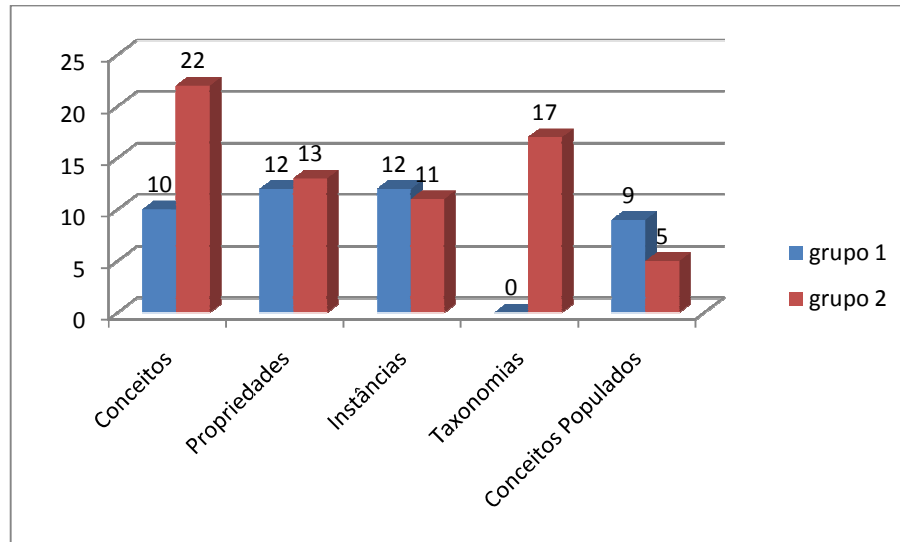


Figura 70 – Gráfico Comparativo dos Elementos Criados no Segundo Experimento.

O grupo 1 utilizou menos *tags* que o grupo 2 porque além das discussões entre os elementos, o grupo 1 precisava responder ao questionário eletrônico na conceitualização. Mesmo assim, o número de elementos foi muito semelhante e as métricas de qualidade foram superiores ao grupo 2. A única métrica em que o grupo 2 foi superior foi a *schema deepness*. Isto devido a quantidade de relações taxonômicas. Semelhante ao primeiro experimento, as ontologias produzidas pelo grupo 1 foram mais específicas (maior profundidade) com uma descrição mais detalhada dos conceitos.

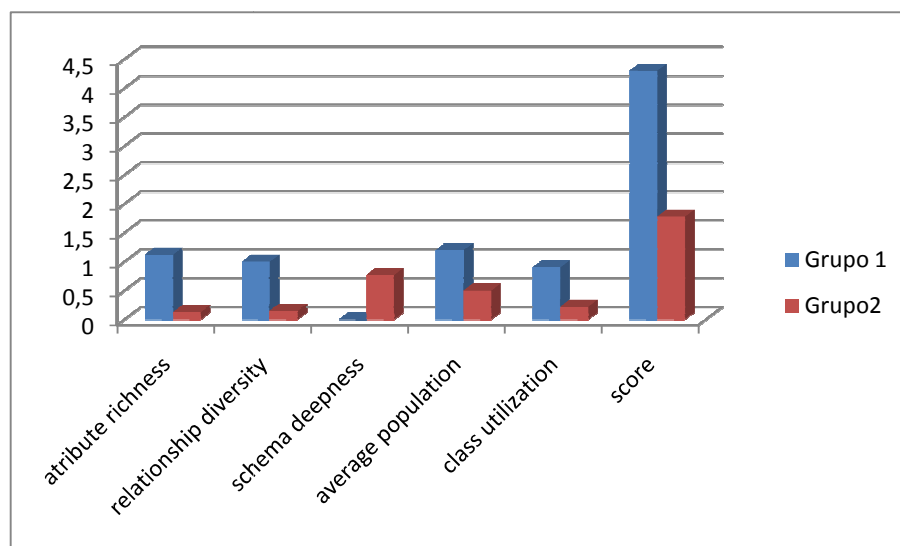


Figura 71 – Gráfico Comparativo da Avaliação do Segundo Experimento em relação as Métricas *OntoQA*.

A Figura 72 ilustra o aproveitamento das tags relacionado ao total de elementos criados a partir das mesmas. Percebe-se a semelhança no número de elemento, porém o grupo 2 utilizou um número superior de tags. Conclui-se que o aproveitamento das tags pelo grupo 1 foi superior ao grupo 2.

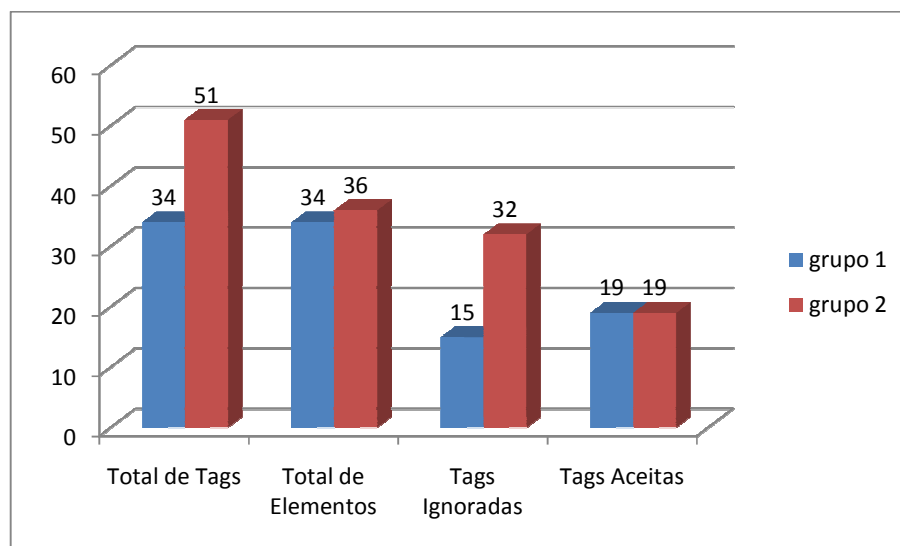


Figura 72 – Gráfico do Aproveitamento de Tags no Segundo Experimento.

A Figura 73 ilustra um gráfico relacionado apenas ao aproveitamento das tags no grupo 1. Verificou-se onde as tags foram ignoradas e aceitas. Das 34 tags aceitas, 19 foram aceitas no alinhamento. Destas 19, que geraram 19 sugestões, apenas 2 sugestões foram ignoradas no questionário eletrônico.

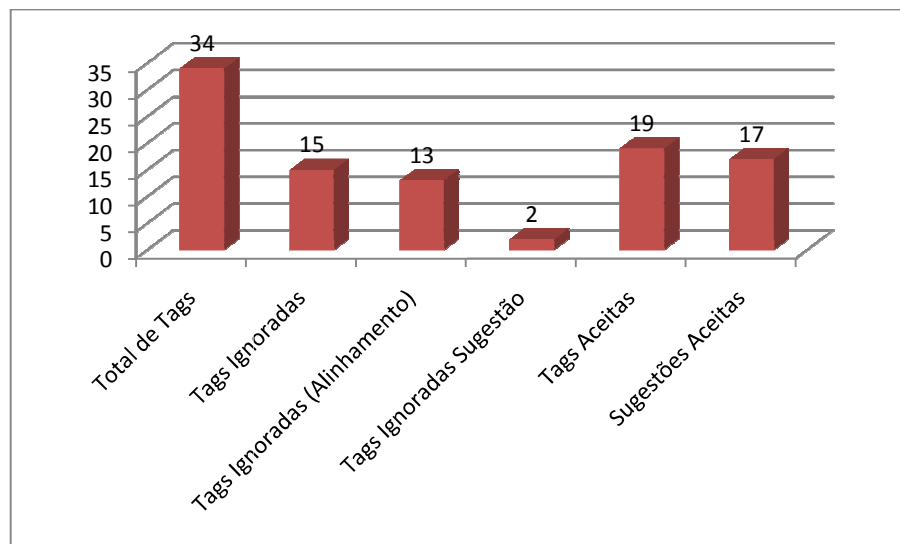


Figura 73 – Gráfico do Aproveitamento de Tags para o Grupo 1 no Segundo Experimento.

#### 4.4 DISCUSSÕES DOS EXPERIMENTOS

Realizado os experimentos, procurou-se responder as questões estipuladas no início deste capítulo 4, ou seja, validar as seguintes hipóteses:

- O método desenvolvido ajuda na capacidade cognitiva dos atores diminuindo o problema de aquisição de conhecimentos e tornando mais fácil a conceitualização e explicitação de tal conceitualização?
- O método desenvolvido (aquisição de conhecimentos e conceitualização com a integração de um processo de avaliação) baseado nas folksonomias ajuda na elaboração de ontologias mais descritivas e com formalismo correto baseado nas consequências ontológicas?

Na **primeira questão**, foi avaliado se o Folkconcept auxilia os atores a organizarem seus conhecimentos, mais precisamente, se o processo de anotação e consequente geração de uma folksonomia, auxilia na conceitualização e na especificação do modelo conceitual.

Ao verificar o desvio padrão relativo ao número de elementos criados a partir das *tags* da folksonomia, os projetistas do grupo 1 não apresentaram grande dispersão em relação à média, ou seja, todos tiveram suporte para criar os vários tipos de elementos na ontologia de forma similar. Percebe-se que os projetistas do grupo 2 foram muito dispersos em relação ao



número de elementos criados, sendo que uns se concentraram mais em criar conceitos, outros em populá-los e assim por diante. Conclui-se que Folkconcept, por meio do **alinhamento** de *tags* da folksonomia com elementos da ontologia, auxiliou os atores do grupo 1 a refletir sobre os diversos tipos de elementos ontológicos de forma homogênea, auxiliando-os na tarefa de identificar classes, instâncias, relações taxonômicas e propriedades.

Em relação às folksonomias propriamente ditas, percebe-se que auxiliam na tarefa de conceitualização considerando que, em ambos experimentos independentemente do grupo, a aceitação das *tags* na geração de elementos da ontologia foi significativa, ficando no intervalo de 55% a 65% em três dos quatro experimentos. Além disso, em três dos quatro experimentos o índice de elementos criados na ontologia por *tag* aceita ficou no intervalo de 1,53 a 1,89. Isto mostra que as folksonomias são importantes não só na seleção de termos candidatos a conceitos, mas também na descoberta de instâncias, relações taxonômicas e propriedades. Ainda sobre a importância da folksonomia na conceitualização, recorda-se que no experimento 2 os participantes do grupo 1 relataram por meio de um questionário que as *tags* os auxiliaram na criação de elementos na ontologia, sendo a segunda opção mais escolhida dentre 4 outras.

Ressalta-se que os números que revelam a importância das folksonomias na conceitualização poderiam ter sido melhores se os participantes dos experimentos tivessem sido informados sobre como tais *tags* seriam utilizadas na especificação do modelo conceitual. No momento da etiquetagem, eles não possuíam uma visão panorâmica do método. A abordagem utilizada simula, portanto, o caso de uma comunidade que converge para um vocabulário por meio de uma aplicação social e decide sistematizar seus conhecimentos por meio de uma ontologia.

A **segunda questão** é se a aquisição de conhecimentos baseada nas folksonomias ajuda na elaboração de ontologias mais descritivas e corretas do ponto de vista das relações taxonômicas idealizadas por *OntoClean*.

Pelas métricas da *OntoQA*, os projetistas do grupo 1, que utilizaram o Folkconcept, criaram ontologias mais específicas com um maior nível de detalhe para os conceitos, principalmente nos relacionamentos de propriedade e na população dos mesmos. Percebe-se que o grupo 1 procurou conectar os conceitos entre si por relações taxonômicas e, principalmente, pela criação de propriedades evitando que conceitos ficassem isolados na hierarquia. Isto se deve particularmente ao questionário eletrônico do Folkconcept que induziu os projetistas do grupo 1 a agirem desta forma.

O grupo 2 criou mais conceitos, porém alguns permaneceram isolados na hierarquia *Thing*, ou seja, sem um relacionamento hierárquico e sem um relacionamento de propriedade. Os conceitos foram criados livremente e o projetista não se preocupava em relacioná-los com o restante da ontologia. No grupo 1 o projetista também não possuía esta preocupação, porém o método auxiliou-o nesta tarefa. Logo, conforme tabelas (Tabela 9 e Tabela 16), nos dois experimentos a qualidade (em relação a descrição dos conceitos e população) das ontologias avaliada pelas métricas foi superior no grupo 1.

Na parte extensional da ontologia, o grupo 1, que utilizou o Folkconcept nos dois experimentos, populou mais conceitos e criou mais instâncias. Assim as métricas *average population* e *class utilization* foram superiores as do grupo 2. Conclui-se que o experimento apresentou indícios de que o método auxilia na população da ontologia pelo número de instâncias e conceitos populados. Outro fato é que o método auxilia a distribuição de instâncias entre os conceitos, ou seja, exemplifica o conhecimento extensional incluindo um número adequado de instâncias em relação ao número de conceitos.

Com relação à parte intencional da ontologia, o grupo 1 criou mais propriedades, procurou enriquecer a ontologia raiz de forma que a mesma seja mais descritiva, ou seja que todos os conceitos estejam relacionados por propriedades ou por relações taxonômicas.

Por meio das métricas de largura, profundidade e emaranhamento dos nós, foi observado que o grupo 1 teve ontologias mais profundas e sem emaranhamentos. O grupo 2 gerou ontologias mais rasas e com erros de emaranhamento. Tanto o grupo 1 quanto o grupo 2 geraram ontologias de larguras similares. Logo as ontologias do grupo 1 possuem uma natureza vertical (conhecimento mais detalhado) e com uma variedade de conhecimento (natureza horizontal) semelhante as ontologias do grupo 2. As ontologias do grupo 2, por não apresentar uma profundidade inferior às ontologias do grupo 1, adequada, apresentam conceitos descendentes com pouco similaridade, logo a ontologia possui menor nível de detalhes.

Com relação ao tipo de ontologia produzida, verifica-se que as ontologias do grupo 1 são mais detalhadas, mais densas (conceitos conectados por propriedades), mais específicas e mais descritivas (GANGEMI et al., 2006). Isto pode ser verificado pela proporção de propriedades em relação aos conceitos identificados e pela profundidade das relações taxonômicas. Logo, são ontologias não leves. Em contraposição, foi observado que o grupo 2 desenvolveu ontologias mais gerais com pouco nível de detalhe, ou seja, ontologias leves. Portanto, o método pode auxiliar no desenvolvimento de ontologias de domínio e

principalmente aplicação com superioridade nos níveis de detalhes (diferente de ontologias leves).

Outra avaliação realizada foi relacionada aos erros de modelagem. As ontologias do grupo 1 possuem menos erros de modelagem quanto à taxonomia idealizada *OntoClean*. Estes erros indicam uma taxonomia correta em relação à natureza ontológica dos conceitos.

Conclui-se que o Folkconcept é indicado para construir ontologias de domínio (conhecer um domínio e construir suas relações, especializar conceitos, construir novas relações, descrever conceitos – capacidade de representar um conhecimento), continuar o desenvolvimento de ontologias de nível superior (expandir e popular a ontologia, popular um grande número de classes, muitas instâncias que pertencem as classes) e, para construir ontologias de aplicação (incluir conhecimentos mais específicos, descrever melhor os conceitos – específicos - pela inclusão de propriedades – incluir toda a informação necessária à aplicação em particular) de uma forma coerente a natureza ontológica dos elementos.



## 5. CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho, apresentou-se um método, denominado Folkconcept, para o desenvolvimento de ontologias que identifica novos elementos lógicos formais a partir de dados informais presentes em uma folksonomia. Esta é uma alternativa para atenuar o gargalo na aquisição de conhecimentos.

O trabalho desenvolvido avança o estado da arte no desenvolvimento de ontologias, mais especificamente nas metodologias, no **aprendizado de ontologias** a partir de textos, no aprendizado de ontologias a partir de folksonomias e no desenvolvimento de ontologias a partir de modelos.

Com relação as **metodologias**, procurou-se fornecer suporte metodológico no desenvolvimento individual de ontologias, visto que as atividades de aquisição, conceitualização e avaliação estão presentes em metodologias de desenvolvimento.

