

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO

FABRÍCIO MARTINS MENDONÇA

**ONTOFORINFOSCIENCE: METODOLOGIA PARA CONSTRUÇÃO DE
ONTOLOGIAS PELOS CIENTISTAS DA INFORMAÇÃO**
**Uma aplicação prática no desenvolvimento da ontologia
sobre componentes do sangue humano (HEMONTO)**

Belo Horizonte

2015

FABRÍCIO MARTINS MENDONÇA

**ONTOFORINFOSCIENCE: METODOLOGIA PARA CONSTRUÇÃO DE
ONTOLOGIAS PELOS CIENTISTAS DA INFORMAÇÃO**
**Uma aplicação prática no desenvolvimento da ontologia
sobre componentes do sangue humano (HEMONTO)**

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Escola de Ciência da Informação da Universidade Federal de Minas Gerais para obtenção do grau de Doutor em Ciência da Informação.

Linha de Pesquisa: Organização e Uso da Informação

Orientador: Maurício Barcellos Almeida

Co-orientador: Renato Rocha Souza

BELO HORIZONTE

2015

Mendonça, Fabrício Martins

M539o Ontoforinfoscience: metodologia para construção de ontologias pelos
cientistas da informação: uma aplicação prática no desenvolvimento da ontologia
sobre componentes do sangue humano (HEMONTA) / Fabrício Martins
Mendonça. – 2015.
320 f.: il.

Orientador: Maurício Barcellos Almeida.

Tese (Doutorado) – Universidade Federal de Minas Gerais, Escola de Ciência
da Informação.

Referências: f. 283-297.

Apêndices: f. 298-320.

1. Ciência da Informação – Teses. 2. Ontologia. 3. Ontologias biomédicas. 3.
Hematologia. I. Título. II. Almeida, Maurício Barcellos. III. Universidade Federal
de Minas Gerais, Escola de Ciência da Informação.

CDU: 111:616.15

Ficha catalográfica: Bibliotecária Raquel de Oliveira Torres, CRB 6/2786.



UFMG

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

FOLHA DE APROVAÇÃO


"ONTOFORINFOSCIENCE: METODOLOGIA PARA CONTRUÇÃO DE ONTOLOGIAS PELOS CIENTISTAS DA INFORMAÇÃO: UMA APLICAÇÃO PRÁTICA NO DESENVOLVIMENTO DA ONTOLOGIA SOBRE COMPONENTES DO SANGUE HUMANO (HEMONTO)"

Fabício Martins Mendonça

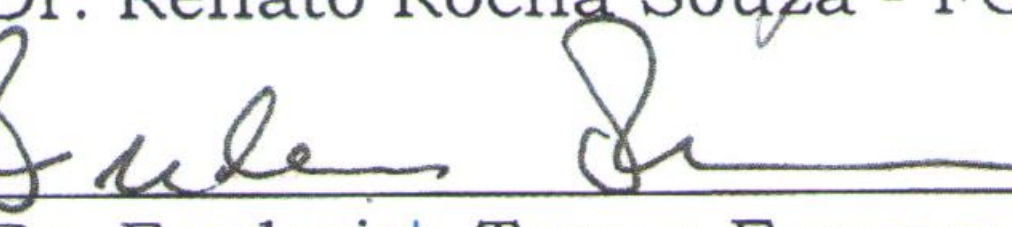
Tese submetida à Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais, como parte dos requisitos à obtenção do título de "**Doutor em Ciência da Informação**", linha de pesquisa "**Organização e Uso da Informação**".

Tese aprovada em: 01 de julho de 2015.

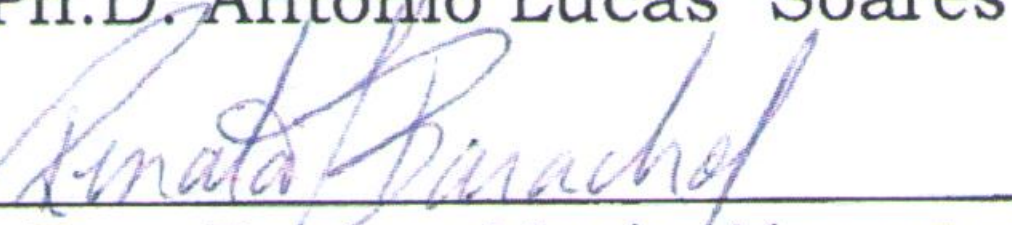
Por:

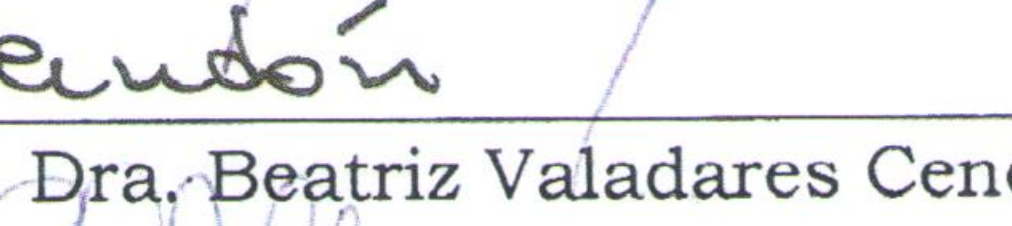

Prof. Dr. Maurício Barcellos Almeida - ECI/UFMG (Orientador)


Prof. Dr. Renato Rocha Souza - FGV/RJ (Co-orientador)


Prof. Dr. Frederico Torres Fonseca - Pennsylvania State University

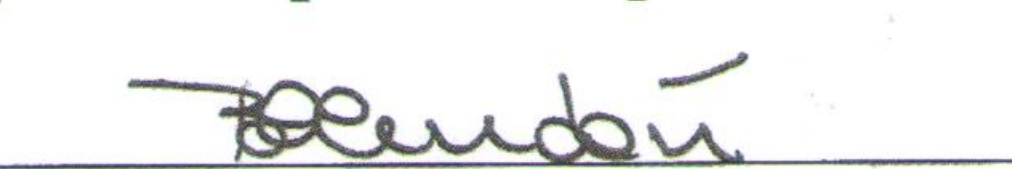

Prof. Ph.D. António Lucas Soares - Faculdade do Porto (por videoconferência)


Profa. Dra. Renata Maria Abrantes Baracho Porto - ECI/UFMG

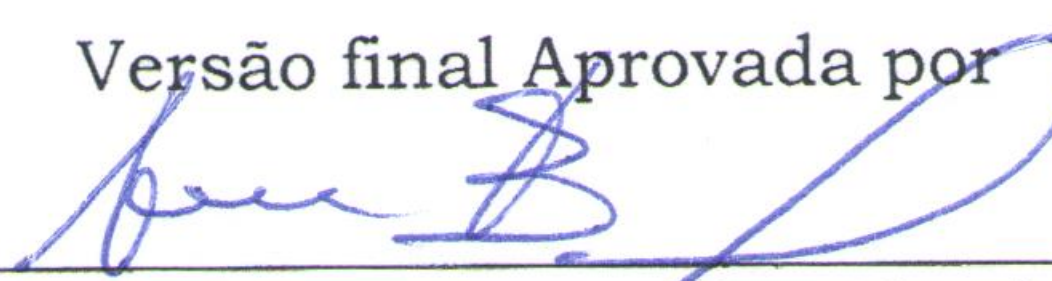

Profa. Dra. Beatriz Valadares Cendón - ECI/UFMG


Profa. Célia da Consolação Dias - ECI/ UFMG

Aprovada pelo Colegiado do PPGCI


Profa. Beatriz Valadares Cendón
Coordenadora

Versão final Aprovada por


Prof. Maurício Barcellos Almeida
Orientador



UFMG

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

ATA DA DEFESA DE TESE DE **FABRÍCIO MARTINS MENDONÇA**, matrícula: 2011710248

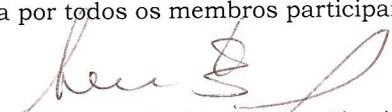
Às 14:00 horas do dia 01 de julho de 2015, reuniu-se na Escola de Ciência da Informação da UFMG a Comissão Examinadora aprovada *ad referendum* pela Coordenadora do Programa de Pós-Graduação em Ciência da Informação em 02/06/2015, para julgar, em exame final, o trabalho intitulado ***Ontoforinfoscience: metodologia para construção de ontologias pelos cientistas da informação: uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano (hemonto)***, requisito final para obtenção do Grau de DOUTOR em CIÊNCIA DA INFORMAÇÃO, área de concentração: Produção, Organização e Utilização da Informação, Linha de Pesquisa: Organização e Uso da Informação. Abrindo a sessão, o Presidente da Comissão, Prof. Dr. Maurício Barcellos Almeida, após dar conhecimento aos presentes do teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Logo após, a Comissão se reuniu sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:


Prof. Dr. Maurício Barcellos Almeida - Orientador	APROVADO
Prof. Dr. Renato Rocha Souza - Co-orientador	APROVADO
Prof. Dr. Frederico Torres Fonseca	APROVADO
Prof. Ph.D. António Lucas Soares (por videoconferência)	APROVADO
Profa. Dra. Renata Maria Abrantes Baracho Porto	APROVADO
Profa. Dra. Beatriz Valadares Cendón	APROVADO
Profa. Célia da Consolação Dias	APROVADO


Pelas indicações, o candidato foi considerado APROVADO.

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

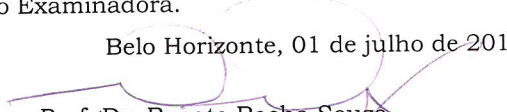
Belo Horizonte, 01 de julho de 2015

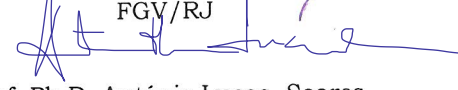

Prof. Dr. Maurício Barcellos Almeida
ECI/UFMG


Prof. Dr. Frederico Torres Fonseca
(Pennsylvania State University)



Profa. Dra. Renata Maria Abrantes Baracho Porto
ECI/UFMG


Profa. Célia da Consolação Dias
ECI/UFMG


Prof. Dr. Renato Rocha Souza
FGV/RJ


Prof. Ph.D. António Lucas Soares
Faculdade do Porto


Profa. Dra. Beatriz Valadares Cendón
ECI/UFMG


Profa. Beatriz Valadares Cendón
Coordenadora do Programa Pós-Graduação
em Ciência da Informação - ECI/ UFMG

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

AGRADECIMENTOS

Primeiramente à Deus, que nos dá força e sua luz nos momentos mais difíceis, para mostrar que a realização de sonhos depende de nossa força interior.

À minha mãe, Marilza, pelo amor, afeto, carinho e a dedicação inigualável que tem por mim.

À meu pai, José Carlos, pelo incentivo e apoio em me conduzir no caminho acadêmico.

Ao professor Maurício, não apenas pelos seus saberes científicos, mas sobretudo pelos seus conselhos e amizade, além da polidez e educação na condução de uma orientação.

Ao professor António, pela grande acolhida em Porto, por seus saberes técnico-científicos em uma orientação no exterior, ensinamentos pessoais e culturais.

À Juliana, pelo amor, carinho, companherismo e compreensão do que é fazer pesquisa e se dedicar horas aos estudos.

À minha irmã, Carla, sempre na torcida e incentivo para a realização de mais esse sonho.

“Cada sonho que você deixa para trás é um
pedaço do seu futuro que deixa de existir.”

(Steve Jobs)

“A sabedoria é um adorno na prosperidade e
um refúgio na adversidade.”

(Aristóteles)

RESUMO

Considerando as pesquisas sobre organização e uso da informação relacionadas à área de Ciência da Informação é possível notar, nos últimos anos, um crescimento significativo na utilização de ontologias como instrumentos de representação do conhecimento e em outras aplicações da área. No contexto biomédico, por exemplo, as ontologias são usadas na padronização dos vocabulários médicos como instrumento com maior poder de expressividade em relação às terminologias médicas. Entretanto, mesmo com o crescimento notável do uso de ontologias, inclusive no contexto da Ciência da Informação, algumas necessidades e problemas permanecem em aberto, talvez em função da proximidade recente deste tema com esta área. Dentre os problemas identificados na área de ontologias, conhecida como engenharia ontológica, que impactam de alguma forma em sua relação com a Ciência da Informação, destacam-se aqueles relacionados com o processo de construção de uma ontologia, muitas vezes, obscuro para especialistas em Organização do Conhecimento. A presente pesquisa realizou uma investigação desses problemas e, a partir desse estudo, propõe uma metodologia própria para a construção de ontologias - a *OntoForInfoScience* -, direcionada, principalmente, para especialistas em Organização do Conhecimento, provenientes do campo de Ciência da Informação. Tal metodologia tem como principal diferencial em relação às metodologias atuais o detalhamento das atividades executadas ao longo do processo de construção de ontologias, as quais não são bem explicadas nessas metodologias. É importante destacar também que a metodologia proposta não é totalmente inédita, já que sua elaboração foi baseada em algumas etapas das conhecidas metodologias NeOn, Methontology e método 101, detalhando etapas destas metodologias e fazendo adaptações necessárias para Ciência da Informação. Como resultados da metodologia *OntoForInfoScience*, realizou-se uma aplicação prática de tal metodologia no desenvolvimento de uma ontologia sobre componentes do sangue humano, denominada de HEMONTO. Tal ontologia possui um caráter científico e social relevantes, uma vez que se propõe a ser um instrumento de apoio a médicos, laboratoristas e profissionais da área em geral, na extração, manipulação, armazenamento e uso adequados dos componentes de sangue no tratamento de pacientes, visando salvar vidas. Ao final do processo de construção da HEMONTO, conclui-se que a metodologia desenvolvida é de grande utilidade para cientistas da informação, e mesmo para outros profissionais que atuam na organização do conhecimento, para a criação de representações ontológicas formais. Como trabalhos futuros, sugere-se a elaboração de um handbook detalhado sobre a metodologia *OntoForInfoScience* e também o desenvolvimento de um software que guie o processo de construção de ontologias a partir da metodologia desenvolvida. Essas medidas são importantes para divulgar melhor a metodologia *OntoForInfoScience* e permitir que mais desenvolvedores possam utilizá-la na construção de ontologias dos mais diferentes domínios do conhecimento.

Palavras-chave: Ontologia. Desenvolvimento de ontologias. Metodologia para construção de ontologias. Ontologias biomédicas. Hematologia. Sangue.

ABSTRACT

Considering the researches about organization and use of information related to Information Science, it is possible to notice, in the last years, a meaning growing in using of ontologies as knowledge representation instruments and in other applications of the area. In the biomedical context, for example, the ontologies are used in the standardization of the medical vocabularies as instrument with higher power of expressivity over medical terminologies. However, even with the remarkable growing of the using of ontologies, including the context of Information Science, some needs and problems remain opened, possibly because of the recent proximity of this topic with this area. Among the problems identified in the area of ontologies, named as ontological engineering, that impact of some way in your relation with the Information Science, the problems related with the process of building of an ontology stand out more, because, more times, they are obscure for specialists in Knowledge Organization. This research held an investigation of these problems and, from that study, it is proposed own methodology for building of ontologies – the *OntoForInfoScience* -, oriented, mainly, specialists in Knowledge Organization, from the field of Information Science. Such methodology brings as novelty the detailing of the steps of the ontology development cycle, which are not well explained in the current methodologies for building of ontologies. It is also important to highlight that the methodology proposed is not entirely new, because its development was based in some steps of the know methodologies NeOn, Methontology and method 101, detailing steps of these methodologies and making adaptations required for Information Science. As results of the *OntoForInfoScience* methodology, it was performed a practical application of this methodology in the development of an ontology about human blood components, named HEMONTO. This ontology has a scientific and social nature relevant, because it proposes to be a support instrument for physicians, laboratorists and specialists of this area in general, in the extraction, manipulation, storage and use appropriate of the blood components in the health care, helping save lifes. At the end of the development process of HEMONTO, it concludes that the new methodology is useful for information scientists, and even for other professional of the area of Knowledge Organization, for the creating formal ontological representations. As future works, it is suggested the elaboration of detailed handbook about the *OntoForInfoScience Methodology* and also the development of the software that supports the building process of ontologies from methodology developed. These suggestions are important to spread the *OntoForInfoScience Methodology* and to allow that more developers can used it in the development of ontologies of different knowledge domains.