Os seguintes pontos foram explorados e contribuem em relação as metodologias:

- **Folksonomias como alternativa para atenuar o gargalo na aquisição de conhecimentos:** embora as metodologias de desenvolvimento procurem reduzir o gargalo na aquisição de conhecimentos, o mesmo persiste na dificuldade em capturar e estruturar o conhecimento do domínio. Além disso, adquirir conhecimentos para construir ontologias ainda requer tempo e recursos. Logo, folksonomias podem ser consideradas artefatos intermediários de modelagem conceitual resultantes da leitura e anotação de recursos. Leitura e anotação são técnicas simples de aquisição de conhecimentos e podem ser incorporadas nas metodologias. Algumas metodologias (*e.g.* METHONTOLOGY, *On-To-Knowledge*) preconizam a leitura de documentos e textos relativos ao domínio para a especificação da ontologia e conhecimento do domínio. Outras utilizam modelagem conceitual (*e.g.* OntoUML, UPON). A NEON e a DILIGENT não especificam qual técnica de aquisição. Enfim, os artefatos na aquisição de conhecimentos presente nas metodologias incluem tabelas, questões de competência, modelos conceituais, entre outros. As metodologias preconizam a utilização de alguma técnica de aquisição, mas nenhuma metodologia utilizam como técnicas de aquisição a leitura e anotação de recursos para produzir o artefato folksonomia. Logo, o proposto para aquisição de conhecimentos é uma alternativa,

pois o método fornece diretrizes de como desenvolver ontologias a partir de folksonomias.

- **Inclusão de um método semiautomático de aprendizado como objetivo de reduzir esforços na aquisição de conhecimentos:** A única metodologia que preconiza a utilização de alguma técnica semiautomática de aprendizado é a *On-To-Knowledge*, mesmo assim, não especifica qual a técnica de aprendizado utilizada e também como será a validação da ontologia gerada. No método desenvolvido, além da aquisição automática de conhecimentos, há um processo de estruturação e validação da ontologia gerada. As folksonomias são compostas por termos sem semântica, que pelo processo de aprendizado desenvolvido, são enriquecidas, representadas e utilizadas com o objetivo de induzir elementos na ontologia em desenvolvimento. Além disso o método semiautomático também possui como objetivo reduzir esforços no processo de aquisição de conhecimentos.
- **Desenvolvimento de uma ontologia a partir de uma já existente:** uma ontologia pode ser desenvolvida ao longo de um período. Durante este período a ontologia recebe constantes alterações. A METHONTOLGY preconiza o desenvolvimento de uma ontologia em vários ciclos. A cada ciclo, a ontologia é enriquecida e uma nova versão é produzida. O método desenvolvido mostra como enriquecer uma ontologia a medida que novos conhecimentos são aprendidos a partir da leitura e anotação de recursos e conseqüentemente geração das folksonomias. Logo, possui semelhança nas metodologias (METHONTOLOGY, *On-To-Knowledge*) pelo desenvolvimento iterativo, sendo diferente pelas técnicas de enriquecimento atreladas a aquisição de conhecimentos com a produção de uma nova folksonomia, que pode encadear alterações na ontologia em desenvolvimento.
- **O desenvolvimento a partir de uma ontologia de nível superior:** esta alternativa não foi trabalhada nos experimentos, porém ao considerar a ontologia raiz como uma ontologia de nível superior, pode-se utilizar o método como auxílio na poda e escolha da região da ontologia de nível superior que será utilizada na ontologia em desenvolvimento. Cyc, SENSUS e *Enterprise* descrevem o desenvolvimento a partir de uma ontologia de nível superior, porém ocorrem dificuldades na seleção extensão da ontologia. OntoUML utiliza definições da UFO para desenvolver os modelos conceituais manualmente. Diferente das metodologias, o método propõe o aterramento tanto da ontologia em desenvolvimento quanto da ontologia de nível superior a fontes de informação, nesse caso folksonomias, pelo aterramento de

*tags*. Logo, o conhecimento utilizado para desenvolver a ontologia (identificar novos elementos) também é utilizado para selecionar elementos da ontologia de nível superior que podem ser utilizados.

- **Desenvolvimento colaborativo de ontologias:** nesta opção há algumas metodologias (*e.g.* DILIGENTE, NEON) que preconizam este tipo de desenvolvimento. Porém o que estas metodologias proporcionam é um ambiente colaborativo, no qual os atores desenvolvem ontologias locais e as integram numa única versão por meio de um processo de argumentação e conceitualização conjunta de uma ontologia. Além disso, estas metodologias corporativas preconizam a utilização de metodologias individuais (*i.e* atividades) que estão presentes em outras metodologias. Por exemplo, a NEON possui atividades de aquisição de conhecimentos que estão presente na METHONTOLOGY. No método desenvolvido, pelo fato da ontologia raiz ser desenvolvida por uma folksonomia inicial que pode ser produzida colaborativamente, o mesmo é uma alternativa ao desenvolvimento colaborativo. Os projetistas utilizam suas *tags* e também *tags* de outros projetistas na aquisição e conceitualização da ontologia. Portanto, os recursos e conhecimentos existentes (folksonomias) podem ser resultantes de um processo consensual utilizado no desenvolvimento colaborativo de ontologias.
- **Avaliação no desenvolvimento das ontologias:** as metodologias (METHONTOLOGY, *On-To-Knowledge*) utilizam uma avaliação funcional relacionada a cobertura da ontologia. Realiza-se esta avaliação confrontando a ontologia com questões de competência ou simulando cenários. Procurou-se no método integrar um processo de avaliação que garantisse a corretude da ontologia relacionada as taxonomias produzidas. Esta corretude está relacionada a natureza ontológica dos elementos preconizada pela *OntoClean*. Ao capturar a natureza ontológica dos elementos, busca-se o significado pretendido do conceito e suas relações. Além disso, utiliza-se o processo de avaliação integrado à conceitualização para evitar a propagação de erros em atividades subsequentes do ciclo de vida da ontologia e das próximas ontologias raiz que serão enriquecidas.
- **Desenvolvimento a partir do reuso:** no método desenvolvido, considera-se a utilização de conhecimentos ontológicos que incluem uma ontologia de domínio original que será enriquecida e a utilização da taxonomia idealizada pela *OntoClean* como modelo para a estruturação dos novos elementos na ontologia.

Diferente das abordagens que utilizam padrões (*e.g.* OntoUML, NEON), utiliza-se a metodologia *OntoClean* para verificar se a ontologia é válida de uma perspectiva formal (regras ontológicas). OntoUML é uma linguagem de modelagem conceitual comprometida com uma meta-conceitualização (representada na ontologia de referência UFO) que permite especificar modelos conceituais que retratem de maneira apropriada uma conceitualização. O reuso, neste caso, se refere à meta-conceitualização. Na OntoUML utiliza-se a UFO, inspirada nos princípios metodológicos de *OntoClean*, enquanto *Folkconcept* é totalmente baseada na *OntoClean*. Embora a NEON reforça a idéia de desenvolvimento por reuso, não aborda em profundidade a aplicação de padrões por se tratar de uma metodologia genérica (MAASS; JANZEN, 2009).

O Quadro 11 mostra as diferenças do método *Folkconcept* em relação a técnicas e artefatos existentes na literatura (conforme apresentado na fundamentação teórica). O quadro mostra as técnicas e artefatos correspondentes empregados no *Folkconcept* na aquisição de conhecimentos, nas técnicas de aprendizado de ontologias, no enriquecimento de uma ontologia, no desenvolvimento a partir de uma ontologia de nível superior, no desenvolvimento colaborativo e pelo reuso.

Quadro 11 – Comparação do Métodos em relação as Metodologias nas Técnicas utilizadas e Artefatos produzidos.

<b>O que diferencia Folkconcept das abordagens</b>		
	<b>Técnica</b>	<b>Artefato</b>
<b>Aquisição de Conhecimentos</b>	<b>Técnicas de Leitura e Anotação de recursos.</b>	<b>Folksonomias</b>
<b>Técnicas de Aprendizado de Ontologias</b>	Método de aprendizado de ontologias semiautomático a partir de folksonomias que consiste no enriquecimento, representação das folksonomias e indução de novos elementos pela utilização do alinhamento e questionário eletrônico.	<ul style="list-style-type: none"> <li>- Folksonomia enriquecida (incluir semântica aos termos presentes na folksonomia) como artefato de entrada para conceitualização de novos elementos.</li> <li>- Indução/Sugestão de novos elementos na ontologia.</li> </ul>
<b>Enriquecimento de uma Ontologia</b>	Técnicas de aquisição de conhecimentos e conceitualização. Conexão de elementos a tags na folksonomia.	<ul style="list-style-type: none"> <li>- Nova folksonomia resultante do aprendizado de novos conhecimentos pelos projetistas.</li> <li>- Nova ontologia resultante do processo</li> </ul>



		semiautomático de aquisição e conceitualização.
<b>Desenvolvimento a partir de uma Ontologia de nível superior (considerando a raiz) (alternativa)</b>	Atrair as tags a folksonomias, método como auxílio na escolha das definições ontológicas para a ontologia de domínio em desenvolvimento.	- Ontologia com elementos conectados a tags e consequente elementos em uma ontologia de nível superior.
<b>Desenvolvimento colaborativo (alternativa)</b>	Técnicas de leitura anotação de recursos. Aquisição e conceitualização.	- Folksonomias resultantes de um processo consensual de anotação (correlação de tags na anotação de recursos). - Ontologia resultante pela aplicação do métodos as folksonomias geradas no processo consensual de anotação.
<b>Desenvolvimento pelo reuso</b>	Utilização de conhecimento ontológico (ontologia de domínio existente e taxonomia idealizada pela <i>OntoClean</i> ) e conhecimento não ontológico (folksonomias).	Evolução da folksonomia Corretude taxonômica da ontologia.

O **aprendizado de ontologias** consiste de um conjunto de métodos e técnicas para se desenvolver ontologias de maneira semiautomática utilizando diversas fontes de informação (GÓMEZ-PÉREZ; MANZANO-MACHO, 2003; MAEDCHE; STAAB, 2000a; CIMIANO, 2006). Por semiautomático entende-se que o projetista precisa validar e avaliar os novos elementos sugeridos pelos métodos de aprendizado. Para tal, deve conhecer o domínio. Logo, mesmo utilizando métodos de aprendizado para aquisição de conhecimentos, a dificuldade em estruturar o conhecimento do domínio persiste e, consequentemente, persiste o gargalo na aquisição de conhecimentos. Todos os métodos de aprendizado procuram seguir o mesmo padrão: transformar/traduzir termos em conceitos e relações. No método desenvolvido inclui-se o aprendizado de ontologias que possui como objetivo auxiliar a geração de novos elementos na ontologia em desenvolvimento. O objetivo foi fornecer diretrizes de como enriquecer uma ontologia existente a partir de um processo de aprendizado de ontologias.

No método desenvolvido, a identificação dos novos elementos e relações, segue um fluxo direcionado pelas respostas do projetista a um questionário eletrônico. Este questionário induz o projetista a criar novos elementos bem como algumas relações que podem ser formalizadas e implementadas na ontologia. Estes elementos e suas relações são consistentes

porque obedecem as restrições ontológicas da *OntoClean*. Logo, além de ajudar a adquirir conhecimentos, ou seja, conhecer o domínio, o método também ajuda os projetistas a conceitualizarem os elementos de forma consistente.

Outra questão levantada nos métodos semiautomáticos é que os projetistas precisam validar os elementos e suas relações, visto que ocorrem vários erros de conceitualização. No método desenvolvido, o projetista valida apenas a relação da *tag* com um elemento da ontologia raiz, ou seja, ele verifica se a relação e a região da ontologia induzem alguma estruturação ou uma conceitualização. Quem pode afirmar se uma *tag* é um novo elemento e que produz uma relação com outro elemento é o próprio projetista. Com isso, o método procura atacar o problema de conceitualização no momento do desenvolvimento da ontologia interagindo com o projetista e guiando o mesmo a produzir novos elementos de forma consistente.

O método apresentado para o aprendizado de ontologias é diferente dos existentes atualmente (GÓMEZ-PÉREZ; MANZANO-MACHO, 2003; AUSSENAC-GILLES, BIÉBOW; SZULMAN, 2000; KIETZ; MAEDCHE; VOLZ, 2000; MAEDCHE; STAAB, 2001; SÁNCHEZ; MORENO, 2008; ALANI et al., 2003; KHAN; LUO, 2002; CIMIANO; WENDEROTH, 2005; VILLAVERDE et al., 2009), por utilizar folksonomia, pelo processo de representação e enriquecimento de *tags*, pela similaridade entre as *tags* calculada por medidas de similaridade, pela interação com o projetista que identifica as noções filosóficas subjacentes aos elementos da ontologia e pelas alterações conceituais que obedecem as consequências ontológicas.

Outra diferença do método desenvolvido para os métodos de aprendizado a partir de textos, é que estes últimos são abordagens supervisionadas que utilizam alguma tecnologia, como redes neurais e árvores de decisão, para identificar elementos e suas relações. Isto exige grandes quantidades de informação como exemplos de aprendizado. Com isso, textos que não foram criados pelos projetistas, muito menos verificados, são utilizados. Além disso, os métodos verificam relações que precisam ser validadas pelos projetistas. As técnicas utilizadas nestes métodos adotam o pressuposto que documentos geralmente possuem a mesma estrutura, bem como conteúdo. Assim, não consideram a realidade, ou seja, a grande quantidade e heterogeneidade de informação na *Web* atual (RUNG-CHING; JUI-YUAN; REN-HAO, 2008).

O Quadro 12 resume as contribuições seguintes em relação aos trabalhos relacionados: fonte de conhecimento utilizado pelo método de aprendizado, validação dos elementos e

relações, conceitualização, técnicas, estrutura gerada, integração a um processo de desenvolvimento, rastreabilidade com a origem do conhecimento.

Quadro 12 – Resumo das Contribuições de Folkconcept em relação aos Métodos de Aprendizado a partir de Textos.

	Diferença para as abordagens de aprendizado
<b>Fonte</b>	Folksonomias
<b>Avaliação</b>	Avaliação integrada na conceitualização. Aplicação de forma transparente da <i>OntoClean</i> pela utilização de um questionário.
<b>Conceitualização</b>	Alinhamento de tags e Questionário
<b>Técnicas que utiliza</b>	Relação com outras fontes, Alinhamento, Similaridades e Quesitonário eletrônico, PLN.
<b>Estrutura gerada</b>	Conceitos, taxonomias, propriedades e instâncias.
<b>Integrado a um processo de desenvolvimento</b>	Estruturado em atividades presentes nas metodologias, entre elas: aquisição de conhecimentos, conceitualização e avaliação.
<b>Conexão da fonte a ontologia</b>	Rastreabilidade entre a origem do conhecimento (folksonomia) e ontologia pela referência das <i>tags</i> a elementos na ontologia.
<b>Enriquecimento e representação da fonte</b>	Importação das tags associadas a outras fontes com o objetivo de enriquecer a representação das <i>tags</i> (fornecer semântica).

Após comparar o método desenvolvido com parte das metodologias e com métodos de aprendizado a partir de textos, compara-se o trabalho desenvolvido com trabalhos relacionados ao desenvolvimento de ontologias a partir de folksonomias (MIKA, 2005; WU; ZHANG; YU, 2006; BEGELMAN; KELLER; SMADJA, 2006; SCHMITZ, 2006; WU; ZUBAIR; MALY, 2006; LUX; DSINGER, 2007; ANGELETOU et al., 2007; SPECIA; MOTTA, 2007; LANIADO; EYNARD; COLOMBETTI, 2007; SILVA, 2009; DAMME; HEPP; SIORPAES, 2007). O Quadro 13 ilustra os principais pontos abordados na comparação com abordagens que procuram gerar algum tipo de estrutura a partir de folksonomias. Os pontos abordados são, o uso das folksonomias, a estrutura desenvolvida, técnicas utilizadas e validação dos resultados.

Quadro 13 – Comparação em relação ao desenvolvimento de ontologias a partir de folkonomias.

Abordagens existentes	Folkconcept

<b>Uso das folksonomias</b>	Indexação de Recursos	Conhecimento do domínio e geração de ontologias.
<b>Estrutura gerada</b>	Tags correlacionadas, grupos de tags e taxonomias relacionados à correlação.	Geração de conceitos, relações de taxonomia, relações de propriedade e instâncias.
<b>Técnicas que utiliza</b>	Correlação de tags e relações extraídas de dicionários.	Correlação de tags, FCA, relação com outras fontes (dicionários), similaridades, natureza ontológica da <i>OntoClean</i> .
<b>Validação</b>	Não especificam	Questionário eletrônico baseado na taxonomia idealizada pela <i>OntoClean</i> .

A Figura 74 ilustra o escopo do trabalho apresentado. Da *Web Social* exploram-se as folksonomias e as aplicações que permitem anotar recursos eletrônicos. Da *Web Semântica*, as ontologias, mais especificamente, ontologias de domínio e ontologias de aplicação. Completando, no método desenvolvido, exploram-se algumas técnicas que incluem processamento de linguagem natural, medidas de similaridade, entre outras. Em outras palavras, o vazio existente entre as tecnologias, a *Web Social* e a *Web Semântica* utiliza o aprendizado/desenvolvimento de ontologias a partir de folksonomias. O aprendizado gera/enriquece ontologias para a *Web Semântica* a partir de folksonomias produzidas por um processo de anotação na *Web Social* e com auxílio de tecnologias/técnicas.

Resumindo, o foco principal desta tese foi um método de desenvolvimento de ontologias a partir de folksonomias que inclui o aprendizado de ontologias. Este possui atividades de aquisição de conhecimentos a partir de folksonomias e conceitualização que estrutura o conhecimento do domínio identificando características filosóficas dos novos elementos para que os mesmos possam ser inseridos, de forma consistente na ontologia em desenvolvimento.



Figura 74 – Escopo do Trabalho Desenvolvido.

O método foi desenvolvido e foram realizados experimentos com projetistas. Pelos experimentos verificou-se quantitativamente as ontologias desenvolvidas. Pelas hipóteses apresentadas, as mesmas foram comprovadas pelo valor das variáveis. Verificou-se que o método auxilia na geração de conhecimentos, mais especificamente ontologias mais descritivas e com menos erros quando relacionados a natureza ontológica dos elementos.

## 5.1 TRABALHOS FUTUROS

Neste trabalho, alguns aspectos não abordados podem originar outros trabalhos. Dentre os aspectos não abordados, pode-se citar: integração da *Web Social* com a *Web Semântica*, desenvolvimento colaborativo de ontologias (FREDDO et al., 2009; TACLA et al., 2010), desenvolvimento de uma ontologia a partir de uma ontologia de topo, evolução de ontologias (*ontology changes*), controle de versões e rastreabilidade.

Na integração da *Web Social* com a *Web Semântica*, as folksonomias representam a *Web Social* pelo processo de anotação de recursos. A conexão da *tag* a elementos em uma ontologia pode trazer novos benefícios tanto para a *Web Social* quanto para a *Web Semântica*. Na *Web Social*, benefícios por possibilitar a anotação de recursos e formulação de *queries* semânticas que aumentam a precisão dos resultados trazidos pelos motores de busca. Na *Web Semântica* por possibilitar o desenvolvimento de novas ontologias, com isso aumentar seu uso. A *Web Social* pode sobrepujar algumas de suas limitações por explorar a tecnologia da *Web Semântica* na pesquisa, navegação e integração de informações na *Web*. Na *Web Social* o

significado das anotações não é especificado. Segundo Gruber (2008), necessita-se de tecnologias para o raciocínio em folksonomias de tal forma que este raciocínio descubra e conclua novas coisas/elementos. Na estrutura das folksonomias não há significado conceitual nas relações entre as *tags*. As ontologias podem atacar o problema de ambiguidade e organização/estrutura plana das folksonomias.

A *Web Semântica* pode se beneficiar com as habilidades das aplicações sociais para anotação de recursos *Web*. O uso de ontologias para anotar recursos *Web* possui algumas desvantagens (WU; ZHANG; YU, 2006): desenvolver um ontologia comum para um grande número de recursos não é comum porque diferentes usuários e aplicações possuem diferentes visões do que significa e do que estes recursos disponibilizam; requerem ao usuário certa habilidade e conhecimento em engenharia de ontologias o qual é um requerimento alto para simples usuários da *Web*. Outro problema é que engenheiros e especialistas possuem dificuldades em manter a consistência entre as ontologias e os recursos anotados pelas mesmas. É necessário capturar automaticamente as alterações nos recursos para manter a ontologia atualizada. Além disso, a falta de uma classe imaginada ou uma anotação errada quando os usuários anotam recursos é um problema recorrente. A *Web* integrada combina as facilidades em anotar recursos por *tags* em aplicações da *Web Social* com ontologias que melhor descrevem estes recursos na *Web Semântica*. Segundo Berners-Lee, Hendler e Lassila (2001), "*the Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation*".

No desenvolvimento colaborativo de ontologias, há várias possibilidades. Uma delas se fundamenta no fato que se uma grande quantidade de usuários anota um mesmo recurso de forma similar, então estas anotações podem refletir o conhecimento consensual. A folksonomia resultante da atividade de uma comunidade de usuários pode apresentar sobreposições de *tags* evidenciando a emergência de conhecimento consensual (*i.e.* vocabulário compartilhado). Outra possibilidade, abordada neste trabalho, é o fato das *tags* iniciais serem aterradas à ontologia raiz. Esta é desenvolvida com a utilização de *tags* de vários projetistas. Neste caso, um projetista pode utilizar *tags* de outros projetistas. Em uma primeira etapa desenvolve-se uma ontologia raiz inicial. Após, cada projetista se aprofunda no domínio realizando novas anotações que dão continuidade ao desenvolvimento colaborativo da ontologia raiz inicial.

Com relação à integração ou desenvolvimento de ontologias a partir de uma ontologia de topo, o método pode ser uma alternativa. O método desenvolvido parte de uma

ontologia raiz. Esta ontologia pode ser uma ontologia inicial desenvolvida pelos projetistas, ou até uma ontologia de alto nível. Isto resolveria alguns problemas relativos à integração e em como posicionar uma ontologia de domínio em uma ontologia de topo.

Evolução ao longo do tempo é um requisito essencial para as ontologias. Evolução de ontologias é definida por (HAASE; STOJANOVIC, 2005) como “*timely adaptation of an ontology to the arisen changes and the consistent management of these changes*”. É um processo que suporta o enriquecimento da ontologia por adicionar novos elementos ou por modificar elementos existentes quando um novo conhecimento é adquirido. Uma ontologia evolui e não pode ser considerada um modelo estático. Alguns elementos podem ser adicionados ou modificados para adaptação às mudanças dinâmicas do ambiente (conhecimento). Outros elementos tornam-se obsoletos ao longo do tempo. O desenvolvimento de uma ontologia inicia-se com uma versão inicial a qual posteriormente deve ser revisada, refinada com novos detalhes (NOY; MCGUINNESS, 2001). Além disso, novas informações que não foram inicialmente previstas precisam ser adicionadas na ontologia. O método desenvolvido pode ser uma forma semiautomática de evolução de ontologias a partir da evolução do conhecimento em folksonomias. As folksonomias possuem uma natureza dinâmica, pois evoluem com o tempo a medida que o usuário acrescenta novas anotações (HAMMOND et al., 2005). O aprendizado de ontologias a partir de folksonomias pode suportar o refinamento ou expansão de ontologias existentes por incorporar novos conhecimentos.

Outro problema relatado na evolução de ontologias é a existência temporal do conhecimento. Com a aquisição de novas informações, novas partes de conhecimento precisam ser incluídas nas ontologias e partes antigas precisam ser verificadas e validadas. Todavia, estas alterações são complexas e precisam ser capturadas pelos engenheiros que as criaram. O atraso nas alterações pode ser inaceitável em aplicações dinâmicas onde o conhecimento é alterado regularmente e dinamicamente. Para ser eficaz, os métodos de evolução precisam alterar as ontologias precisam ser eficazes nas alterações (HAASE; VÖLKER; SURE, 2005; HAASE; SURE, 2004).

Juntamente com a evolução, outra questão é o controle de versões e rastreabilidade. O desenvolvimento de ontologias é um processo longo, pois o conhecimento é aprendido de uma forma incremental e as pessoas devem comprometer-se com uma certa visão do domínio. Pessoas diferentes têm opiniões diferentes sobre a mesma parte do domínio a ser modelado e os conflitos resultantes devem ser gerenciados. Conseqüentemente, há uma grande diferença das versões iniciais da ontologia para as versões finais. Como o método desenvolve várias

ontologias a partir de novas anotações, pode-se manter um controle de versões verificando a evolução desde a primeira ontologia até as versões finais. As versões anteriores podem ser mantidas para registrar alterações na ontologia ao longo do tempo, ligando-se às anotações das folksonomias antigas e atuais para permitir a rastreabilidade, isto é, conhecer a anotação que deu origem a um elemento da ontologia.

Melhorias também podem ser adicionadas ao método desenvolvido. Alguns aspectos que podem ser explorados:

- Explorar outras técnicas de extração de conhecimentos a partir de folksonomias;
- Incluir outras atividades de desenvolvimento como formalização implementação (integração do método em uma metodologia de desenvolvimento);
- Incluir outras técnicas de processamento de linguagem natural;
- Incluir alguma abordagem de aprendizado de máquina como redes neurais, ou criação de grupos.



## REFERÊNCIAS

ALANI, Harith.; KIM, Sanghee.; MILLARD, David.; WEAL, Mark J.; HALL, Wendy.; LEWIS, Paul H.; SHADBOLT, Nigel R. Automatic ontology-based knowledge extraction from web documents. **IEEE Intelligent Systems**, v.18, n. 1, p. 14–21, jan. 2003.

ALEXIEV, Vladimir.; BREU, Michael.; DE BRUIJN, Jos.; FENSEL, Dieter.; LARA, Ruben.; LAUSEN, Holger. **Information Integration with Ontologies: Experiences from an Industrial Showcase**. Wiley, 2005.

ANGELETOU, Sofia.; SABOU, Marta.; SPECIA, Lucia.; MOTTA, Enrico. Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report. **In: Workshop Bridging the Gap between Semantic Web and Web 2.0**, 4th European Semantic Web Conference, p. 30-43, Innsbruck, Austria, 2007.

ARPINAR, I.B., Giriloganathan, K., and Aleman-Meza, B. Ontology Quality by Detection of Conflicts in Metadata. **In Proceedings of the 4th International EON Workshop**. May 22nd, 2006.

ASSADI H. Construction of a regional ontology from text and its use within a documentary system. **In: Proceedings of the International Conference on Formal Ontology and Information Systems (FOIS-98)**. Trento, Itália, Amsterdam, 1998.

AUSSENAC-GILLES, Nathalie.; BIÉBOW, Brigitte.; SZULMAN, Sylvie. Revisiting Ontology Design: A Methodology Based on Corpus Analysis. **In: Dieng R, Corby O (eds) 12th International Conference in Knowledge Engineering and Knowledge Management (EKAW'00)**. Juan-Les-Pins, France, 2000.

BAADER, Franz.; CALVANESE, Diego.; MCGUINNESS, Deborah L.; NARDI, Daniele.; PATEL-SCHNEIDER, Peter. **The Description Logic Handbook**. Cambridge University Press, 2003.

BALDONI, Matteo.; BAROGLIO, Cristina.; HORVÁTH, András.; PATTI, Viviana.; PORTIS, Flavio.; AVILIA, Maurizio.; GRILLO, Pierluigi. Folksonomies meet ontologies in ARSMETEO: From social descriptions of artifacts to emotional concepts. **In: Proc. of Formal Ontologies Meet Industry (FOMI)**. Torino, Italy, jun. 2008.

BANERJEE, Stanjev; PEDERSEN, Ted. An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet. **In: Proceedings of the Third international Conference on Computational Linguistics and intelligent Text Processing**. A. F. Gelbukh, Ed. Lecture Notes In Computer Science, v. 2276, p. 136-145, Springer-Verlag, London, 2002.

BATEMAN, John A.; FABRIS, Giovanni.; MAGNINI, Bernardo. The Generalized Upper Model Knowledge Base: Organization and Use. **In: Second International Conference on Building and Sharing of Very Large-Scale Knowledge Bases (KBKS '95)**. University of Twente, Enschede, Netherlands. IOS Press, Amsterdam. p. 60–72, 1995.

BEGELMAN, Grigory.; KELLER, Philipp; SMADJA, Frank. Automated *Tag* Clustering: Improving search and exploration in the *tag* space. **In: Collaborative Web Tagging Workshop at WWW2006**. Edinburgh, Scotland, 2006.

BERNERS-LEE, Tim.; HENDLER, James.; LASSILA, Ora. **The Semantic Web. Scientific American Magazine**. 2001. Disponível em: <<http://www.sciam.com/article.cfm?id=the-semantic-web&print=true>> Acesso em: jun. 2010.

BLOMQVIST. E.. Fully automatic construction of enterprise ontologies using design patterns: Initial method and first experiences. **In Proc of ODBASE'05**, 2005.

BOTTAZZI, E.; FERRARIO, R. **Preliminaries to a DOLCE Ontology of Organizations**, International Journal of Business Process Integration and Management, 2006.

BREWSTER, Christopher.; ALANI, Harith.; DASMAHAPATRA, Srinandan.; WILKS, Yorick. Data-driven ontology evaluation. **In: Proceedings of the 4th International Conference on Language Resources and Evaluation**. Lisbon, 2004.

BUITELAAR, Paul.; CIMIANO, Philipp., MAGNINI, Bernardo. Ontology learning from text: An overview. **Artificial Intelligence and Applications Series**, p. 3-12, 2005.

BURTON-JONES, Andrew.; STOREY, Veda C.; SUGUMARAN, Vijayan.; AHLUWALIA, Punit. A semiotic metrics suite for assessing the quality of ontologies. **Data Knowledge Engineering**, v. 5, p. 84-102, 2005.

CIMIANO, P.; VOELKER, J.: Text2Onto - A Framework for Ontology Learning and Data-driven Change Discovery. **In: Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB)**, Alicante, Spain (2005)

CIMIANO, Philipp. Ontology Learning and Population from Text: Algorithms, Evaluation and Applications. **Springer-Verlag**. New York, 2006.

CIMIANO, Philipp.; WENDEROTH, Johanna. Automatically Learning Qualia Structure from the Web. **In: Proceedings of the ACL Workshop on Deep Lexical Acquisition**. ACL Workshops. Association for Computational Linguistics, Morristown, p. 28-37, jun. 2005.

CIMIANO, Philipp.; HANDSCHUH, Siegfried.; STAAB, Steffen. (2004). Towards the self-annotating Web. **In: Proceedings of the 13<sup>th</sup> International Conference on World Wide Web**. New York. p. 462-471, 2004.

CIMIANO, Philipp.; STAAB, Steffen.; TANE, Julien. Automatic acquisition of taxonomies from text: FCA meets NLP. **In: Proceedings of the PKDD/ECML'03 International Workshop on Adaptive Text Extraction and Mining**. 2003.

CLARK, P.; THOMPSON, J.; PORTER, B. Knowledge patterns. **In Proceedings of 7th International Conference Principles of Knowledge Representation and Reasoning (KR)**, Breckenridge, Colorado, USA: 591-600, 2000.

DAMME, Celine. V.; HEPP, Marting.; SIORPAES, Katharina. **FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies**. 2007. Disponível em <<http://www.kde.cs.uni-kassel.de/ws/eswc2007/proc/FolksOntology.pdf>>. Acesso em: nov. 2007.

DEBENHAM, J. K. Constructing the Functional Model. **In: Proceedings of the 8th international Conference on Database and Expert Systems Applications**. A. Hameurlain and A. M. Tjoa, Eds. Lecture Notes In Computer Science, v. 1308, p. 48-57. Springer-Verlag, London, 1997.

DELLSCHAFT, Klass.; STAAB, Steffen. On how to perform a gold standard based evaluation of ontology learning. **In: Internation Semantic Web Conference**, v. 4273, p. 228-241, Lecture Notes in Computer Science. 2006.

DE NICOLA, A.; MISSIKOF, M., NAVIGLI, R.: A software engineering approach to ontology building. **Inf. Syst.** 34(2) (2009) 258–275

DIETTERICH, Thomas G. Machine learning research: Four current directions. **AI Magazine**, v. 18, n. 4, p. 97–136, 1997.

DOTSIKA, Fefie. Uniting formal and informal descriptive power: Reconciling ontologies with folksonomies. **International Journal of Information Management**, v. 9, n. 5, p. 407-415, out. 2009.

ECHARTE, Francisco.; ASTRAIN, José.; CÓRDOBA, Alberto.; VILLADANGOS, Jesús. Ontology of Folksonomy: A New Modeling Method. **In: Proc. of Semantic Authoring, Annotation and Knowledge Markup**, 2007

EHRIG, Marc. **Ontology Alignment: Bridging the Semantic Gap (Semantic Web and Beyond)**. Springer Verlag, 2006.

EHRIG, Marc.; SURE, York. Ontology Mapping: an Integrated Approach. **In: Proceedings of the First European Semantic Web Symposium**. Lecture Notes in Computer Science, v. 3053, p. 76-91, Heraklion, Greece, Springer Verlag, may 2004.

FALBO, R. A.; GUIZZARDI, G.; DUARTE, K. C. An Ontological Approach to Domain Engineering. **In Proceedings of the 14th International Conference on Software Engineering and Knowledge Engineering (SEKE)**, Ischia, Italy, 2002.