Keywords: Ontology. Ontologies development. Methodology for building of ontologies. Biomedical ontologies. Hematology. Blood components.

LISTA DE FIGURAS

Figura 1 - Diagrama esquemático de um sistema geral de comunicação.....	38
Figura 2 – Representação de uma ontologia e sua relação com a conceitualização	53
Figura 3 - Os vários subdomínios biomédicos integrados ao UMLS.....	63
Figura 4 - Exemplos de identificadores no MetaThesaurus UMLS.	64
Figura 5 - Extrato da rede semântica UMLS.	64
Figura 6 - O sexteto ontológico e suas relações formais-ontológicas.	73
Figura 7 - Cronologia das ontologias de fundamentação pesquisadas.....	75
Figura 8 - Página de vocabulários da Open-CYC.	77
Figura 9 - Taxonomia das categorias mais básicas da DOLCE.....	79
Figura 10 - Relações primitivas básicas da DOLCE representadas em FOL.	80
Figura 11 - Hierarquia das categorias alto-nível da ontologia de Sowa.....	82
Figura 12 - Busca pelo termo "record" na interface da SUMO.....	84
Figura 13 - Busca pelo termo "system" na Wordnet e suas relações com a SUMO.....	84
Figura 14 - Diagrama das categorias mais básicas da GFO.....	87
Figura 15 - Representação esquemática das categorias básicas da UFO.	90
Figura 16 - Parte da taxonomia da IAO contendo entidades dependentes continuanes.	93
Figura 17 - Parte mais detalhada da taxonomia da IAO contendo dependentes continuanes.	93
Figura 18 - Taxonomia das entidades mais básicas da BFO.....	97
Figura 19 – Taxonomia geral da BFO 2.0	99
Figura 20 – Ontologia SNAP da BFO 2.0.....	100
Figura 21 - Ontologia SPAN da BFO 2.0.....	100
Figura 22 - Arquitetura ontológica do projeto BIOTOP.....	110
Figura 23 - Registro da classe Célula na Gene Ontology (GO).	117
Figura 24 - Hierarquia <i>is-a</i> das principais classes da Anatomy taxonomy do FMA.	119
Figura 25 - Taxonomia FMA da entidade anatômica do sangue.....	121
Figura 26 - Estrutura taxonômica da Cell Ontology apresentada por meio do BioPortal.....	123
Figura 27 - Arquitetura geral da Protein Ontology (PRO).....	125
Figura 28 - Relação das etapas das metodologias de Uschold e Gruninger e Methontology	137
Figura 29 - Conjunto de representações intermediárias da Methontology	139
Figura 30 - Cenários para construção de redes ontológicas pela NeOn.....	141
Figura 31 - Taxonomia de relações part-of nas visões mereológica e meronímica	159
Figura 32 - Esquema de representação da metodologia de pesquisa adotada.	176

Figura 33 - Etapas da metodologia <i>OntoForInfoScience</i>	185
Figura 34 - A etapa de aquisição e extração do conhecimento da metodologia <i>OntoForInfoScience</i>	195
Figura 35 - A etapa de conceitualização da metodologia <i>OntoForInfoScience</i>	199
Figura 36 - Partonomia dos produtos derivados do sangue.....	208
Figura 37 - Processos de obtenção dos hemocomponentes Concentrado de Hemácias e Concentrado de Plaquetas.....	209
Figura 38 - Ontologias importadas para a HEMONTO no Protégé 4.3.....	213
Figura 39 - Resultado da busca pelo termo "portion of body"na ferramenta OntoBee	217
Figura 40 - Parte da taxonomia geral da ontologia HEMONTO	219
Figura 41 - Preenchimento das propriedades da classe "portion of body substance" no Protégé 4.3	223
Figura 42 - Preenchimento das restrições de tipo e cardinalidade no Protégé 4.3.....	227
Figura 43 - Criação da definição formal de uma classe no Protégé 4.3.....	228
Figura 44 - Preenchimento das propriedades de dados de uma classe no Protégé 4.3	231
Figura 45 - Exemplo de hierarquia natural no domínio do sangue.....	234
Figura 46 - Hierarquia de tipos da relação part_of na ontologia HEMONTO	241
Figura 47 - Preenchimento das propriedades da relação "occurs_in" no Protégé 4.3	244
Figura 48 - Preenchimento do IRI de acesso à ontologia, homepage e anotações gerais.....	252
Figura 49 - Partonomia de uma porção de sangue.	258
Figura 50 - Taxonomia dos tipos de leucócitos.	259
Figura 51 - Partonomia dos produtos derivados do sangue.....	260
Figura 52 - Análise do processo básico de amostras de sangue.....	261
Figura 53 - Tipos de células sanguíneas patológicas.	262
Figura 54 - Parte da taxonomia de entidades continuantes da HEMONTO.....	264
Figura 55 - Extrato da taxonomia de entidades ocorrentes da HEMONTO	266
Figura 56 – Representação do processo de punção venosa.	267
Figura 57 – Taxonomia dos tipos de anticoagulantes.	268
Figura 58 - Parâmetros básicos do sangue como elementos sanguíneos.....	268
Figura 59 - Processos de obtenção dos hemocomponentes Concentrado de Hemácias e Concentrado de Plaquetas	269
Figura 60 - Processos de obtenção dos hemocomponentes e hemoderivados do plasma.....	271
Figura 61 - Processo de obtenção do hemocomponente Concentrado de Granulócitos	272

LISTA DE TABELAS

Tabela 1 - Classificação das relações segundo a Teoria Geral da Terminologia	49
Tabela 2 - Ontologias biomédicas.....	66
Tabela 3 - Ontologias biomédicas integradas a OBO.....	68
Tabela 4 - Quadrado ontológico Aristotélico.....	72
Tabela 5 - Exemplos (instâncias) de categorias básicas da DOLCE.....	79
Tabela 6 - Mapeamento das categorias DOLCE para GFO.....	87
Tabela 7 - Mapeamento das categorias da ontologia de Sowa para a GFO	88
Tabela 8 - Tipos de entidades da ontologia de continuantes da BFO 2.0.....	101
Tabela 9 - Tipos de entidades da ontologia de ocorrentes da BFO 2.0.....	103
Tabela 10 - As dez relações formais da RO em sua primeira versão.....	106
Tabela 11 - Atributos e exemplos das relações da RO	108
Tabela 12 - Extrato da taxonomia das categorias principais da OpenGALEN.....	114
Tabela 13 - Estrutura da ChEBI Ontology.....	127
Tabela 14 - Metodologias para construção de ontologias.....	130
Tabela 15 - Etapas das metodologias para construção de ontologias.....	133
Tabela 16 - Informações ontológicas indispensáveis às classes de uma ontologia	153
Tabela 17 - Definição semi-formal das relações <i>part_of</i> e <i>is_a</i> usadas na HEMONTA	157
Tabela 18 - Tipos ontológicos da relação <i>is_a</i>	158
Tabela 19 - Tipos ontológicos da relação <i>part-of</i>	159
Tabela 20 - Outros tipos de relações <i>part-of</i> : espaciais, temporais e imateriais.....	161
Tabela 21 - Tipos de <i>pitfalls</i> encontrados na construção de ontologias.....	166
Tabela 22 - Conjunto de relações usado no esquema de representação da metodologia.....	177
Tabela 23 - Plano de ação para elaboração da metodologia <i>OntoForInfoScience</i>	180
Tabela 24 - Passos realizados e artefatos gerados a cada etapa da <i>OntoForInfoScience</i>	186
Tabela 25 - Parte do Glossário de Conceitos da HEMONTA	200
Tabela 26 - Parte do Dicionário de Conceitos da HEMONTA	200
Tabela 27 - Parte da Tabela de Conceitos e Valores da HEMONTA	201
Tabela 28 – Alguns hemocomponentes do sangue humano e seus usos indicados.....	203
Tabela 29 - Parte do Dicionário de Verbos da HEMONTA	205
Tabela 30 - Propriedades descritivas das classes de uma ontologia.....	221
Tabela 31 - Propriedades de dados (data properties) da classe "erythrocytes concentrate"	231