FALBO, R. A.; MENEZES, C. S.; ROCHA, A. R. C. .A Systematic Approach for Building Ontologies., **In Proceedings of the Sixth Iberoamerican Conference on Artificial Intelligence (IBERAMIA)**, Lisboa, Portugal, 1998.

FAURE, David.; NEDELLEC, Claire. A corpus-based conceptual clustering method for verb frames and ontology acquisition. **In: LREC workshop on adapting lexical and corpus resources to sublanguages and applications**. Granada, Spain, 1998.

FELLBAUM, Christiane. **WordNet: An Electronic Lexical Database**. Cambridge: The MIT Press, 1998.

FERNÁNDEZ-LÓPEZ, Mariano.; GÓMEZ-PÉREZ, Asunción.; SIERRA, Juan P. .; SIERRA, Alejandro P. Building a Chemical Ontology Using Methontology and the Ontology Design Environment. **IEEE Intelligent Systems & their applications**, v. 4, n. 1, p. 37–46, jan. 1999.

FERNÁNDEZ-LÓPEZ, Mariano.; GÓMEZ-PÉREZ, Asunción.; JURISTO, Natalia. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. **In: Spring Symposium on Ontological Engineering of AAAI**. Stanford University, California, 1997.

FININ, Tim., FRITZSON, Richard., MCKAY, D., and MCENTIRE, R. KQML as an agent communication language. **In: Proceedings of the Third international Conference on information and Knowledge Management** (Gaithersburg, Maryland, United States, November 29 - December 02, 1994). N. R. Adam, B. K. Bhargava, and Y. Yesha, Eds. CIKM '94. ACM, New York, NY, p. 456-463, 1994.

FREDDO, Ademir R.; TACLA, Cesar A.; BRITO, Robison.; GIMENEZ-LUGO, Gustavo. Partial and Dynamic Ontology Mapping Model in Dialogs of Agents. **EPIA workshops. Artificial Intelligence Magazine**, v. 26, n. 1, p. 71–82, 2006.

FREDDO, Ademir R.; TACLA, Cesar A. Evaluation of a method for partial ontology alignment in multi-agent system. **International Journal of Reasoning-based Intelligent Systems**, v.1, n. 3/4, p. 132-146, 2009.

FREDDO, Ademir. R.; YAMAMOTO, Anderson.; PINTO, Sofia.; TACLA, Cesar A. Arquitetura de um serviço de registro baseado em alinhamento de ontologias para interoperabilidade dinâmica de processos. **In: Proceedings of the 14th Brazilian Symposium on Multimedia and the Web** (Vila Velha, Brazil, October 26 - 29, 2008). WebMedia '08. ACM, New York, NY, p. 234-241, 2008.

FREDDO, Ademir R.; TACLA, Cesar A. Integrating Social Web with Semantic Web - Ontology Learning and Ontology Evolution from Folksonomies. **In: Proceedings of the International Conference on Knowledge Engineering and Ontology Development (KEOD)**. Funchal - Madeira, Portugal, October 6-8, 2009. INSTICC Press, 2009.

FREDDO, Ademir R.; SATO Gilson.; TACLA, Cesar A.; PARAISO Emerson.; CAMPAGNOLO, Bruno.; RAMOS Milton. Uma Arquitetura para o Suporte ao Desenvolvimento Colaborativo de Software em Pequenas Equipes. **In: VI Simposio Brasileiro de Sistemas Colaborativos**. Fortaleza, CE, VI Simposio Brasileiro de Sistemas Colaborativo, 2009.

FREDDO, Ademir R.; TACLA, Cesar A. Semantic Interoperability for Registry Services Using Ontology Alignment. Hamid R. Arabnia, Andy Marsh (Eds.): **Proceedings of the 2010 International Conference on Semantic Web & Web Services**, SWWS 2010, July 12-165 2010, Las Vegas, Nevada, USA.

FREITAS, Fredi. L. G. de. Ontologias e a Web Semântica. In: SBC. III MCIA-XXIII Congresso da SBC. **Anais do XXIII Congresso da SBC**. Campinas, SP, Brasil, 2003.

GAMBA R. **Historia sencilla de la filosofía**, 24th edn. Rialp, Madrid, Spain, 1999.

GANGEMI, A.; C. CATENACCI, C.; M. CIARAMITA, M.; J. LEHMANN, J. Modelling Ontology evaluation and validation. **In proceedings of the 3rd European Semantic Web Conference**, June, 2006.

GANGEMI, A.: Ontology design patterns for semantic web content. **In: M. Musen et al. (eds.): Proceedings of the Fourth International Semantic Web Conference**, Berlin, Springer, 2005.

GANGEMI, Aldo.; GUARINO, Nicola.; OLTRAMARI, Alessandro.; BORGIO, Stefano. Cleaning-up WordNet's top-level. **In: Proceedings of the 1st International WordNet Conference**. Mysore, India, 2002.

GOLDER, Scott A.; HUBERMAN, Bernardo A. **The structure of collaborative tagging systems**. Information Dynamics Lab: HP Labs, Palo Alto, 2005. Disponível em: <http://arxiv.org/abs/cs.DL/0508082>. Acesso em fev/2010.

GÓMEZ-PÉREZ, Asunción.; JURISTO, Natalia.; PAZOS, Juan. Evaluation and assessment of the knowledge sharing technology. **In: Mars N (ed) Towards very large knowledge bases**. IOS Press, Amsterdam, p. 289–296, 1995.

GÓMEZ-PÉREZ, Asunción. Some ideas and examples to evaluate ontologies. **In: Proceedings of the 11th Conference on Artificial Intelligence For Applications**. CAIA. IEEE Computer Society, Washington, DC, 299, fev. 1995.

GÓMEZ-PÉREZ, Asunción. Evaluation of Taxonomic Knowledge on Ontologies and Knowledge-Based Systems. **Int. Workshop on Knowledge Acquisition, Modeling and Management**, 1999.

GÓMEZ-PÉREZ, Asunción.; ROJAS-AMAYA, D. Ontological reengineering for reuse. **In: Proceedings of the 11th European Knowledge Acquisition Workshop EKAW**, 1999, pp. 139-157.

GÓMEZ-PÉREZ, Asunción.; MANZANO-MACHO, David. **OntoWeb Deliverable 1.5: A Survey of Ontology Learning Methods and Techniques**. Universidad Politecnica de Madrid, Spain, 2003.

GÓMEZ-PÉREZ, Asunción.; CORCHO, Oscar.; FERNÁNDEZ-LÓPEZ. **Ontological Engineering : with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web**. First Edition (Advanced Information and Knowledge Processing). Springer, 2004.

GROSSO, William E.; ERIKSSON, Henrik.; FERGERSON, Ray W.; GENNARI, John H.; TU, Samson W.; MUSEN, Mark A.. Knowledge modeling at the millennium — the design and evolution of Protégé-2000. **In: Proceedings of KAW-99**. Banff, Canada, 1999.

GRUBER, Thomas. **Folksonomy of Ontology: A Mash-up of Apples and Oranges**. First on-Line conference on Metadata and Semantics Research (MSTR'05), 2005.

GRUBER, Thomas. Collective knowledge systems: **Where the Social Web meets the Semantic Web, Web Semantics: Science, Services and Agents on the World Wide Web**, v. 6, n.1, p. 4-13, 2008

GRUBER Thomas. Toward principles for the design of ontologies used for knowledge sharing. In: Guarino N, Poli R (eds) **International Workshop on Formal Ontology in Conceptual Analysis and Knowledge Representation**. Padova, Italy. (Formal Ontology in Conceptual Analysis and Knowledge Representation) Kluwer Academic Publishers, Deventer, The Netherlands, 1993.

GRÜNINGER, Michael.; FOX Mark S. Methodology for the design and evaluation of ontologies **In: IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing**. p 6.1–6.10, april. 1995.

GUARINO, Nicola.; CARRARA, Massimiliano.; GIARETTA, Pierdaniela. 1994. An Ontology of Meta-Level Categories. **In: D. J., E. Sandewall and P. Torasso (eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference (KR94)**. Morgan Kaufmann, San Mateo, p. 270-280, 1994.

GUARINO, Nicola.; GIARETTA, Pierdaniela. Ontologies and Knowledge Bases: Towards a Terminological Clarification. **Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing**, p. 25-32, 1995.

GUARINO, N. Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. **In: M. Pazienza, Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology**, International Summer School, SCIE-97, Frascati, Italy, p. 139-170, 1997.

GUARINO, Nicola. Formal Ontology in Information Systems. **In: Guarino N (ed) 1<sup>st</sup> International Conference on Formal Ontology in Information Systems (FOIS'98)**. Trento, Italy. IOS Press, Amsterdam, p. 3–15, 1998.

GUARINO, Nicola.; MASOLO, Claudio.; VETERE, Guido. Ontoseek: Content-based access to the web. **IEEE Intelligent Systems**, v. 14, n. 3, p. 70-80, 1999.

GUARINO, Nicola.; WELTY, Christopher A. **Evaluating Ontological Decisions with OntoClean**. Communications of The ACM. v. 45, n. 2, p. 61-65, feb. 2002.

GUARINO, Nicola.; WELTY, Christopher A. An Overview of OntoClean. **In: Steffen Staab and Rudi Studer, eds., The Handbook on Ontologies**. Chapter 8, p. 151-172. Berlin:Springer-Verlag, 2004

GUIZZARDI, G. **Ontological Foundations for Structural Conceptual Models**. Telematica Instituut Fundamental Research Series no. 15, Universal Press, The Netherlands, 2005, ISBN 90-75176-81-3.

GUIZZARDI, G.; LOPES, M.; BAIÃO, F. A.; FALBO, R. A.: On the Importance of Truly Ontological Distinctions for Ontology Representation Languages: **An Industrial Case Study in the Domain of Oil and Gas**. BMMDS/EMMSAD 2009: 224-236

- GUIZZARDI, G., FALBO, R.A., GUIZZARDI, R.S.S. **A importância de Ontologias de Fundamentação para a Engenharia de Ontologias de Domínio: o caso do domínio de Processos de Software**, Revista IEEE América Latina, v. 6, p. 244-251, 2008
- HAASE P., van Harmelen F., Huang Z., Stuckenschmidt H., and Sure Y. A framework for handling inconsistency in changing ontologies. In Proceedings of ISWC2005, 2005.
- HAASE, Peter.; VÖLKER, Johanna. Ontology learning and reasoning- dealing with uncertainty and inconsistency. **In: Proceedings of the Workshop on Uncertainty Reasoning of the Semantic Web (URSW)**, p 45-55, 2005.
- HAASE, Peter.; VÖLKER, Johanna.; SURE, York. Management of dynamic knowledge. **Journal of Knowledge Management**, v. 9, n. 5, p. 97-107, 2005.
- HAASE, Peter.; SURE, York. **D3.1.1.b State of the Art on Ontology Evolution**. 2004. Disponível em: <http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/SEKT-D3.1.1.b.pdf>. Acesso em fev. 2008.
- HAASE, Peter.; STOJANOVIC, Ljiljana. Consistent Evolution of OWL Ontologies. **In: Proceedings of the 2nd European Semantic Web Conference**. v. 3532, of Lecture Notes in Computer Science, p. 182-197, Heraklion, Greece, 2005.
- HAMMOND, Tony.; HANNAY, Timo.; LUND, Ben.; SCOTT, Joanna. Social Bookmarking Tools (I): A General Review. **D-Lib Magazine**, n. 11, v. 4, 2005.
- HARTMANN, P. Spyns, A. Giboin, D. Maynard, R. Cuel, M. Suarez-Figueroa, and Y. Sure, “**D1.2.3 Methods for Ontology Evaluation**”, EU-IST Network of Excellence (NoE) IST-2004-507482, Knowledge Web Consortium, Jan. 2005.
- HEARST, Marti. A. Automatic acquisition of hyponyms from large text corpora. **In: Proceedings of the 14th Conference on Computational Linguistics**. (Nantes, France, August 23 - 28, 1992). International Conference On Computational Linguistics. Association for Computational Linguistics, Morristown, NJ, v.2, p. 539-545, 1992.
- HOLSAPPLE, Clyde W.; JOSHI, K. D. A collaborative approach to ontology design. *Commun. ACM*, v. 45, n. 2, p. 42-47, 2002.
- HOTHO, Andreas.; MAEDCHE, Alexander.; STAAB, Steffen. Ontology- based text clustering. **In: Proceedings of the IJCAI-2001 Workshop Text Learning: Beyond Supervision**, Seattle, WA: Springer, 2001.
- HOTHO, Andreas.; STAAB, Steffen. Learning concept hierarchies from text corpora using formal concept analysis. **Jornal of Artificial Intelligence Researc.**, v. 24, p. 305–339, 2005.
- HOTHO, Andreas.; JÄSCHKE, Robert.; SCHMITZ, Christoph.; STUMME, Gerd. Information retrieval in folksonomies: Search and ranking. **In: Sure, Y. and Domingue, J., editors, The Semantic Web: Research and Applications**, v. 4011 of Lecture Notes in Computer Science, chapter 31, p. 411-426. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- IEEE (1990) *IEEE Standard Glossary of Software Engineering Terminology*. IEEE

Computer Society. New York. IEEE Std 610.121990

JÄSCHKE, Roberto.; HOTHO, Andreas.; SCHMITZ Christoph.; GANTER, Bernhard.; STUMME Gerd. Discovering shared conceptualizations in folksonomies. **In: Web Semantics: Science, Services and Agents on the World Wide Web**, v. 6, n. 1, Semantic Web and Web 2.0, p. 38-53, 2008.

KAVALEC, Martin.; MAEDCHE, Alexander.; SVÁTEK, Vojtech. Discovery of lexical entries for non-taxonomic relations in ontology learning. **In: Proc. Of SOFSEM 2004**. Lecture Notes in Computer Science, v. 2932, p. 249-256, 2004.

KENDAL, Simon.; CREEN, Malcolm. **An Introduction to Knowledge Engineering**. Springer-Verlag London Limited, 2007.

KHAN, Latifur.; LUO, Feng. Ontology construction for information selection. **In: Proceedings of the 14th IEEE international conference on tools with artificial intelligence**. Washington, DC p. 122-127, 2002.

KIETZ, Joerg-Uwe.; MAEDCHE, Alexander.; VOLZ, Raphael. A Method for Semi-Automatic Ontology Acquisition from a Corporate Intranet. **In: Aussenac-Gilles N, Biébow B, Szulman S (eds) EKAW'00 Workshop on Ontologies and Texts**. Juan-Les-Pins, France. CEUR Workshop Proceedings 51:4.1-4.14. Amsterdam, The Netherlands, 2000.

KIKAS, T. e TREUMUTH, M. **Word Sense Disambiguation: WordNet SenseRelate AllWords**. 2007. Disponível em: <<http://math.ut.ee/~treumuth/NLP/semantics2.pdf>>. Acesso em: março, 2009.

KIM, Hak-Lae.; YANG, Sung-Kwon.; BRESLIN, John. G.; KIM Hong-Gee. Simple algorithms for representing *tag* frequencies in the scot exporter. **In: Proceedings of the 2007 IEEE/WIC/ACM international Conference on intelligent Agent Technology**. IAT. IEEE Computer Society, Washington, DC, 2007.

KLEIN, Michel. Combining and Relating Ontologies: An Analysis of Problems and Solutions. **In: Workshop on Ontologies and Information Sharing at the 17th International Joint Conference on Artificial Intelligence**. Seattle, USA, p. 53-62, 2001.

KNERR, Torben. **Tagging Ontology- Towards a Common Ontology for Folksonomies**. 2006. Disponível em: <<http://tagont.googlecode.com/files/TagOntPaper.pdf>>. Acesso em maio, 2009.

KROSKI, Ellyssa. **The Hive Mind: Folksonomies and User-Based Tagging**. 2005. Disponível em: <<http://infotangle.blogspot.com/category/folksonomies>>. Acesso em dezembro, 2009.

LANIADO, David.; EYNARD, Davide.; COLOMBETTI, Marco. A Semantic Tool to Support Navigation in a Folksonomy. **In: Proceedings of the Eighteenth Conference on Hypertext and Hypermedia**. ACM, New York, NY, p. 153-154, 2007.



LENAT, Dougla B.; GUHA R.V. **Building Large Knowledge-based Systems: Representation and Inference in the CYC Project**. Addison-Wesley, Boston, Massachusetts, 1990.

LIMPENZ, Freddy.; GANDON, Fabien.; BUFFA, Michel. Bridging Ontologies and Folksonomies to Leverage Knowledge Sharing on the Social Web: a Brief Survey. **Proc. 1st International Workshop on Social Software Engineering and Applications (SoSEA)**, 2008.

LIN, Dekang.; PANTEL, Patrick. Discovery of inference rules from text. **In: Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining**, p. 323-328, 2001.

LONSDALE, D.; EMBLEY, W.; DING, Y.; XU, L.; HEPP, M. Reusing ontologies and language components for ontology generation. **Data & Knowledge Engineering**, 69:318–330, 2010.

LOWE, E. J. 1989. **Kinds of Being. A Study of Individuation, Identity and the Logic of Sortal Terms**. Basil Blackwell, Oxford, 1989.

LUX, Mathias.; DOSINGER, Gisela. From folksonomies to ontologies: employing wisdom of the crowds to serve learning purposes. **International Journal of Learning Technology**, v. 3 n. 4/5, p. 515–528, 2007.

MAEDCHE, Alexander.; STAAB, Steffen. Semi-automatic engineering of ontologies from text. **In: Proceedings of the 12th international conference on software engineering and knowledge engineering**. 2000a.

MAEDCHE, Alexander.; STAAB, Steffen. Discovering conceptual relations from text. **In: Proc. of the 14th European Conference on Artificial Intelligence**. IOS Press, Amsterdam, p. 321-325, 2000b.

MAEDCHE, Alexander.; STAAB, Steffen. Ontology Learning for the Semantic Web. **IEEE Intelligent Systems. Special Issue on the Semantic Web**, v. 16, n. 2, p. 72-79, 2001.

MAEDCHE, Alexander. **Ontology learning for the semantic Web**. Norwell, MA, USA: Kluwer Academic Publishers, 2002.

MAEDCHE, Alexander.; STAAB, Steffen. Measuring similarity between ontologies. **In: Proceedings of the 13th international Conference on Knowledge Engineering and Knowledge Management. ontologies and the Semantic Web**. A. Gómez-Pérez and V. R. Benjamins, Eds. Lecture Notes In Computer Science, v. 2473, p. 251-263. Springer-Verlag, London, 2002.

MAEDCHE, Alexander.; STAAB, Steffen. **Ontology learning**. In Staab, S., and Studer, R., eds., Handbook on Ontologies. Springer. p. 173–189, 2004.

MAHESH, Kavi. **Ontology development for machine translation: Ideology and Methodology**. Technical Report MCCS-96-292. Computing Research Laboratory, New Mexico State University, Las Cruces, New Mexico, 1996

MAHESH, Kavi.; NIRENBURG, Sergei. Semantic classification for practical natural language processing. In: **Schwartz RP, Kwasnik BH, Beghtol C, Smith PJ, Jacob E (eds) 6th ASIS SIG/CR Classification Research Workshop: An Interdisciplinary Meeting**. Chicago, Illinois, p 79–94, 1995.

MANNING, Christopher D.; SCHÜTZE, Hinrich. **Foundations of Statistical Natural Language Processing**. Cambridge: The MIT Press, 1999.

MAASS, W.; JANZEN, S.: A Pattern-based Ontology Building Method for Ambient Environments Workshop on Ontology Patterns - **WOP2009 at the 8th International Semantic Web Conference (ISWC 2009)**, Washington, DC, 2009.

MATHES, Adam. **Folksonomies - Cooperative Classification and Communication Through Shared Metadata**. 2004. Disponível em <<http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>>. Acesso em outubro, 2009.

MAYNARD, D.; FUNK, A.; PETERS, W. Using Lexico-Syntactic Ontology Design Patterns for ontology creation and population. In **Proceedings of WOP2009 collocated with ISWC2009**, CEUR-WS.org, 2009.

MCGUINNESS, Deborah. L. **Ontologies Come of Age**. 2002 Disponível em: [http://www-ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-\(with-citation\).htm](http://www-ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm). Acesso em: Outubro 2006.

MERRIAM-WEBSTER **Dictionary**, [[online:www.m-w.com](http://www.m-w.com)], captured in 2004.

MILLER, George A.; BECKWITH, Richard.; FELBAUM, Christiane.; GROSS, Derek.; MILLER, Katherine. Introduction to WordNet: An on-line lexical database. **International Journal of Lexicography**, v. 3, n. 4, p. 235–244, 1990.

MILLER, George A. (1995) WordNet: a lexical database for English. **Communications of the ACM**, v. 38, n. 11, p. 39–41, 1995.

MIKA, Peter. Ontologies are us: A unified model of social networks and semantics. In: **Proceedings of the 4th International Semantic Web Conference (ISWC)**. LNCS 3729, p. 522-536, Galway, Ireland, 2005.

MILTON, Nicholas.R. **Knowledge Acquisition in Practice**. Springer-Verlag London Limited, 2007.

MINSKY, Marvin A. Framework for Representing Knowledge. In: **The Psychology of Computer Vision**, p. 211-281, McGraw-Hill, New York. USA, 1975.

MISSIKOFF, Michele.; NAVIGLI, Roberto.; VELARDI, Paola. Integrated approach for web ontology learning and engineering. **IEEE Computer**, v. 35, n. 11, p. 60–63, 2002.

MORIN, Emmanuel. Automatic acquisition of semantic relations between terms from technical corpora. **In: Proc. Of the Fifth Int. Congress on Terminology and Knowledge Engineering (TKE-99)**. Vienna. TermNet-Verlag, 1999.

NAVARRO, Gonzalo. 2001. A guided tour to approximate string matching. **ACM Comput. Surv**, v. 33, n. 1, p. 31-88, 2001.

NAVIGLI, Roberto.; VELARDI, Paola.; GANGEMI, Aldo. Ontology learning and its application to automated terminology translation. **IEEE Intelligent Systems**, v. 18, n. 1, p. 22–31, 2003.

NECHES, Robert.; FIKES, Richard.; FINIM, Tim.; GRUBER Tom.; PATIL, Ramesh.; SENATOR. Ted.; SWARTOUT, William. Enabling technology for knowledge sharing. **AI Magazine**, v. 12, n. 3, p. 36–56, 1991.

NEWMAN, Richard. *Tag Ontology Writeup*. 2005. Disponível em <<http://www.holygoat.co.uk/projects/tags/>>. Acesso em novembro, 2008.

NILES, I., PEASE, A. Towards a standard upper ontology. **In Proceedings of the international Conference on Formal ontology in information Systems (FOIS '01)**, Ogunquit, Maine, USA, 2001.

NIRENBURG, Sergei.; RASKIN, Victor. **Ontological Semantics**. MIT Press, Cambridge, Mass, 2004.

NOY, N., “Evaluation by Ontology Consumers”. *IEEE Intelligent Systems* 1541-1672, 2004.

NOY Natalya F.; MUSEN, Mark A. Anchor-PROMPT: Using Non-Local Context for Semantic Matching. **In: Gómez-Pérez A, Grüninger M, Stuckenschmidt H, Uschold M (eds) IJCAI'01 Workshop on Ontologies and Information Sharing**. Seattle, Washington, p. 63–70, 2001.

NOY, Natalya F.; MCGUINNESS, Deborah. **Ontology development 101: A guide to creating your first ontology**. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, 2001.

NOY, Natalya F.; SINTEK, Michael.; DECKER, Stefan.; CRUBEZY, Monica.; FERGERSON, Ray.; MUSEN, Mark A. Creating semantic web contents with Protege-2000. **IEEE Intelligent Systems**, v. 16, n. 2, p. 60–71, 2001.

OGDEN, C. K.; RICHARDS, Ivor. A. **The Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism**. London: Routledge & Kegan Paul, 1923.

PAOLUCCI, Massimo.; KAWAMURA, Takahiro.; PAYNE, Terry. R.; SYCARA, Katia. P. Semantic Matching of Web Services Capabilities. **In: Horrocks, I. and Hendler, J., editors, The Semantic Web — ISWC 2002**. Lecture Notes in Computer Science, chapter 26, v. 2342, p. 333-347, Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.

PETERSON, Elaine. Beneath the Metadata: Some Philosophical Problems with Folksonomy. **D-Lib Magazine**, v. 12, n. 11, nov. 2006.

PINTO, Helena S.; GÓMEZ-PÉREZ, Asunción.; MARTINS, João P. Some issues on ontology integration. **In: Benjamins VR (ed) IJCAI'99 Workshop on Ontology and Problem Solving Methods: Lessons Learned and Future Trends**. Stockholm, Sweden. CEUR Workshop Proceedings 18:7.1–7.12. Amsterdam, The Netherlands, 1999.

PINTO, Helena. S.; MARTINS, João. P. Ontologies: How can They be Built?. **Knowledge Information Systems**, v. 6, n. 4, p. 441-464, 2004.

PINTO, Helena. S.; TEMPICH, Christoph.; SURE, York. Diligent: Towards a fine-grained methodology for distributed. **In: Proceedings of the European Conference on Artificial Intelligence (ECAI)**, p. 393-397, 2004.

PLESSERS P. and De Troyer O. Ontology Change Detection Using a Version Log. **In Proceedings of the 4th International Semantic Web Conference (ISWC-05)**, 2005.

PLISSON, J.; LAVRAC, N.; MLADENIC, D. **A Rule based Approach to Word Lemmatization**. In: Proceedings of IS04, 2004.

POESIO, Massimo.; ALMUHAREB, Abdulrahman. Identifying concept attributes using a classifier. **In: Proceedings of the ACL-SIGLEX Workshop on Deep Lexical Acquisition**. ACL Workshops. Association for Computational Linguistics, Morristown, NJ, p. 18-27, 2005.

PORZEL, Robert.; MALAKA, Rainer. A task-based approach for ontology evaluation. **In: ECAI 2004 Workshop Ontology Learning and Population**. 2004.

PRESUTTI V.; GANGEMI A. Content Ontology Design Patterns as practical building blocks for web ontologies. **In Proceedings of ER2008**, Barcelona, Spain, 2008.

PRESUTTI V.; GANGEMI A.; DAVID S.; AGUADO, G.; SUAREZ-FIGUEROA, M.C.; MONTIEL E.; POVEDA M. Neon deliverable D2.5.1. A Library of Ontology Design Patterns. Disponível em <<http://www.neon-project.org>>.

QUINE, Willard. V. O. **Ontological Relativity and Other Essays**. Columbia University Press, New York, London, 1969.

QUINTARELLI, Emanuele. **Folksonomies: power to the people**. ISKO Italy-UniMIB meeting, Milan, Italy, 2005. Disponível em: <<http://www.iskoi.org/doc/folksonomies.htm>>. Acesso em outubro, 2009.

RECTOR, Alan.; DRUMMOND, Nick.; HORRIDGE, Matthew.; ROGERS, Jeremy.; KNUBLAUCH, Holger.; STEVENS, Robert.; WANG, Hai.; WROE, Chris. OWL Pizzas: Practical Experience of Teaching OWL-DL: Common Errors & Common Patterns. **In: 14th International Conference on Knowledge Engineering and Knowledge Management - EKAW 2004**. Whittlebury Hall, Northamptonshire, UK, 2004.

RECH, J.; FELDMANN R.L.; RAS, E. Knowledge Patterns. In M. E. Jennex (Ed.), *Encyclopedia of Knowledge Management (2nd Edition)*, IGI Global, USA, (2009).

ROLSTON, David.W. **Principles of Artificial Intelligence and Expert Systems Development**. McGraw-Hill Book Co, 1988.

RUIZ-CASADO, Maria.; ALFONSECA, Enrique.; CASTELLS, Pablo. Automatising the learning of lexical patterns: An application to the enrichment of WordNet by extracting semantic relationships from Wikipedia. **Data & Knowledge Engineering**, v. 61, n. 3, p. 484-499, jun. 2007.

RUNG-CHING, Chen.; JUI-YUAN, Liang.; REN-HAO, Pan. Using recursive ART network to construction domain ontology based on term frequency and inverse document frequency. **Expert Systems with Applications**, v. 34, n. 1, p. 488-501, jan. 2008.

RUI-LING, Zhang.; HONG-SHENG, Xu. 2009. Using Bayesian Network and Neural Network Constructing Domain Ontology. **In: Proceedings of the WRI World Congress on Computer Science and information Engineering - CSIE**. IEEE Computer Society, Washington, DC, v. 6, p. 116-120, 2009.

SABOU, Marta.; D'AQUIN, Mathieu.; MOTTA, Enrico. Using the Semantic Web as Background Knowledge for Ontology Mapping. **In: Proc. of the Int. Workshop on Ontology Matching**. 2006.

SÁNCHEZ, David.; MORENO, Antonio. Learning non-taxonomic relationships from web documents for domain ontology construction. **Data & Knowledge Engineering**, v. 64, n. 3, p. 600-623, 2008.

SHAMSFARD, Mehrnoush.; BARFOROUSH, Ahmad A. The state of the art in ontology learning: A framework for comparison. **Knowledge Engineering Review**, v. 18, n. 4, p. 293-316, 2003.

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

SCHMITZ, Patrick. Inducing ontology from Flickr *tags*. **In: Proceedings of the Collaborative Web Tagging Workshop, 15th WWW Conference**. Edinburgh, Scotland, 2006.

SCHUTZ, Alexander.; BUITELAAR, Paul. RelExt: A tool for relation extraction in ontology extension. **In: Proceedings of the Fourth International Semantic Web Conference**. p. 593-606, 2005.

SHIRKY, Clay. **Ontology is Overrated: Categories, Links, and Tags**. 2005. Disponível em: <[http://shirky.com/writings/ontology\\_overrated.html](http://shirky.com/writings/ontology_overrated.html)> Acesso em agosto, 2008.

SILVA, José. V. **Gerenciamento do vocabulário do usuário em sistemas baseados em tagging**. Dissertação (Mestrado em Ciência da Computação) – Universidade Estadual de Maringá, Maringá-PR, 2009.

SIMPERL, E. Reusing ontologies on the Semantic Web: a feasibility study, **Data & Knowledge Engineering (DKE)** 68 (10) (2009), pp. 905-925

SINTEK, Michael.; BUITELAAR, Paul.; OLEJNIK, Daniel. A protégé e plug-in for ontology extraction from text based on linguistic analysis. **In: Proceedings of the 1st European semantic web symposium (ESWS), 2004.**

SIRIN, Everen.; HENDLER, James.; PARSIA, Bijan. Semi-automatic composition of web services using semantic descriptions. **In: Web Services: Modeling, Architecture and Infrastructure Workshop in Conjunction with ICEIS, 2003.**

SMITH, B.; WELTY, C. Ontology: Towards a new synthesis., **In Chris Welty and Barry Smith, eds., Formal Ontology in Information Systems.** Pp. iii-x. Ogunquit, Maine: ACM Press, 2001.

SPYNS P.. **Evalexon: Assessing triples mined from texts.** Technical report, STAR Lab, Brussel, 2005.

SRIKANT, Ramakrishman.; AGRAWAL, Rakesh. Mining Generalized Association Rules. **In: Proceedings of the 21th international Conference on Very Large Data Bases.** U. Dayal, P. M. Gray, and S. Nishio, Eds. Very Large Data Bases. Morgan Kaufmann Publishers, San Francisco, CA, p. 407-419, sep. 1995.

SOWA, John F. Ontology, metadata and semiotics. **In: Ganter, B. Mineau G., editors, Conceptual structures: Logic, linguistic and Computational issues.** Lecture Notes in Artificial Intelligence, p. 55-81, Springer, 2000.

SPECIA, Lucia.; MOTTA, Enrico. Integrating Folksonomies with the Semantic Web. **In: European Semantic Web Conference (ESWC).** The Semantic Web: Research and Applications. LNCS 4519, p. 503-17, Springer, Heidelberg, 2007.

STAAB, Steffen.; ERDMANN, Michael.; MADCHE, Alexander. Engineering ontologies using semantic patterns. **In: Proceedings of the IJCAI Workshop on E-Business and Intelligent Web, 2001.**

STAAB, Steffen.; STUDER, Rudi.; SCHNURR, Hans-Peter.; SURE, York. Knowledge Processes and Ontologies. **IEEE Intelligent Systems**, v. 16, n. 1, p. 26–34, 2001.

STUDER, Rudi.; BENJAMINS, Richard.; FENSEL, Dieter. Knowledge Engineering: Principles and Methods. **IEEE Transactions on Data and Knowledge Engineering**, v. 25 n.1-2, p. 161–197, 1998.

STUMME Gerd.; MAEDCHE, Alexander. FCA-MERGE: Bottom-Up Merging of Ontologies. **Bernhard Nebel (ed) Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI 2001).** Seattle, Washington. Morgan Kaufmann Publishers, San Francisco, California, p. 225–234, 2001.

SUÁREZ-FIGUEROA, M.C.; K. DELLSCHAFT, E. MONTIEL-PONSODA, B. VILLAZÓN-TERRAZAS, Z. YUFEI, G. AGUADO DE CEA, A. GARCÍA, M. FERNÁNDEZ-LÓPEZ, A. GÓMEZ-PÉREZ, M. ESPINOZA, M. SABOU. NeOn

**Deliverable D5.4.1. NeOn Methodology for Building Contextualized Ontology Networks.** NeOn Project. <http://www.neon-project.org>. February 2008.