Tabela 32 - Exemplos de conceitos com propriedades específicas que são classes	233
Tabela 33 - Relações ontológicas da RO e BFO utilizadas na construção da HEMONTO	236
Tabela 34 - Algumas relações específicas do domínio do sangue na HEMONTO	238
Tabela 35 - Tipos ontológicos das relações <i>is_a</i> e <i>part_of</i> utilizados na HEMONTO	239
Tabela 36 - Propriedades descritivas e lógicas de uma relação ontológica.....	242
Tabela 37 - Parâmetros e critérios avaliativos da metodologia <i>OntoForInfoScience</i>	246
Tabela 38 - Correspondência entre os atributos de uma classe e seu nome no Protégé.....	273
Tabela 39 - Correspondência entre os atributos de uma relação e seus nomes no Protégé.....	274
Tabela 40 - Avaliação da HEMONTO pelos critérios da metodologia <i>OntoForInfoScience</i>	275
Tabela 41 - Dicionário de Classes Continuanes da HEMONTO	298
Tabela 42 - Dicionário de Classes Ocorrentes da HEMONTO	308
Tabela 43 - Dicionário de Relações da ontologia HEMONTO	311

LISTA DE QUADROS

Quadro 1 - Template de Especificação de uma ontologia.....	189
Quadro 2 – Definição e forma de obtenção do componente Concentrado de Hemácias	203
Quadro 3 - Template de Documentação de uma ontologia.....	249
Quadro 4 - Template de Especificação da ontologia HEMONTO	254

LISTA DE ABREVIATURAS

AMO - ABSTRACT MEDIATING OCCURRENT
ANSI – AMERICAN NATIONAL STANDARDS INSTITUTE
ANVISA - AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA
APO - AGENTIVE PHYSICAL OBJECT
ASA - ANATOMICAL STRUCTURAL ABSTRACTION
AT - ANATOMY TAXONOMY
ATA - ANATOMICAL TRANSFORMATION ABSTRACTION
ATCC - AMERICAN TYPE CELL COLLECTION
ATO - ABSTRACT TOP ONTOLOGY
BFO - BASIC FORMAL ONTOLOGY
BIOTOP - BIOLOGICAL TOP-LEVEL
BIREME - BIBLIOTECA REGIONAL DE MEDICINA
CABIG - NATIONAL CANCER INSTITUTE'S BIOMEDICAL INFORMATICS GRID
CDD - CLASSIFICAÇÃO DECIMAL DE DEWEY
CDU - CLASSIFICAÇÃO DECIMAL UNIVERSAL
CHEBI - CHEMICAL ENTITIES OF BIOLOGICAL INTEREST
CIA - CENTRAL INTELLIGENCE AGENCY
CID - CLASSIFICAÇÃO INTERNACIONAL DE DOENÇAS
CL - CELL ONTOLOGY
CONCEPTME - CONCEPTUALIZON MODELLING ENVIROMENMT
CRG - CLASSIFICATION RESEARCH GROUP
CUI - CONCEPT UNIQUE IDENTIFIER
CYCL - CYC LANGUAGE
DDI - ONTOLOGY FOR DRUG DISCOVERY INVESTIGATIONS
DECS - DESCRITORES EM CIÊNCIAS DA SAÚDE
DOLCE - DESCRIPTIVE ONTOLOGY FOR LINGUISTIC AND COGNITIVE
ENGINEERING
FIPS - FEDERAL INFORMATION PROCESSING STANDARDS
FMA - FOUNDATIONAL MODEL ANATOMY
FOL - FIRST ORDER LOGICAL
GALEN - GENERALISED ARCHITECTURE FOR LANGUAGES, ENCYCLOPAEDIAS
AND NOMENCLATURES IN MEDICINE
GFO - GENERAL FORMAL ONTOLOGY
GO - GENE ONTOLOGY

GRAIL - GALEN REPRESENTATION AND INTEGRATION LANGUAGE
HL7-RIM - HL7'S REFERENCE INFORMATION MODEL
IAO - INFORMATION ARTIFACT ONTOLOGY
IFDAO - INTEGRATED FRAMEWORK FOR DEVELOPMENT AND APPLICATION OF ONTOLOGIES
IFOMIS - INSTITUTE FOR ONTOLOGY AND MEDICAL INFORMATION SCIENCE
IEEE - INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS
IHTSDO - INTERNATIONAL HEALTH TERMINOLOGY STANDARDS DEVELOPMENT ORGANIZATION
ISFO - INTEGRATED SYSTEM OF FOUNDATIONAL ONTOLOGIES
ISO - INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
KIF - KNOWLEDGE INTERCHANGE FORMAT
LILACS - LITERATURA LATINO-AMERICANA E DO CARIBE EM CIÊNCIAS DA SAÚDE
MESH - MEDICAL SUBJECT HEADINGS
MILO - MIDLEVEL ONTOLOGY
MK - META-KNOWLEDGE
MMO - MEASUREMENT METHOD ONTOLOGY
NCBO - NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY
NCI THESAURUS - NATIONAL CENTER INSTITUTE'S THESAURUS
NDF-RT - NATIONAL DRUG FILE REFERENCE TERMINOLOGY
NEMO - NEURAL ELECTROMAGNETIC ONTOLOGIES
NISO - NATIONAL INFORMATION STANDARDS ORGANIZATION
NIST - NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
NLM - NATIONAL LIBRARY OF MEDICINE
OBI - ONTOLOGY FOR BIOMEDICAL INVESTIGATIONS
OBO - OPEN BIOLOGICAL ONTOLOGIES
OBO-RO - OPEN BIOMEDICAL ONTOLOGIES-RELATION ONTOLOGY
OCHRE - OBJECT-CENTERED HIGH-LEVEL REFERENCE ONTOLOGY
ODPS - ONTOLOGY DESIGN PATTERNS
OLS - ONTOLOGY LOOKUP SERVICE
OGMS - ONTOLOGY FOR GENERAL MEDICAL SCIENCE
OMIM - ONLINE MENDELIAN INHERITANCE IN MAN
OMS - ORGANIZAÇÃO MUNDIAL DA SAÚDE
ONTOMED - ONTOLOGIAS EM MEDICINA
OPAS - ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE

OTKM - ON-TO-KNOWLEDGE METHODOLOGY
OWL - ONTOLOGY WEB LANGUAGE
OWL-DL - ONTOLOGY WEB LANGUAGE DESCRIPTION LOGIC
P24 – PLASMA DE 24 HORAS
PATO - PHENOTYPIC TRAIT ONTOLOGY
PFC – PLASMA FRESCO CONGELADO
PLN - PROCESSAMENTO DE LINGUAGEM NATURAL
POB - PHYSICAL OBJECTS
PR - PROTEIN ONTOLOGY
PRAXIS - PROCESSO PARA APLICATIVOS EXTENSÍVEIS E INTERATIVOS
PSI-MOD - PROTEIN MODIFICATION ONTOLOGY
RUP - RATIONAL UNIFIED PROCESS
RBC – RED BLOOD CELL
RDF - RESOURCE DESCRIPTION FRAMEWORK
RDF-S - RESOURCE DESCRIPTION FRAMEWORK-SCHEMA
RO - RELATION ONTOLOGY
SGD - SACCHAROMYCES GENOME DATABASE
SNOMED - SYSTEMATIZED NOMENCLATURE OF MEDICINE
SNOMED-CT - SYSTEMATIZED NOMENCLATURE OF MEDICINE-CLINICAL TERMS
SNOMED-RT - SNOMED -REFERENCE TERMINOLOGY
SNOP - STANDARD NOMENCLATURE PATHOLOGY
SO - SEQUENCY ONTOLOGY
SOB - NON-PHYSICAL SOCIAL OBJECTS
SRI - SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÃO
SUI - STRING UNIQUE IDENTIFIER
SUMO - STANDARD UPPER MERGED ONTOLOGY
UFO - UNIFIED FORMAL ONTOLOGY
UML - UNIFIED MODEL LANGUAGE
UMLS - UNIFIED MEDICAL LANGUAGE SYSTEM
UP - UNIFIED PROCESS
UPON - UP FOR ONTOLOGY
UWDA - DIGITAL ANATOMIST SYMBOLIC KNOWLEDGE BASE
W3C - WORLD WEB WIDE CONSORTIUM
WBC – WHITE BLOOD CELL
WFOL - WONDERWEB FOUNDATIONAL ONTOLOGIES LIBRARY
ZFA - ZEBRAFISH ANATOMY AND DEVELOPMENT

SUMÁRIO

1 INTRODUÇÃO	19
1.1 Problema de pesquisa	22
1.2 Objetivos.....	25
1.2.1 Objetivo geral	25
1.2.2 Objetivos específicos.....	26
1.3 Justificativa	26
2 TEORIAS CIENTÍFICAS	30
2.1 O que são teorias?	30
2.2 Teorias no âmbito da Ciência.....	32
2.3 Das teorias terminológicas às teorias ontológicas	36
3 ORGANIZAÇÃO DA INFORMAÇÃO	42
3.1 Instrumentos de organização da informação	43
3.1.1 Vocabulários controlados	44
3.1.2 Terminologias	46
3.1.3 Ontologias	51
3.2 Organização da informação médica.....	57
3.2.1 Terminologias e ontologias biomédicas	58
3.2.2 Repositórios de ontologias	67
4 ONTOLOGIAS: FUNDAMENTOS E APLICAÇÕES	71
4.1 Fundamentos ontológicos básicos	71
4.2 Ontologias de fundamentação.....	74
4.3 Ontologias de fundamentação para biomedicina	94
4.4 Ontologias biomédicas de domínio	113
5 ONTOLOGIAS: CONSTRUÇÃO, AVALIAÇÃO E IMPLEMENTAÇÃO	129
5.1 Metodologias para construção de ontologias	129
5.1.1 Methontology	136
5.1.2 Metodologia NeOn.....	140

5.2 Análise do cenário atual das metodologias	143
5.3 Aspectos essenciais na construção de ontologias	146
5.3.1 Aspectos gerais	147
5.3.2 Caracterização das relações ontológicas.....	154
5.4 Critérios de avaliação de conteúdos ontológicos	161
5.5 Implementação de ontologias por linguagens lógicas.....	169
6 METODOLOGIA	174
6.1 Caracterização e tipologia da pesquisa.....	174
6.2 Descrição geral da metodologia de pesquisa.....	176
6.3 Passos adotados na elaboração da metodologia <i>OntoForInfoScience</i>	178
7 METODOLOGIA ONTOFORINFOSCIENCE.....	184
7.0 Avaliação da necessidade da ontologia (Etapa 0)	187
7.1 Especificação da ontologia (Etapa 1).....	188
7.2 Aquisição e Extração de conhecimento (Etapa 2).....	191
7.3 Conceitualização (Etapa 3)	197
7.4 Fundamentação ontológica (Etapa 4)	209
7.5 Formalização da ontologia (Etapa 5).....	214
7.6 Avaliação da ontologia (Etapa 6).....	244
7.7 Documentação da ontologia (Etapa 7)	249
7.8 Disponibilização da ontologia (Etapa 8)	250
8 HEMONTO: ONTOLOGIA SOBRE OS COMPONENTES DO SANGUE HUMANO	254
8.1 Documento de Especificação	254
8.2 Documentos de Referência	256
8.3 Modelos Conceituais	258
8.4 Ontologias reutilizadas	262
8.5 Conteúdo ontológico.....	263
8.6 Métricas de Avaliação	274
9 CONCLUSÕES E TRABALHOS FUTUROS.....	278

REFERÊNCIAS	283
APÊNDICE A – DICIONÁRIO DE CLASSES CONTINUANTES	298
APÊNDICE B – DICIONÁRIO DE CLASSES OCORRENTES	308
APÊNDICE C – DICIONÁRIO DE RELAÇÕES ONTOLÓGICAS	311

1 Introdução

Com o intuito de designar o problema do crescimento exponencial da informação e seus registros em função das novas tecnologias desenvolvidas, já em 1945, o pesquisador Vannevar Bush, do *Massachusetts Institute of Technology* (MIT), referiu-se a esse fenômeno como o “problema da explosão informacional”, afim de chamar a atenção da sociedade da época para esse ocorrido, como relatado em Bush (1945).

Desde essa época até os dias atuais, muita coisa mudou e diversos outros fenômenos sócio-informacionais ocorreram, mas ainda sim o problema da explosão informacional continua sendo um desafio para a sociedade contemporânea. Esse problema, inclusive, está presente em dimensões ainda maiores nos dias de hoje, muito em função do uso intensivo das novas tecnologias da informação e comunicação, tais como os sistemas de informação eletrônicos, os dispositivos móveis, sites de e-commerce, redes sociais, tecnologias semânticas, entre outros. O uso de tais tecnologias pode tanto contribuir com a explosão informacional, impondo dificuldades a seu usuário diante do grande volume de informação, quanto apoiar o usuário em seu processo de busca da informação, tornando mais simples e eficiente o acesso a informação requerida.

Frente a este cenário, as pesquisas em Ciência da Informação têm muito a contribuir na tarefa de organização da informação, de forma que ela possa ser usada, efetivamente e eficientemente, por seus usuários. Para tanto, a área desenvolveu, ao longo de sua trajetória como ciência, um arcabouço teórico e metodológico afim de possibilitar a organização e uso da informação. Esse repertório da Ciência da Informação inclui instrumentos terminológicos para organização da informação, que incluem os vocabulários controlados, os tesouros, as terminologias e, a partir da década de 90 (VICKERY, 1997; SOERGEL, 1997), também as ontologias.

Atualmente, as potencialidades de uso das ontologias em Ciência da Informação e em áreas correlacionadas estão reforçadas pelo grande volume e diversidade de dados manipulados na *web*, e também pelo avanço no uso de tecnologias semânticas na *web*, as quais fazem parte do que se convencionou chamar de *web semântica* (BREITMAN, 2005). Nesse sentido, Almeida (2013) afirma que as representações formais, como as ontologias, cresceram em importância para uso em computadores e, atualmente, não tem sido usadas apenas em sistemas computacionais. Um grande exemplo disso é o uso de ontologias na estruturação de conhecimento em domínios biomédicos. Smith (2004) destaca que o papel fundamental desempenhado pelas ontologias na área biomédica é servir de iniciativa para padronização do vocabulário médico. Embora essa função seja semelhante àquela das terminologias médicas, as

ontologias biomédicas são instrumentos de representação mais expressivos que as terminologias, devido ao seu uso por computadores, e possibilitam também uma representação do conhecimento do domínio mais próxima da realidade, o que é essencial no caso da biomedicina (FREITAS e SCHULZ, 2009).

Considerando esse cenário atual, favorável ao uso de ontologias para organização e representação do conhecimento, além da evidente relação do tema “ontologia” com a área de Ciência da Informação, o presente trabalho descreve uma pesquisa inédita, realizada nos últimos quatro anos, sobre os principais aspectos e atividades envolvidas no desenvolvimento de ontologias como instrumentos de representação do conhecimento.

De maneira mais específica, o que a presente pesquisa propõe é uma metodologia própria para o desenvolvimento de ontologias, direcionada, principalmente, para especialistas em Organização do Conhecimento, provenientes do campo de Ciência da Informação. Tal metodologia tem como principal diferencial em relação às metodologias atuais o detalhamento das atividades executadas ao longo do processo de construção de ontologias, as quais não são bem explicadas nessas metodologias.

A presente pesquisa inclui ainda uma aplicabilidade prática da metodologia proposta, denominada de *OntoForInfoScience*, através do desenvolvimento de uma ontologia em um domínio biomédico caracterizado pelos componentes do sangue humano usados para fins terapêuticos, os quais são chamados de hemocomponentes e hemoderivados do sangue. Tal ontologia, denominada de HEMONTO, está inserida no âmbito de um projeto de escopo maior – o *Blood Project* – que tem por objetivo gerar uma linguagem de representação formal sobre o sangue humano (ALMEIDA et al., 2010) para propiciar a organização deste domínio.

Sobre a metodologia de construção proposta, é importante afirmar que a mesma não representa uma metodologia completamente inédita, uma vez que reutiliza algumas etapas de outras metodologias já existentes, e nem uma metodologia unificada que servirá como padrão de desenvolvimento ontológico, suprindo o problema da não existência de um padrão de construção na área (JONES, BENCH-CAPON e VISSER, 1998) (FERNANDÉZ-LOPEZ, 1999).

O que se pretende com a metodologia *OntoForInfoScience* é que ela seja um instrumento útil de suporte aos desenvolvedores de ontologias, principalmente aqueles ligados à área de Ciência da Informação, ao detalhar, através de uma linguagem simples, etapas essenciais no ciclo de desenvolvimento ontológico, que não estão bem explicadas ou explicitadas nas metodologias de construção atuais. Seguindo essa abordagem, a metodologia *OntoForInfoScience* fez uso de algumas etapas das conhecidas metodologias

NeOn (SUÁREZ-FIGUEROA et al., 2008), Methontology (FERNÁNDEZ et al., 1997) e método 101 (NOY e MCGUINNES, 2001), detalhando etapas destas metodologias e fazendo adaptações necessárias de cada uma em uma linguagem mais adequada aos desenvolvedores de ontologia, incluindo os cientistas da informação.

A criação de metodologias próprias para construção de ontologias, unificando etapas de metodologias existentes e fazendo adaptações a essas metodologias para que se adequem as necessidades da área, vem ganhando espaço na literatura da área (SANTOS, 2014). O diferencial da presente pesquisa é a proposição de uma nova metodologia de construção que possibilite aos especialistas em Organização do Conhecimento, provenientes do campo da Ciência da Informação, superar problemas relativos ao jargão técnico e a questões lógicas e filosóficas no desenvolvimento de ontologias.

Um último registro nessa parte introdutória da tese vai para a ontologia sobre componentes do sangue humano (HEMONTO), desenvolvida a partir da metodologia *OntoForInfoScience*. A opção pela construção de uma ontologia nesse domínio está atrelada tanto à necessidade de seu desenvolvimento no escopo *Blood Project* quanto ao seu ineditismo no domínio biomédico. Pelas pesquisas realizadas até o momento não foi encontrada nenhuma uma ontologia específica para tratar tal tipo de conhecimento, que é de extrema relevância social para o estado de saúde da população, pois através da extração, manipulação e armazenamento adequados dos componentes do sangue humano torna-se possível realizar importantes tratamentos hemoterapêuticos e evitar mortes.

Para descrever a presente pesquisa, ela foi estruturada da seguinte forma: no restante do capítulo 1, apresentam-se o problema de pesquisa, objetivos geral e específicos, além da justificativa; o referencial teórico e empírico da pesquisa são descritos nos capítulos seguintes, do 2 (dois) ao 5 (cinco): o capítulo 2 apresenta uma explicação clara e sucinta sobre teorias científicas e sua relação com as ontologias, uma vez que a ontologia sobre componentes do sangue humano desenvolvida (HEMONTO) é compreendida como uma teoria ontológica; o capítulo 3 discorre sobre a organização da informação no âmbito da Ciência da Informação, com o enfoque para organização da informação biomédica, destacando alguns instrumentos usados nesta área do conhecimento, tais como terminologias e ontologias biomédicas; o capítulo 4 descreve ontologias de fundamentação e ontologias de domínio biomédico atualmente utilizadas na área de engenharia ontológica e que são relevantes para a temática desta pesquisa; o capítulo 5 apresenta, analisa e discute aspectos fundamentais para os processos de construção, avaliação e implementação de ontologias, com enfoque para as metodologias

de construção. O capítulo 6 traz a metodologia de pesquisa adotada na elaboração da metodologia para construção de ontologias proposta e sua aplicação no desenvolvimento da ontologia sobre componentes do sangue humano; o capítulo 7 descreve passo a passo a metodologia *OntoForInfoScience* proposta, incluindo suas etapas e exemplos de sua aplicação no desenvolvimento da ontologia HEMONTO; o capítulo 8 traz a documentação formal da ontologia HEMONTO, incluindo elementos textuais, artefatos gráficos e representações formais da ontologia desenvolvida; e, por fim, o capítulo 9 apresenta conclusões e considerações finais acerca desta pesquisa, além de possíveis trabalhos futuros sobre o estudo aqui desenvolvido.

1.1 Problema de pesquisa

Em razão das características e potencialidades das ontologias como instrumentos de representação do conhecimento e diante de um cenário atual favorável ao seu uso para organização da informação, é possível notar, nos últimos anos, uma grande expansão na construção e aplicação de ontologias em diferentes tarefas das mais diversas áreas do conhecimento.

No contexto biomédico, como citado anteriormente, as ontologias têm sido utilizadas na padronização dos vocabulários médicos como uma alternativa às terminologias médicas, trazendo maior poder de expressividade na representação do conhecimento do domínio a partir de seus formalismos lógicos e também possibilitando uma representação deste domínio mais próxima da realidade. O aumento na construção e aplicação de ontologias biomédicas pode ser notada, principalmente, pelo surgimento de repositórios especializados, tal como o *Open Biomedical Ontologies (OBO)* (SMITH et al., 2007), visando à integração e padronização de ontologias biomédicas, disponíveis *on line* em domínio público.

Mesmo com o crescimento notável das atividades de pesquisa sobre ontologias e de suas aplicações práticas nos últimos anos, algumas necessidades e problemas da área permanecem em aberto, talvez em função da imaturidade de ser uma área de pesquisa recente, a qual é denominada de engenharia ontológica. Dentre os problemas identificados, atualmente, na área de engenharia ontológica, aqueles que são objeto de estudo da presente pesquisa correspondem aos problemas que surgem ao longo do processo de construção de uma ontologia, os quais serão explorados de forma aprofundada ao longo da investigação aqui conduzida.