SURE, York.; GÓMEZ-PÉREZ, Asunción.; DAELEMANS, Walter.; REINBERGER, Marie-Laure.; GUARINO, Nicola.; NOY, Nataly. F. Why Evaluate Ontology Technologies? Because It Works!. **IEEE Intelligent Systems**, v. 19, n. 4, p. 74-81, jul. 2004.

SURE, York.; STAAB, Steffen.; STUDER, Rudi. On-To-Knowledge Methodology (OTKM). **In Steffen Staab & Rudi Studer, ed., 'Handbook on Ontologies: International Handbook on Information Systems'**. Springer, p. 117-132, 2003.

SURE, York.; STAAB, Steffen.; STUDER, Rudi. Methodology for development and employment of ontology based knowledge management applications. **ACM SIGMOD Record**, v. 31, n. 4, p. 18-23, 2002.

SURE, York.; STUDER, Rudi. **On-To-Knowledge Methodology: final version.** Institute AIFB, University of Karlsruhe, 2002.

SURE, York.; ERDMANN, Michael.; ANGELE, Juergen.; STAAB, Steffen.; STUDER, Rudi.; WENKE, Dirk. OntoEdit: collaborative ontology development for the semantic web. **In: Proceedings of the First International Semantic Web Conference**, Sardinia, Italia, 2002.

SURE, York. **Methodology, Tools and Cases Studies for Ontology based Knowledge Management.** PhD thesis, University of Karlsruhe, 2003.

SURYANTO, Hendra.; COMPTON, Paul. Discovery of Ontologies from Knowledge Bases. **In: Proceedings of the First International Conference on Knowledge Capture**, Eds. Yolanda Gil; Mark Musen; Jude Shavlik, Victoria, British Columbia Canada. The Association for Computing Machinery, New York, USA, p. 171-178, oct. 2001.

SVÁTEK, V.: Design patterns for semantic web ontologies: Motivation and discussion. **In: Proceedings of the 7th Conference on Business Information Systems**, Poznan, 2004.

SWARTOUT, Bill.; RAMESH, Patil.; KNIGHT, Kevin.; RUSS, Tom. Toward Distributed Use of Large- Scale Ontologies. **In: Farquhar A, Gruninger M, Gómez-Pérez A, Uschold M, van der Vet P (eds) AAAI'97 Spring Symposium on Ontological Engineering.** Stanford University, California, p. 138-148, 1997.

TACLA, Cesar A. FREDDO, Ademir R. DALL AGNOL, Josiane M. H. Desenvolvimento de Ontologias Limpas a partir de Etiquetagens (pôster). **ONTOBRÁS. Terceiro Seminário de Pesquisa em Ontologia no Brasil.** 30 e 31 de agosto, 2010. Florianópolis, SC.

TACLA, Cesar A.; FREDDO, Ademir R.; PARAISO, Emerson C.; RAMOS, Milton P.; SATO, Gilson, Y. Supporting small teams in cooperatively building application domain models. **Expert Systems with Applications.** In Press, Corrected Proof, Available online 1 June 2010, ISSN 0957-4174.

TARTIR, Samir.; BUDAK ARPINAR, I. Ontology Evaluation and Ranking using OntoQA. **In: Proceedings of the First IEEE International Conference on Semantic Computing.** Irvine, California, USA, 2007.

TARTIR, Samir.; BUDAK ARPINAR, I., MOORE, Michael.; SHETH, Amit. P.; ALEMAN-MEZA, Boanerges. OntoQA: Metric-based ontology quality analysis. **In: Proceedings of IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources.** Houston, Texas, 2005.

THOMPSON, Cynthia A.; MOONEY, Raymond J. Automatic construction of semantic lexicons for learning natural language interfaces. **In: Proceedings of 16th National Conference on Artificial Intelligence (AAAI'99).** Orlando, FL, USA, p. 487–493, 1999.

USCHOLD, Mike.; KING, Martin. Towards a Methodology for Building Ontologies. **In: Skuce D (eds) IJCAI'95 Workshop on Basic Ontological Issues in Knowledge Sharing.** Montreal, Canada, p. 6.1–6.10, 1995.

USCHOLD, Mike. Building ontologies: Towards a unified methodology. **In 16th Annual Conf. of the British Computer Society Specialist Group on Expert Systems.** Cambridge, UK, 1996.

USCHOLD, Mike.; GRÜNINGER, Michael. Ontologies: Principles, Methods and Applications. **Knowledge Engineering Review**, v. 11, n. 2, p. 93–155, 1996.

VELARDI, Paola.; NAVIGLI, Roberto.; CUCHIARELLI, Alessandro.; NERI, Francesca. Evaluation of ontolearn, a methodology for automatic population of domain ontologies. **In: Ontology Learning from Text: Methods.** Applications and Evaluation. IOS Press, 2005.

VILLAVERDE, Jorge.; PERSSON, Agustin.; GODOY, Daniela.; AMANDI, Analia. Supporting the discovery and labeling of non-taxonomic relationships in ontology learning. **Expert Systems with Application**, v. 36, n. 7, p. 10288-10294, sep. 2009.

VÖLKER, Johana.; VRANDECIC, Denny.; SURE, York.; HOTHÖ, Andreas. 2008. AEON - An approach to the automatic evaluation of ontologies. **Applied Ontology**, v. 3, n. 1-2, p. 41-62, jan. 2008.

VOSSSEN P. EuroWordNet: **A Multilingual Database with Lexical Semantic Networks.** Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

WAL, Vander T. **Explaining and Showing Broad and Narrow Folksonomies.** Disponível em: <[http://www.personalinfocloud.com/2005/02/explaining\\_and\\_.html](http://www.personalinfocloud.com/2005/02/explaining_and_.html)>. Acesso em março, 2008.

WAL, Vander, T. **Folksonomy.** 2007. Disponível em: <<http://vanderwal.net/folksonomy.html>>. Acesso em abril, 2008.

WANG, Yimin.; VÖLKER, Johana.; HAASE, Peter. Towards semi-automatic ontology building supported by large-scale knowledge acquisition. **In AAI Fall Symposium On Semantic Web for Collaborative Knowledge Acquisition**, v. FS-06-06, p. 70-77, Arlington, VA, USA. AAI, AAI Press, 2006.



WELTY Christopher.; GUARINO Nicola. Supporting ontological analysis of taxonomic relation. **Data & Knowledge Engineering**, v. 39, n. 1, p. 51– 74, 2001.

WELTY, Chistopher.; ANDERSEN, W. Towards OntoClean 2.0: A framework for rigidity. **Applied Ontology**, v. 1, n. 1, p. 107-116, jan. 2005.

WENG, Sung-Shung.; TSAI, Hsine.-Jen.; LIU, Shang-Chia.; HSU, Cheng-Hsin. Ontology construction for information classification. **Expert Systems with Applications**, v. 31, n. 1, p. 1-12, 2006.

WIEDERHOLD, Gio. Interoperation, Mediation and Ontologies. **In: International Symposium on Fifth Generation Computer Systems (FGCS94)**. Tokyo, Japan, p. 33-48, 1994.

WILLE, Rudolf. Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts. **In: RIVAL, I. (ed.) Ordered Sets**. Vol. 83 of NATO Advanced Study Institute Series C. Reidel, Dordrecht, p. 445-470, 1982.

WITTEN I. H. et al. KEA: Practical automatic keyphrase extraction. **In: Proceedings of the Fourth ACM Conference on Digital Libraries**, 1999, p. 254-255.

WITTEN, Ian. H.; PAYNTER, Gordon.; FRANK, Eibe.; GUTWIN, Carl.; NEVILANLMANNING, Craig G. Kea: Practical automatic keyphrase extraction. **In: Proceedings of Digital Libraries**, 1999.

WOODS, William. A. What's in a link: Foundations for semantic networks. **In: Representation and Understanding Studies in Cognitive Science**. Academic Press, New York, NY. USA, 1975.

WORDNET. **WNStats – WordNet 3.0 Database Statistics**. Cognitive Science Laboratory. Princeton University, 2009. Disponível em <<http://wordnet.princeton.edu/man/wnstats.7WN>>. Acesso em: abril, 2009.

WU, Xian.; ZHANG, Lei.; YU, Yong. Exploring social annotations for the semantic web. **In: 15th WWW Conference**. Edinburgh, 2006.

WU, Harris.; ZUBAIR, Mohammad.; MALY, Kurt. Harvesting Social Knowledge from Folksonomies. **In: Proc. of HYPERTEXT '06**, 2006.

YAMADA, Ichiro.; BALDWIN, Timothy. Automatic discovery of telic and agentive roles from corpus data. **In: Proc. of the 18th Pacific Asia Conference on Language, Information and Computation**. Tokyo, Japan, 2004.

ZHOU, Lina. Ontology learning: state of the art and open issues. **Inf. Technol. and Management**, v. 8, n. 3, p. 241-252, sep. 2007.



## APÊNDICE A

Neste apêndice as questões utilizadas do questionário eletrônico utilizadas na fase de conceitualização no método de desenvolvimento de ontologias desenvolvido. Nas questões, além de alternativas de múltipla escolha, há textos explicativos, caixas de textos a serem preenchidas e figuras explicativas. Na descrição das questões a seguinte simbologia é utilizada:

- <<tag>>: tag utilizada no processo de anotação pelo projetista;
- <<elemento>>: pode ser um conceito ou uma instância na ontologia raiz;
- <<conceito>>: caixa de texto no formulário/conceito.

### Questão Q1

A tag <<tag>> está conectada ao elemento <<elemento>> na ontologia inicial. Esta observação é importante para a resposta da questão abaixo. Você pode ignorar a tag <<tag>> caso a conexão com o elemento <<elemento>> não seja relevante.

Primeira questão para a tag <<tag>>.

Nesta questão você deve classificar a tag <<tag>> como um conceito ou instância.

<<tag>> é um conceito ou uma instância?

- (1) Conceito
- (2) Instância

### Questão Q1I

A tag <<tag>> é instância de qual conceito? Escolha um da ontologia raiz (opção 1) ou sugira um novo conceito (opção 2).

- (1) Selecione um conceito da ontologia raiz: << lista de conceitos da ontologia raiz >>
- (2) Sugira um novo conceito: << caixa de texto >>

### Questão Q2C

Nesta questão há três alternativas. Para cada alternativa além de um texto explicativo há uma figura que exemplifica.

Escolha uma das alternativas para a tag <<tag>> que você classificou como conceito:

- (1) CONCEITO, CUJO INDIVÍDUOS REPRESENTAM POSSÍVEIS VALORES DE UM ATRIBUTO (TIPO DE). O conceito “Cor” possui dois indivíduos ou valores “Branco” e “Tinto” que classificam o “Vinho” como “Tinto” ou “Branco”. Logo

indivíduos de “Cor” preenchem atributos ou características de indivíduos do conceito “Vinho”. O conceito “Gênero” possui dois indivíduos “Fêmea” e “Macho” que classificam o gênero de uma Pessoa. Outros valores como “Velho”, “Luxo”, “Grande” fornecem valores para atributos de outros indivíduos. As instâncias ou indivíduos do conceito <<tag>> representam valores de atributos para outros indivíduos, ou seja, classificam indivíduos de outros conceitos. Nesse item, o conceito <<tag>> é semelhante aos conceitos "Gênero" e "Cor".

- (2) CONCEITO QUE NÃO POSSUI INDIVÍDUOS, NÃO POSSUI ATRIBUTOS. UTILIZADO PARA ORGANIZAR UMA TAXONOMIA. Alguns conceitos organizam uma taxonomia em uma ontologia. Estes conceitos não possuem instâncias. As instâncias são criadas nos subconceitos, ou conceitos mais específicos. Por exemplo, o conceitos “Animal”, “Mamífero”, “Ave” e “Réptil” são conceitos que organizam uma taxonomia. O conceito mais específico “Gato” possui a instância “Garfield”. Logo, semelhante a “Animal”, “Mamífero”, “Ave” e “Réptil”, o conceito <<tag>> pode ser um conceito de alto nível, apenas para organização de uma taxonomia. O conceito <<tag>> não possui instâncias ou indivíduos. Nesse item, o conceito <<tag>> é semelhante aos conceitos "Animais", "Mamífero", "Ave" e "Réptil".
- (3) CONCEITO QUE PODE SER UMA ESPECIALIZAÇÃO DE UM CONCEITO MAIS GERAL (HERDA ATRIBUTOS) OU CONCEITO QUE FORNECE ATRIBUTOS A OUTROS CONCEITOS ABAIXO DE SUA HIERARQUIA. Por exemplo, os conceitos “Estudante” e “Professor” herdam atributos do conceito “Pessoa”. Se você considera <<tag>> semelhante a “Estudante” e “Professor” (herda atributos) ou semelhante a “Pessoa” (fornece atributos).

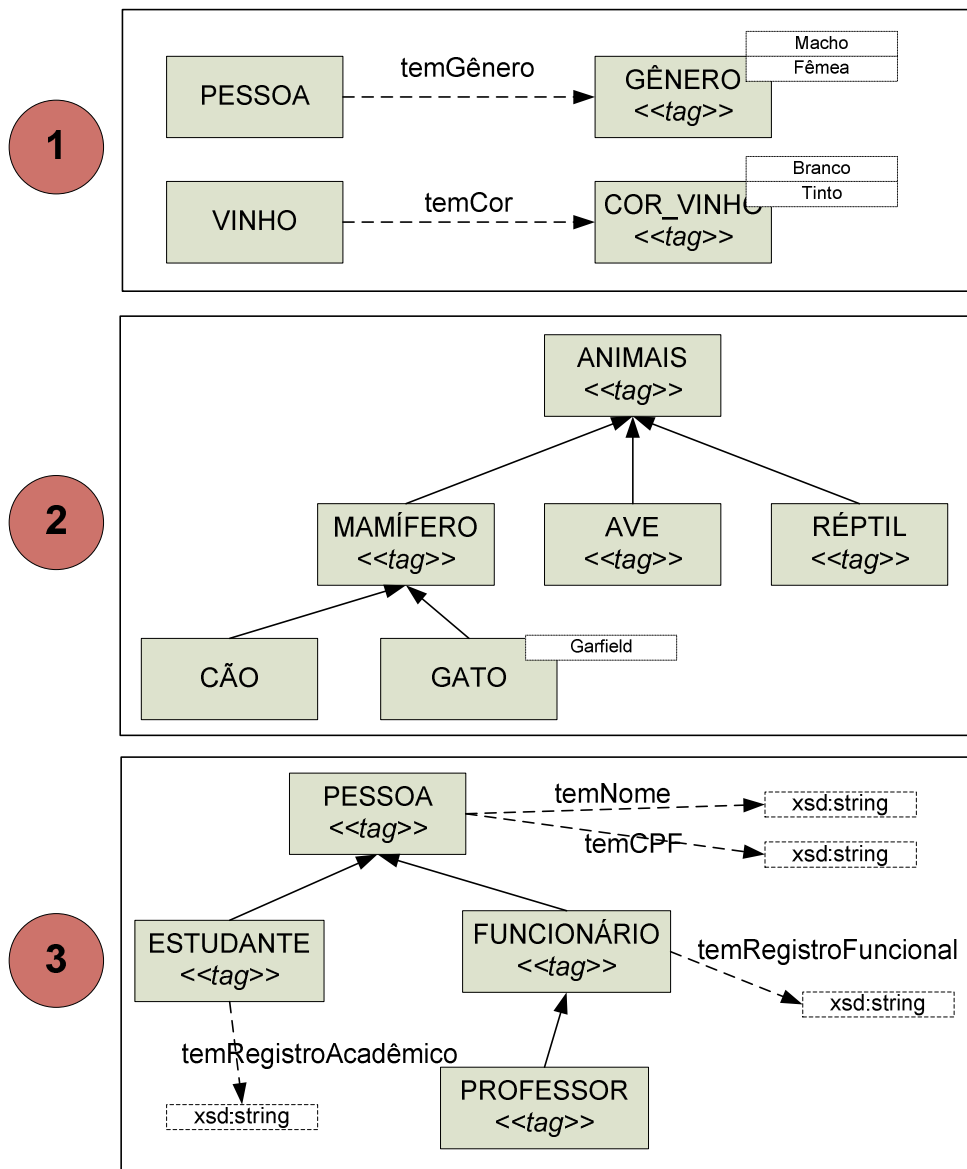


Figura 75 – Figura da Questão Q2C.