Inicialmente, nesta seção sobre a problemática de pesquisa, destacam-se duas classes mais gerais de problemas no desenvolvimento de ontologias: a primeira está relacionada com as limitações das metodologias atuais para construção de ontologias e a

segunda engloba os problemas nas ontologias decorrentes de um processo de construção mal conduzido.

Sobre as limitações atuais das metodologias para construção de ontologias pode-se afirmar que:

- **Não há um padrão de construção ou uma metodologia unificada** que seja amplamente aceita para o desenvolvimento de ontologias (USCHOLD e GRUNINGER (1996); (JONES, BENCH-CAPON e VISSER (1998); FERNANDÉZ-LOPEZ (1999); FERNANDÉZ-LOPEZ e CORCHO (2004); BREITMAN (2005); CARDOSO (2007); SILVA, SOUZA e ALMEIDA (2008). Tais metodologias, ainda hoje, não são consideradas suficientemente “maduras” como as metodologias empregadas na área de Engenharia de Software.
- As metodologias atuais apresentam **abordagens e características diversas, sendo direcionadas a propósitos e aplicações específicas** (FERNANDÉZ-LOPEZ (1999)), tais como aquelas dedicadas ao domínio dos negócios.
- Em relação a detalhes das atividades e dos procedimentos para sua condução, algumas **metodologias e métodos mostram-se superficiais na elucidação dos passos para construção de ontologias**, parecendo até que considerar que o ontologista já domina o assunto sobre construção de ontologias e não necessita de detalhes acerca de atividades e procedimentos envolvidos (SILVA, SOUZA e ALMEIDA, 2008). Um exemplo disso é a ausência de orientações para ajudar o ontologista a fazer escolhas ontológicas sobre a inclusão ou não de um dado conhecimento na ontologia.
- **Algumas abordagens dão mais ênfase em atividades de desenvolvimento**, especialmente a implementação da ontologia, por exemplo os métodos CYC e 101, desconsiderando aspectos importantes relacionados a gerenciamento do projeto, a estudo de viabilidade, à manutenção e à avaliação de ontologias (SILVA, SOUZA e ALMEIDA, 2008).
- **O uso recorrente de conceitos específicos da área de Ciência da Computação não explicados adequadamente aos demais desenvolvedores de ontologias** dificulta seriamente o trabalho de construção de ontologias por parte de profissionais que não são da área computacional. Herança múltipla, cardinalidade, tipos de valores: string, inteiro, real, booleano, classe abstrata, classe disjunta são apenas alguns exemplos de conceitos comuns em Ciência da Computação e nas metodologias para construção de ontologias que podem

trazer dificuldades a desenvolvedores que não são da área, como é o caso de muitos cientistas da informação.

Os problemas relacionados com as metodologias de construção impactam diretamente a segunda classe de problemas, a qual engloba problemas no conteúdo das ontologias desenvolvidas. Entretanto, é importante considerar que nem todos problemas desta classe são culpa da metodologia de construção, pois há também erros decorrentes do projeto ontológico, de especificação, de não avaliação e não documentação, além de desconhecimento ou falha humana dos desenvolvedores de ontologias. Alguns exemplos de problemas relacionados ao conteúdo de uma ontologia são:

- Inclusão de relações que não são genuinamente ontológicas (MUNN e SMITH, 2008);
- Criação e definição de relações de maneira intuitiva, com base em nossa própria experiência de vida (MUNN e SMITH, 2008);
- Criação de instâncias como classes, tornando a hierarquia da ontologia excessivamente especializada (NOY e MCGUINNES, 2001);
- Inclusão de heranças múltiplas: uma subclasse da ontologia que está ligada a mais de uma superclasse (SMITH, 2005);
- Utilização imprópria da relação *is_a* entre classes da ontologia, misturando os diferentes tipos de relação *is_a* que podem existir, tais como instanciação, especialização e equivalência (POVEDA-VILLALÓN, 2010);
- Criação e utilização de relações todo-parte (*has_part* e *part_of*) ambíguas e inconsistentes (BITTNER e DONNELLY, 2007);
- Não caracterização ou não utilização dos diferentes tipos de relação *part_of* existentes no domínio representado na ontologia (KEET e ARTALE, 2008);
- Definição de relações muito genéricas (CEUSTERS et al., 2003);
- Construção de definições formais de classes e de relações de forma imprecisa (SCHULZ, 2006).

Diante deste cenário atual na área de engenharia ontológica, surge como oportunidade de pesquisa a investigação por princípios, métodos, técnicas e processos mais detalhados que possam contribuir com o desenvolvimento de ontologias. Ao vislumbrar tal oportunidade, a presente pesquisa apresenta uma nova metodologia para construção de ontologias direcionada, principalmente, a especialistas em Organização do

Conhecimento e que pretende encontrar respostas para as seguintes questões de pesquisa:

- ✓ A elaboração de uma metodologia própria para construção de ontologias direcionada, principalmente, a especialistas em Organização do Conhecimento, provenientes do campo de Ciência da Informação, é uma solução viável para tratar os problemas enfrentados por estes desenvolvedores ao usarem as metodologias de construção tradicionais da área?
- ✓ O detalhamento de atividades e processos previstos no desenvolvimento de ontologias, em uma metodologia para este fim, é suficiente para o entendimento e apreensão de todas as atividades necessárias do processo de construção ontológico por parte dos desenvolvedores, incluindo aqueles menos familiarizados com conceitos da área?
- ✓ A aplicabilidade prática de uma metodologia própria para construção de ontologias pode ser medida a partir do desenvolvimento de ontologias de domínio usando tal metodologia?

1.2 Objetivos

Considerando a atual importância e relevância das aplicações de ontologias, especialmente em domínios biomédicos, esta pesquisa pretende trazer contribuições à área de engenharia ontológica, especificamente às metodologias de construção, ao atender o objetivo geral e aos objetivos específicos descritos a seguir.

1.2.1 Objetivo geral

O **objetivo geral** da presente pesquisa é propor uma nova metodologia para a construção de ontologias, direcionada, principalmente, a especialistas em Organização do Conhecimento, provenientes do campo da Ciência da Informação, auxiliando-os superar problemas relativos ao jargão técnico e a questões lógicas e filosóficas no desenvolvimento de ontologias.

1.2.2 Objetivos específicos

Nesta pesquisa, pretende-se cumprir os seguintes objetivos específicos:

- ✓ Levantar alguns aspectos ontológicos essenciais no processo de construção de ontologias, de forma que eles sejam considerados na metodologia de construção proposta nesta pesquisa.
- ✓ Detalhar, na metodologia proposta, todas atividades necessárias no processo de desenvolvimento de uma ontologia.
- ✓ Demonstrar que a metodologia de construção proposta pode ser útil como um instrumento de apoio aos cientistas da informação no desenvolvimento de ontologias.
- ✓ Desenvolver uma ontologia sobre os componentes do sangue humano, denominada de HEMONTO, que represente, conceitualmente e formalmente, o conhecimento relativo aos aspectos fisiológicos do sangue, seus hemocomponentes e hemoderivados, além dos processos envolvidos para a obtenção de tais produtos do sangue.
- ✓ Documentar a ontologia do sangue humano por meio de artefatos de representação gráficos, textuais e formais (em alguma linguagem lógica), que possibilite o acesso e o entendimento ao seu conteúdo por parte de especialistas no domínio e cientistas da informação.

1.3 Justificativa

Inserida no contexto de representação, organização e uso da informação através do emprego de ontologias, a presente pesquisa apresenta como argumentos gerais para sua condução como investigação científica, os seguintes:

- O cenário atual na área de engenharia ontológica, que envolve o aumento do número de ontologias desenvolvidas, principalmente no ambiente web, e por consequência a expansão no uso de metodologias para construção de ontologias.
- A necessidade de elaboração de uma metodologia para construção de ontologias que possa detalhar atividades e processos necessários no ciclo de desenvolvimento ontológico, auxiliando os desenvolvedores nos passos que devem ser executados.
- O propósito de colaborar com o aumento da aproximação teórica e prática das pesquisas em Ciência da Informação e o processo de desenvolvimento

de ontologias, buscando incentivar que um número maior de cientistas da informação construa suas próprias ontologias ou reutilizem ontologias existentes em suas atividades na área.