### Questão Q3C

Nesta questão há duas alternativas com texto descritivo e exemplificadas por figuras.

Escolha uma das alternativas para a tag `<<tag>>` que você classificou como conceito:

- (1) CONCEITO QUE APENAS FORNECE ATRIBUTOS OU CARACTERÍSTICAS PARA OUTROS ABAIXO DE SUA HIERARQUIA. Por exemplo, “Pessoa”, apenas fornece atributos para os conceitos “Estudante”, “Funcionário” e “Professor”. Não há

um conceito mais geral na hierarquia. Nesse item, o conceito `<<tag>>` é semelhante ao conceito "Pessoa".

- (2) CONCEITO QUE HERDA E FORNECE ATRIBUTOS A OUTROS CONCEITOS. Por exemplo, "Estudante", e "Funcionário" herdam atributos de "Pessoa". Bem como "Professor" herda atributos de "Pessoa" e "Funcionário". Nesta opção o conceito `<<tag>>` é semelhante a "Estudante", "Professor" e "Funcionário".

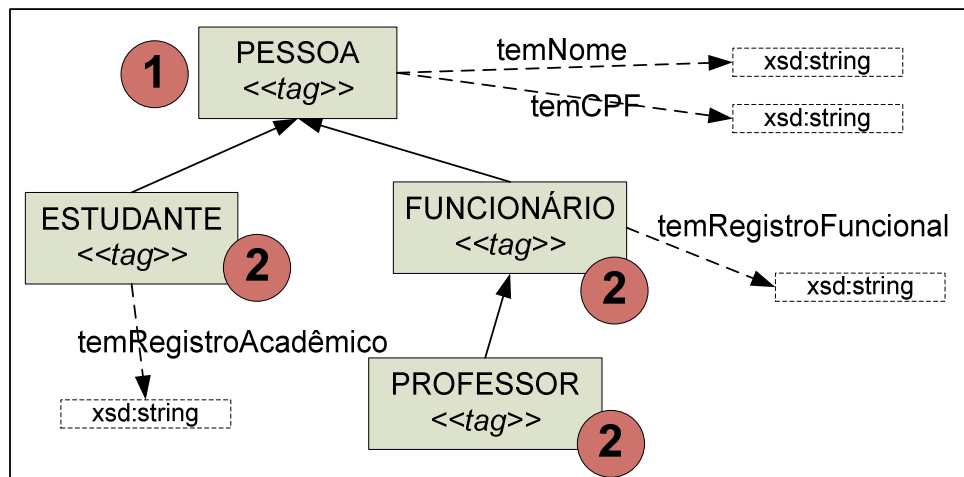


Figura 76 – Figura da Questão Q3C.

### Questão Q4C

Com o passar do tempo (no término dos estudos), um "Estudante" pode deixar de ser "Estudante", mas continua a ser "Pessoa" (estudante é uma especialização de pessoa). O mesmo pode ocorrer com "Professor". O mesmo pode mudar de profissão, mas continua a ser uma "Pessoa" ou um "Funcionário". Com base neste exemplo, escolha uma das opções abaixo para o conceito `<<tag>>`:

- (1) As instâncias do conceito `<<tag>>` podem deixar de ser `<<tag>>`, mas continuam a ser instâncias de um conceito mais geral. Nesse item, o conceito `<<tag>>` é semelhante a "Estudante" e "Professor" no exemplo citado.
- (2) `<<tag>>` sempre será `<<tag>>` em todos os mundos. Nesse item `<<tag>>` é semelhante a "Pessoa" no exemplo citado. Nesse caso não é possível um indivíduo deixar de existir como "Pessoa". Nesse item, o conceito `<<tag>>` é semelhante a "Pessoa".

**Questão Q5C**

Alguns conceitos possuem dependências com outros conceitos. Por exemplo: Para ser "Pai" é necessário ter uma relação com "Filho", para ser "Estudante", é necessário uma relação com "Escola" ou "Faculdade". Com isso, questiona-se: <<tag>> possui alguma dependência com outro conceito, ou seja, para um indivíduo ser <<tag>> é necessário possuir alguma relação com algum indivíduo de outro conceito (semelhante a "Pai" com "Filho" e "Estudante" com "Faculdade")? Se indivíduos de <<tag>> possui alguma dependência com outros conceitos, responda "SIM" a esta questão, caso contrário responda Não.

- (1) Sim
- (2) Não

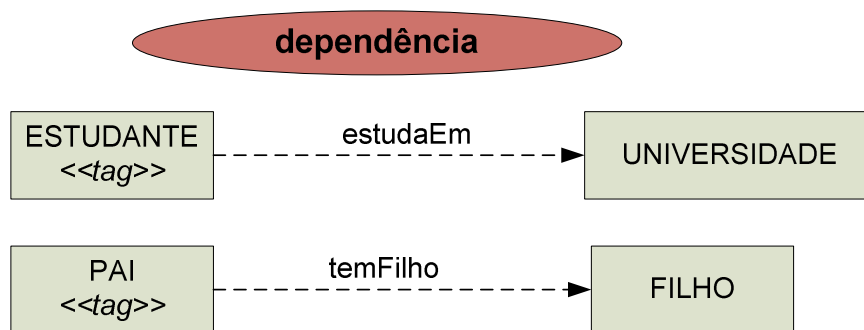


Figura 77 – Figura da Questão Q5C.

**Questão Q5C**

Com base nas suas respostas nas questões anteriores, você deverá nesta questão escolher uma das sugestões abaixo para a *tag* que vc classificou como conceito e para o elemento. Caso nenhuma das sugestões seja relevante, você poderá ignorar as sugestões.

Escolha uma das sugestões abaixo:

- (1) A *tag* <<tag>> é um subconceito do <<elemento>>
- (2) O elemento <<elemento>> é um subconceito da <<tag>>
- (3) A *tag* <<tag>> e o elemento <<elemento>> são irmãos, ou seja, herdam características de um mesmo conceito (*tag* <<tag>> e elemento <<elemento>> são subconceitos de um mesmo conceito).

- (4) A *tag* <<tag>> e o elemento <<elemento>> estão relacionados por uma propriedade. Coloque na caixa de texto ao lado o nome da propriedade.
- (5) Nenhuma das alternativas.

### Questão Q21

Você classificou a *tag* <<tag>> como uma instância ou indivíduo de algum conceito. Este conceito é referenciado como <<conceito>>. Na caixa de texto ao lado você pode incluir o conceito em que a <<tag>> é instância. Você pode escolher um conceito existente na ontologia (relação) ou um novo conceito.

Segunda questão para a *tag* <<tag>>. Escolha uma das opções abaixo:

- (1) INDIVÍDUOS QUE REPRESENTAM POSSÍVEIS VALORES DE UM ATRIBUTO. "Branco" e "Tinto" são indivíduos do conceito "Cor" que classificam o "Vinho" como "Tinto" ou "Branco". Logo indivíduos de "Cor" preenchem atributos ou características de indivíduos do conceito "Vinho". O conceito "Gênero" possui dois indivíduos "Fêmea" e "Macho" que classificam o gênero de uma Pessoa. Outros valores como "Velho", "Luxo", "Grande" fornecem valores para atributos de outros indivíduos. Nesse item a *tag* <<tag>> é semelhante a indivíduos como "Branco", "Tinto", "Fêmea" e "Macho" no exemplo citado. O conceito <<conceito>> em que <<tag>> é instância é semelhante a "Gênero" e "Cor do Vinho" no exemplo.
- (2) INDIVÍDUOS DE ALGUM CONCEITO. Um indivíduo possui atributos ou características que são preenchidas. Por exemplo, indivíduos do Conceito "Pessoa" possuem atributos como CPF, nome, impressão digital. Nesse item, os indivíduos são diferenciados pelos valores de seus atributos. Por exemplo, uma "Pessoa" é diferenciada de outra pelo CPF ou pela impressão digital. Um "Estudante" é caracterizado pelo seu registro acadêmico. Nesse item, <<tag>> é semelhante a indivíduos de "Estudante", "Pessoa", "Funcionário", "Universidade", "Cidades do Paraná", "Teses de Doutorado". O conceito <<conceito>> em que a <<tag>> é instância é semelhante a "Estudante", "Pessoa", "Funcionário", "Universidade", "Cidades do Paraná", "Teses de Doutorado".



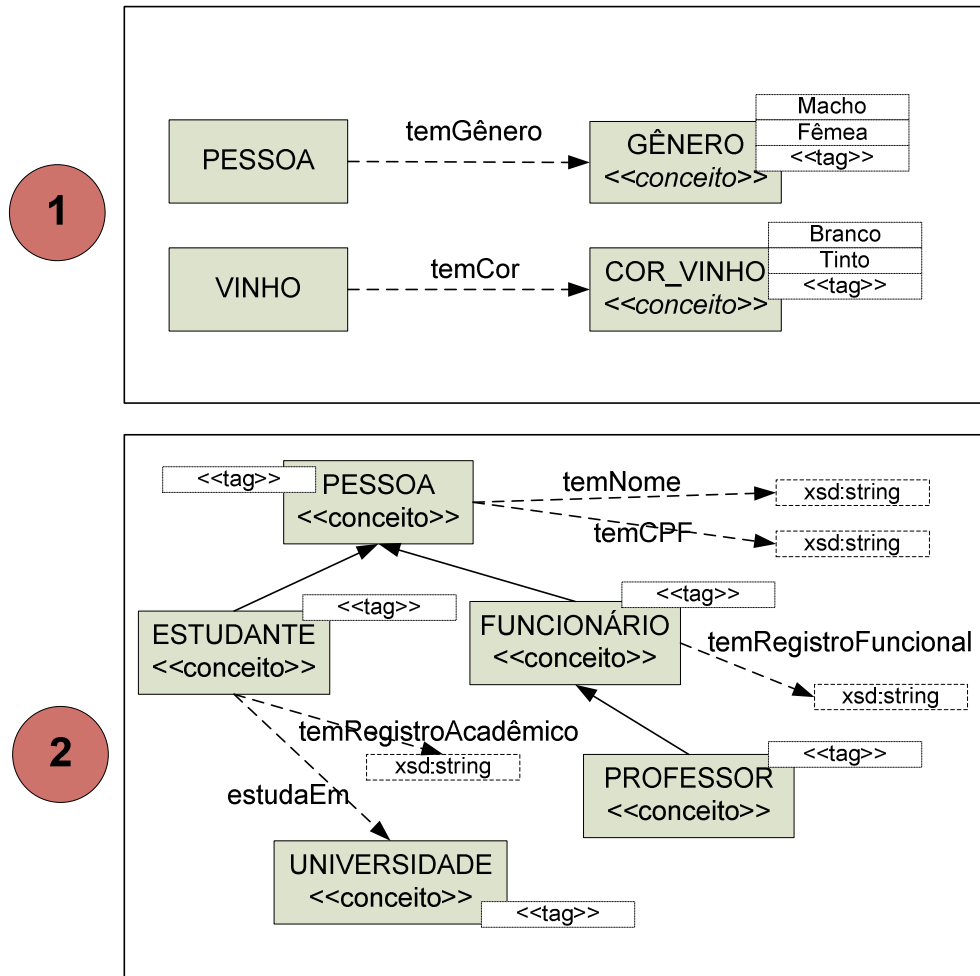


Figura 78 – Figura da Questão Q21.

### Questão Q31

Nesta questão há duas alternativas com texto descritivo e exemplificadas por figuras.

Escolha uma das alternativas para a tag `<<tag>>` que você classificou como instância do conceito `<<conceito>>`:

- (1) INDIVÍDUOS `<<tag>>` DE ALGUM CONCEITO `<<conceito>>` QUE APENAS FORNECEM ATRIBUTOS OU CARACTERÍSTICAS PARA OUTROS INDIVÍDUOS ABAIXO DE SUA HIERARQUIA. Por exemplo, Pessoa, apenas fornece atributos para os conceitos Estudante e Professor. Não há um conceito mais geral na hierarquia. Nesta opção, o conceito `<<conceito>>` que possui como instância a `<<tag>>` é semelhante ao conceito "Pessoa".
- (2) INDIVÍDUOS `<<tag>>` DE ALGUM CONCEITO `<<conceito>>` QUE HERDA E FORNECE ATRIBUTOS A OUTROS CONCEITOS. Por exemplo, "Estudante", e

“Funcionário” herdam atributos de “Pessoa”. Bem como “Professor” herda atributos de “Pessoa” e “Funcionário”. Nesta opção, o conceito <<conceito>> que possui como instância a <<tag>> é semelhante a “Estudante”, “Professor” e “Funcionário”.

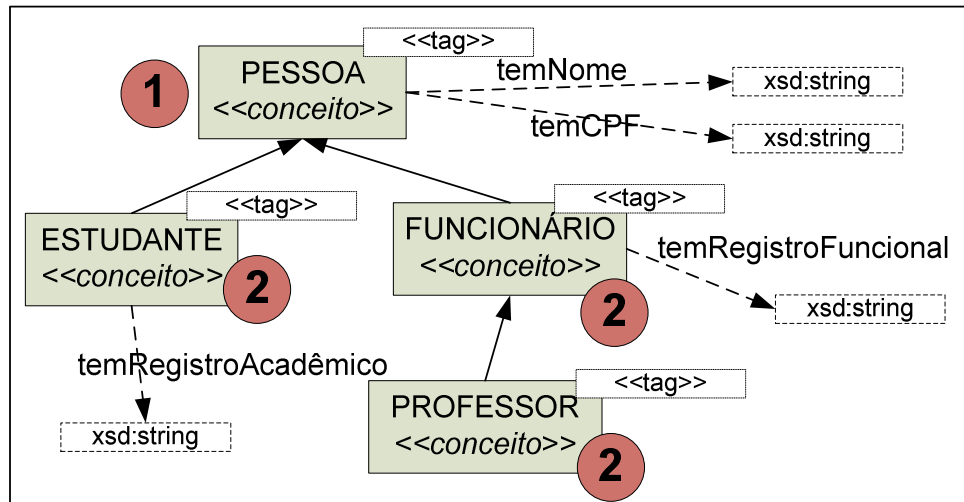


Figura 79 – Figura da Questão Q31.

#### Questão Q41

Lembre-se, você classificou a tag <<tag>> como uma instância de um conceito <<conceito>>.

Alguns indivíduos ou instâncias possuem dependências com outras instâncias. Por exemplo: Para ser "Pai" é necessário ter uma relação com "Filho", para ser "Estudante", é necessário uma relação com "Escola" ou "Faculdade". Com isso, questiona-se: <<tag>> possui alguma dependência com outro indivíduo, ou seja, para um indivíduo ser <<tag>> é necessário possuir alguma relação com algum indivíduo de outro conceito (semelhante a "Pai" com "Filho" e "Estudante" com "Faculdade")? O conceito <<conceito>> em que <<tag>> é instância possui alguma dependência com outro conceito?

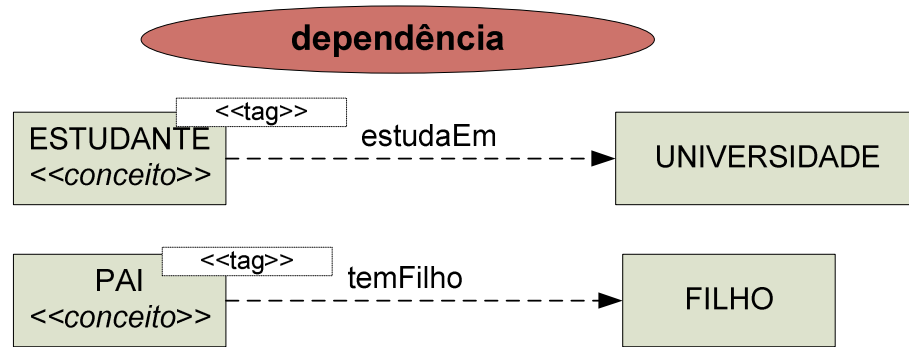


Figura 80 – Figura da Questão Q41.

**Questão Q51**

Lembre-se, você classificou a tag <<tag>> como uma instância de um conceito <<conceito>>.

Com o passar do tempo (no término dos estudos), um "Estudante" pode deixar de ser "Estudante", mas continua a ser uma "Pessoa" ("Estudante" é uma especialização de "Pessoa"). O mesmo pode ocorrer com "Professor". O mesmo pode mudar de profissão, mas continua a ser uma "Pessoa". Uma "Pessoa" nunca pode deixar de ser "Pessoa" em um determinado mundo porque não existe outra classe mais genérica. Com base neste exemplo, escolha uma das opções abaixo:

- (1) O conceito <<conceito>> em que <<tag>> é instância, é semelhante a "Estudante" e "Professor" no exemplo citado.
- (2) O conceito <<conceito>> em que <<tag>> é instância, é semelhante a "Pessoa" no exemplo citado.

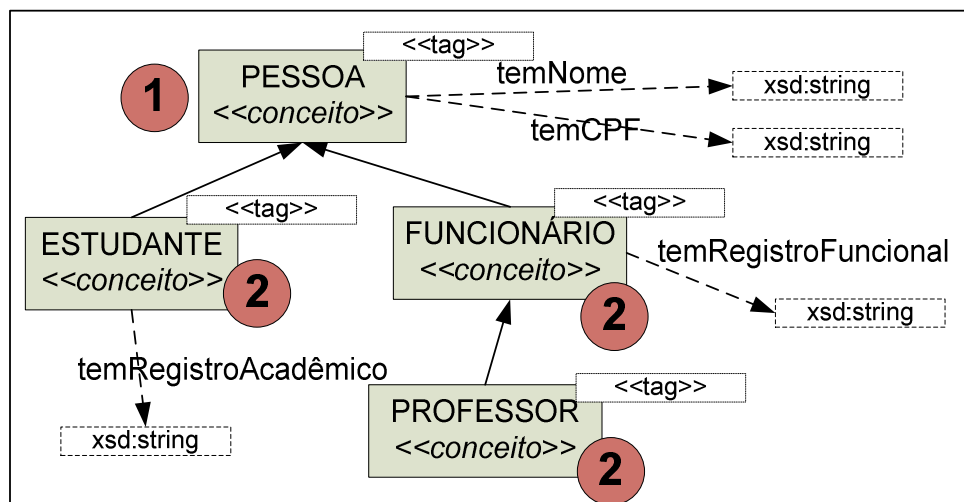


Figura 81 – Figura da Questão Q51.

**Questão QSI**

Com base nas suas respostas nas questões anteriores, você deverá nesta questão escolher uma das sugestões abaixo para a *tag* que você classificou como instância e para o elemento. Caso nenhuma das sugestões seja relevante, você poderá ignorar as sugestões.

Escolha uma das sugestões abaixo:

- (1) A *tag* <<*tag*>> é uma instância do <<*conceito*>>. Este conceito <<*conceito*>> é um subconceito do <<*elemento*>>.
- (2) A *tag* <<*tag*>> é uma instância do <<*conceito*>>. O elemento <<*elemento*>> é um subconceito do conceito <<*conceito*>>.
- (3) A *tag* <<*tag*>> é uma instância do <<*conceito*>>. O conceito <<*conceito*>> e o elemento <<*elemento*>> são irmãos, ou seja, herdam características de um mesmo conceito (conceito <<*conceito*>> e elemento <<*elemento*>> são subconceitos de um mesmo conceito).
- (4) A *tag* <<*tag*>> é uma instância do <<*conceito*>>. O conceito <<*conceito*>> e o elemento <<*elemento*>> estão relacionados por uma propriedade. Coloque na caixa de texto ao lado o nome da propriedade.
- (5) Nenhuma das alternativas.

## APÊNDICE B

O Quadro 14 com as possíveis combinações e sugestões fornecidas no questionamento ao projetista. Na **Erro! Fonte de referência não encontrada.**, tanto A quanto B podem ser *ag* ou elemento da ontologia raiz. Nas sugestões, C pode ser um conceito qualquer da ontologia raiz. As sugestões são fornecidas para as questões *QSI* e *QSC* obedecendo as restrições da filosofia *OntoClean* para os tipos de elementos classificados.

**Quadro 14** - Combinação das Classificações *OntoClean* com Justificativas e Sugestões.

Classificação <i>OntoClean</i>		Justificativa/Restrições	Sugestões
A	B		
+O+I+R+D (type)	+O+I+R+D (type)		$A \sqsubseteq B$ $B \sqsubseteq A$ $(C \supseteq A) \sqcap (C \supseteq B)$
+O+I+R+D (type)	+O+I+R-D (type)	+D subsume -D	$A \sqsubseteq B$ $(C \supseteq A) \sqcap (C \supseteq B)$
+O+I+R-D (type)	+O+I+R+D (type)	+D não subsume -D	$B \sqsubseteq A$ $(C \supseteq A) \sqcap (C \supseteq B)$
+O+I+R-D (type)	+O+I+R-D (type)		$A \sqsubseteq B$ $B \sqsubseteq A$ $(C \supseteq A) \sqcap (C \supseteq B)$
+O+I+R+D (type)	-O+I+R+D (quasi-type)		$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $B \sqsubseteq A$
+O+I+R+D (type)	-O+I+R-D (quasi-type)	+D não subsume -D	$(C \supseteq A) \sqcap (C \supseteq B)$ $B \sqsubseteq A$
+O+I+R-D (type)	-O+I+R+D (quasi-type)	+D não subsume -D	$(C \supseteq A) \sqcap (C \supseteq B)$ $B \sqsubseteq A$
+O+I+R-D (type)	-O+I+R-D (quasi-type)		$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $B \sqsubseteq A$
+O+I+R+D (type)	-O+I~R+D (material role)	-R/~R não subsume +R	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $A \sqsubseteq \exists \text{temPropriedade}.B$
+O+I+R+D (type)	-O+I~R-D (material role)	-R/~R não subsume +R +D não subsume -D	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \sqsubseteq \exists \text{temPropriedade}.B$
+O+I+R-D (type)	-O+I~R+D (material role)	-R/~R não subsume +R +D não subsume -D	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \sqsubseteq \exists \text{temPropriedade}.B$
+O+I+R-D (type)	-O+I~R-D (material role)	-R/~R não subsume +R	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $A \sqsubseteq \exists \text{temPropriedade}.B$
+O+I+R+D (type)	-O+I~R-D (phase sortal)	-R/~R não subsume +R +D não subsume -D	$(C \supseteq A) \sqcap (C \supseteq B)$
+O+I+R-D (type)	-O+I~R-D (phase sortal)	-R/~R não subsume +R	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$
+O+I+R+D (type)	-O+I-R-D (mixin)	-R/~R não subsume +R +D não subsume -D	$(C \supseteq A) \sqcap (C \supseteq B)$
+O+I+R-D (type)	-O+I-R-D (mixin)	-R/~R não subsume +R +D não subsume -D	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$
+O+I+R+D (type)	-O-I+R+D (category)	+I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}.B$
+O+I+R+D (type)	-O-I+R-D (category)	+I não subsume -I +D não subsume -D Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}.B$

+O+I+R-D (type)	-O-I+R+D (category)	+ I não subsume -I +D não subsume -D Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
+O+I+R-D (type)	-O-I+R-D (category)	+ I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
+O+I+R+D (type)	-O-I~R+D (formal role)	-R/~R não subsume +R + I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
+O+I+R+D (type)	-O-I~R-D (formal role)	-R/~R não subsume +R + I não subsume -I +D não subsume -D Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
+O+I+R-D (type)	-O-I~R+D (formal role)	-R/~R não subsume +R + I não subsume -I +D não subsume -D Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
+O+I+R-D (type)	-O-I~R-D (formal role)	-R/~R não subsume +R + I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
+O+I+R+D (type)	-O-I-R/~R-D (attribution)	-R/~R não subsume +R + I não subsume -I +D não subsume -D Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
+O+I+R-D (type)	-O-I-R/~R-D (attribution)	-R/~R não subsume +R + I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I+R+D (quasi-type)	-O+I+R+D (quasi-type)		$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$ $B \sqsubseteq A$
-O+I+R+D (quasi-type)	-O+I+R-D (quasi-type)	+D não subsume -D	$(C \supseteq A) \cap (C \supseteq B)$ $B \sqsubseteq A$
-O+I+R-D (quasi-type)	-O+I+R+D (quasi-type)	+D não subsume -D	$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$
-O+I+R-D (quasi-type)	-O+I+R-D (quasi-type)		$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$ $B \sqsubseteq A$
-O+I+R+D (quasi-type)	-O+I~R+D (material role)	-R/~R não subsume +R	$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I+R+D (quasi-type)	-O+I~R-D (material role)	+D não subsume -D -R/~R não subsume +R	$(C \supseteq A) \cap (C \supseteq B)$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I+R-D (quasi-type)	-O+I~R+D (material role)	+D não subsume -D -R/~R não subsume +R	$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I+R-D (quasi-type)	-O+I~R-D (material role)	-R/~R não subsume +R	$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I+R+D (quasi-type)	-O+I~R-D (phased sortal)	-R/~R não subsume +R +D não subsume -D	$(C \supseteq A) \cap (C \supseteq B)$
-O+I+R-D (quasi-type)	-O+I~R-D (phased sortal)	-R/~R não subsume +R	$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$
-O+I+R+D (quasi-type)	-O+I-R-D (mixin)	-R/~R não subsume +R +D não subsume -D	$(C \supseteq A) \cap (C \supseteq B)$
-O+I+R-D (quasi-type)	-O+I-R-D (mixin)	-R/~R não subsume +R	$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$
-O+I+R+D (quasi-type)	-O-I+R+D (category)	+ I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I+R+D (quasi-type)	-O-I+R-D (category)	+ I não subsume -I Sortal e Non-Sortal +D não subsume -D	$A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I+R-D (quasi-type)	-O-I+R+D (category)	+ I não subsume -I Sortal e Non-Sortal +D não subsume -D	$A \sqsubseteq \exists \text{temPropriedade}. B$

-O+I+R-D (quasi-type)	-O-I+R-D (category)	+ I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I+R+D (quasi-type)	-O-I~R+D (formal role)	+ I não subsume -I -R/~R não subsume +R Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I+R+D (quasi-type)	-O-I~R-D (formal role)	+ I não subsume -I -R/~R não subsume +R +D não subsume -D Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I+R-D (quasi-type)	-O-I~R+D (formal role)	+ I não subsume -I -R/~R não subsume +R +D não subsume -D Sortal e Non-Sortal	$A \supseteq \exists \text{temPropriedade}. B$
-O+I+R-D (quasi-type)	-O-I~R-D (formal role)	+ I não subsume -I -R/~R não subsume +R Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I+R+D (quasi-type)	-O-I~R/~R-D (atribution)	+ I não subsume -I -R/~R não subsume +R +D não subsume -D Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I+R-D (quasi-type)	-O-I~R/~R-D (atribution)	+ I não subsume -I -R/~R não subsume +R Sortal e Non-Sortal	$A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R+D (material role)	-O+I~R+D (material role)		$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R+D (material role)	-O+I~R-D (material role)	+D não subsume -D	$(C \supseteq A) \sqcap (C \supseteq B)$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R-D (material role)	-O+I~R+D (material role)	+D não subsume -D	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R-D (material role)	-O+I~R-D (material role)		$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R+D (material role)	-O+I~R-D (phased sortal)	+D não subsume -D	$(C \supseteq A) \sqcap (C \supseteq B)$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R-D (material role)	-O+I~R-D (phased sortal)		$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R+D (material role)	-O+I~R-D (mixin)	+D não subsume -D	$(C \supseteq A) \sqcap (C \supseteq B)$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R-D (material role)	-O+I~R-D (mixin)		$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R+D (material role)	-O-I+R+D (category)	-R/~R não subsume +R	$(C \supseteq A) \sqcap (C \supseteq B)$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R+D (material role)	-O-I+R-D (category)	+D não subsume -D -R/~R não subsume +R	$(C \supseteq A) \sqcap (C \supseteq B)$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R-D (material role)	-O-I+R+D (category)	+D não subsume -D -R/~R não subsume +R	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R-D (material role)	-O-I+R-D (category)	-R/~R não subsume +R	$(C \supseteq A) \sqcap (C \supseteq B)$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R+D (material role)	-O-I~R+D (formal role)	+ I não subsume -I	$(C \supseteq A) \sqcap (C \supseteq B)$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}. B$
-O+I~R+D	-O-I~R-D	+D não subsume -D	$(C \supseteq A) \sqcap (C \supseteq B)$

(material role)	(formal role)	+ I não subsume -I	$B \sqsubseteq A$ $A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (material role)	-O-I~R+D (formal role)	+D não subsume -D + I não subsume -I	$(C \supseteq A) \cap (C \supseteq B)$ $A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (material role)	-O-I~R-D (formal role)	+ I não subsume -I	$(C \supseteq A) \cap (C \supseteq B)$ $B \sqsubseteq A$ $A \sqsubseteq \exists temPropriedade.B$
-O+I~R+D (material role)	-O-I~R/-R-D (atribution)	+ I não subsume -I +D não subsume -D	$(C \supseteq A) \cap (C \supseteq B)$ $B \sqsubseteq A$ $A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (material role)	-O-I~R/-R-D (atribution)	+ I não subsume -I	$(C \supseteq A) \cap (C \supseteq B)$ $B \sqsubseteq A$ $A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (phased sortal)	-O+I~R-D (phased sortal)		$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$ $B \sqsubseteq A$
-O+I~R-D (phased sortal)	-O+I~R+D (mixin)	+D não subsume -D	$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$
-O+I~R-D (phased sortal)	-O-I+R+D (category)	+D não subsume -D -R/~R não subsume +R + I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (phased sortal)	-O-I+R-D (category)	-R/~R não subsume +R + I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (phased sortal)	-O-I~R+D (formal role)	+D não subsume -D + I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (phased sortal)	-O-I~R-D (formal role)	+ I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (phased sortal)	-O-I~R/-R-D (atribution)	+ I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (mixin)	-O+I~R-D (mixin)		$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$ $B \sqsubseteq A$ $A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (mixin)	-O-I+R+D (category)	+D não subsume -D + I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (mixin)	-O-I+R-D (category)	+D não subsume -D + I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (mixin)	-O-I~R+D (formal role)	+D não subsume -D + I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (mixin)	-O-I~R-D (formal role)	+ I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists temPropriedade.B$
-O+I~R-D (mixin)	-O-I~R/-R+D (atribution)	+D não subsume -D + I não subsume -I Sortal e Non-Sortal	$A \sqsubseteq \exists temPropriedade.B$
-O-I+R+D (category)	-O-I+R+D (category)		$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$ $B \sqsubseteq A$ $A \sqsubseteq \exists temPropriedade.B$
-O-I+R+D (category)	-O-I+R-D (category)	+D não subsume -D	$(C \supseteq A) \cap (C \supseteq B)$ $B \sqsubseteq A$ $A \sqsubseteq \exists temPropriedade.B$
-O-I+R-D (category)	-O-I+R+D (category)	+D não subsume -D	$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$ $A \sqsubseteq \exists temPropriedade.B$
-O-I+R-D (category)	-O-I+R-D (category)		$(C \supseteq A) \cap (C \supseteq B)$ $A \supseteq B$



			$B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}.B$
$-O-I+R+D$ (category)	$-O-I\sim R+D$ (formal role)	$-R/\sim R$ não subsume $+R$	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $A \sqsubseteq \exists \text{temPropriedade}.B$
$-O-I+R+D$ (category)	$-O-I\sim R-D$ (formal role)	$+D$ não subsume $-D$ $-R/\sim R$ não subsume $+R$	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \sqsubseteq \exists \text{temPropriedade}.B$
$-O-I+R-D$ (category)	$-O-I\sim R+D$ (formal role)	$+D$ não subsume $-D$ $-R/\sim R$ não subsume $+R$	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $A \sqsubseteq \exists \text{temPropriedade}.B$
$-O-I+R-D$ (category)	$-O-I\sim R-D$ (formal role)	$-R/\sim R$ não subsume $+R$	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $A \sqsubseteq \exists \text{temPropriedade}.B$
$-O-I+R+D$ (category)	$-O-I\sim R/-R-D$ (attribution)	$-R/\sim R$ não subsume $+R$ $+D$ não subsume $-D$	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \sqsubseteq \exists \text{temPropriedade}.B$
$-O-I+R-D$ (category)	$-O-I\sim R/-R-D$ (attribution)	$-R/\sim R$ não subsume $+R$	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $A \sqsubseteq \exists \text{temPropriedade}.B$
$-O-I\sim R+D$ (formal role)	$-O-I\sim R/-R+D$ (formal role)		$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}.B$
$-O-I\sim R+D$ (formal role)	$-O-I\sim R/-R-D$ (formal role)	$+D$ não subsume $-D$	$(C \supseteq A) \sqcap (C \supseteq B)$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}.B$
$-O-I\sim R-D$ (formal role)	$-O-I\sim R/-R+D$ (formal role)	$+D$ não subsume $-D$	$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $A \sqsubseteq \exists \text{temPropriedade}.B$
$-O-I\sim R-D$ (formal role)	$-O-I\sim R/-R-D$ (formal role)		$(C \supseteq A) \sqcap (C \supseteq B)$ $A \supseteq B$ $B \sqsubseteq A$ $A \sqsubseteq \exists \text{temPropriedade}.B$