- A necessidade de uma análise e representação ontológica do sangue humano, mais representativa, expressiva e abrangente, do que a presente nas terminologias (SNOMED, UMLS, CID) e ontologias biomédicas (FMA, GALEN) atualmente disponíveis.
- A importância do desenvolvimento de teorias científicas no exercício acadêmico, que, de maneira geral, podem representar um indicativo da maturidade na área.

Sobre a **primeira justificativa** apontada, é possível perceber que desde o uso de ontologias como instrumento de representação do conhecimento a partir dos anos de 1990, em pesquisas da área e em algumas aplicações práticas, uma série de metodologias, métodos e propostas de como conduzir a construção de uma ontologia foram desenvolvidos. Algumas metodologias, tais como a *Methontology* (FERNÁNDEZ *et al.*, 1997) ou o método 101 (NOY e GUINNESS, 2001) e, mais recentemente, a metodologia *NeON* (SUARÉZ-FIGUEROA, 2010), passaram a ser bastante utilizadas pelos desenvolvedores de ontologias. Ainda assim, talvez em função da imaturidade de uma área de pesquisa recente, não é possível identificar uma metodologia unificada ou padrão de construção que seja amplamente aceito no desenvolvimento de ontologias.

Em função disso, Santos (2014) destaca que metodologias próprias para construção de ontologias tem sido desenvolvidas para atender aos requisitos de representação de áreas ou domínios específicos. A ideia por trás da elaboração destas metodologias próprias é que elas tem feito o reaproveitamento e unificação de etapas de metodologias existentes e fazendo adaptações às necessidades da área em que são mais utilizadas. Santos (2014) explica esse fenômeno afirmando que as metodologias de construção atuais não estão sendo usadas isoladamente, na maioria das vezes, são usadas em parceria com outras metodologias como forma de suprir as limitações existentes em cada uma delas. Como, em geral, algumas metodologias são mais detalhadas em algumas etapas que outras, o uso em conjunto de várias metodologias para construção ontológica parece ser a solução mais viável. Na presente pesquisa, a metodologia de construção proposta - *OntoForInfoScience* – combina etapas de três metodologias existentes em uma metodologia única.

Entretanto, apenas o reaproveitamento de etapas das metodologias de construção existentes não é suficiente para solucionar alguns problemas no

desenvolvimento de ontologias. Um aspecto essencial que deve ser considerado nesse processo é o detalhamento das atividades necessárias ao ciclo de desenvolvimento de uma ontologia, o qual corresponde à **segunda justificativa** da presente pesquisa.

Como mencionado anteriormente, as atuais metodologias de construção de ontologias, em geral, não detalham suficientemente os passos que um desenvolvedor de ontologias precisa executar no processo de construção. Tais metodologias afirmam que é necessário, por exemplo, que cada classe e relação da ontologia possua uma definição formal em alguma linguagem lógica, porém, não explicam ao desenvolvedor como ele deve proceder na construção de definições formais. Esse detalhamento é um dos objetivos específicos da metodologia *OntoForInfoScience*.

A explicação detalhada dos processos e atividades envolvidos no ciclo de desenvolvimento de uma ontologia contribui positivamente para que um número maior de desenvolvedores, incluindo os cientistas da informação, construa suas próprias ontologias. Esse argumento faz parte da **terceira justificativa** da presente pesquisa, que pretende colaborar com o aumento da aproximação teórica e prática das pesquisas em Ciência da Informação e o desenvolvimento de ontologias. Porém, para atingir tal propósito, além do detalhamento das atividades, torna-se necessário que a metodologia proposta seja elaborada em uma linguagem de fácil compreensão por pesquisadores e profissionais da área de Ciência da Informação ligados à Organização do Conhecimento, o que envolve a explicação de jargões técnicos e questões lógicas e filosóficas do desenvolvimento de ontologias.

A **quarta justificativa** refere-se ao problema de que mesmo com os esforços já empregados na construção de vocabulários, terminologias e ontologias biomédicas conhecidas, eles ainda são insuficientes para atender domínios biomédicos mais específicos, como o do sangue humano. No caso específico do sangue, as terminologias SNOMED, UMLS e CID e também as ontologias FMA e GALEN descrevem alguns aspectos do sangue, por exemplo, seus componentes básicos e características fisiológicas. Entretanto, tem-se muito pouco sobre os tipos de componentes do sangue usados no tratamento hemoterapêutico de pacientes e, além disso, o conteúdo do sangue nesses instrumentos de representação citados está expresso em uma forma com baixo rigor formal, sem a inclusão de axiomas lógicos importantes para tratar o conhecimento deste domínio. Nesse sentido, a ontologia sobre os hemocomponentes do sangue humano (HEMONTO) consegue englobar termos ou conceitos da área ainda não representados nas terminologias e ontologias existentes e também inclui algumas propriedades e axiomas lógicos para representar os conceitos do domínio mais próximos da realidade.

Por fim, a **quinta justificativa** destaca o fato de que ao se desenvolver uma ontologia sobre o sangue humano, compreendida como uma teoria ontológica e também científica, está se praticando o exercício acadêmico em si, a própria ciência. De fato, modelos e teorias são entidades da prática científica indispensáveis para o avanço da ciência, pois funcionam como instrumentos humanos para alcançar a compreensão aproximada da realidade do mundo.

Essa compreensão de uma ontologia como uma teoria ontológica e científica parte do pressuposto de que “uma ontologia deve estabelecer a verdade sobre a realidade, encontrando respostas para a questão ‘o que existe’ ” (SMITH, 2004). Tal perspectiva vai muito além da definição simplista do termo, comum no contexto de sistemas de informação e de computação, que compreende a ontologia como um artefato de software. Afim de alcançar esse status de teoria ontológica, a ontologia HEMONTO foi desenvolvida sob uma base de fundamentação teórica, especificamente, à base de princípios filosóficos do realismo ontológico, que norteia esse desenvolvimento.

Apresentados tais argumentos, acredita-se que a elaboração da metodologia para construção de ontologias proposta e sua aplicação no desenvolvimento da teoria ontológica do sangue humano configuram-se como atividades de pesquisa que, de alguma forma, contribuem para o desenvolvimento científico da área de Ciência da Informação, enquanto artefatos de representação, organização e uso da informação. Além desse argumento mais geral, há de se considerar o caráter social e humano de tal investigação, ao tratar de um domínio do conhecimento fundamental ao estado de saúde da população em geral.

2 Teorias científicas

Ao sugerir que a ontologia desenvolvida sobre os componentes do sangue humano (HEMONTO) pode ser caracterizada como uma teoria ontológica do sangue, torna-se imprescindível à presente pesquisa apresentar a compreensão e o entendimento do termo “teoria ontológica” no contexto científico.

Em função dessa necessidade apontada, o presente capítulo discorre sobre teorias científicas, destacando seu importante papel na compreensão e explicação dos fenômenos estudados pela Ciência e evidenciando sua correlação com ontologias.

Sob a perspectiva de desenvolvimento de ontologias a partir de uma abordagem realista, de modo que se objetive construir esquemas abstratos da realidade a partir da avaliação de fatos e a representação de sua estrutura real, as ontologias são comparáveis, senão equivalentes, às teorias científicas.

Nessa abordagem, uma ontologia é uma teoria representando os principais fatos e regras que governam certa parte da realidade, ou seja, determinados domínios do conhecimento.

Considerando tais argumentos, a presente pesquisa propõe a elaboração de uma teoria ontológica, que é também científica, como forma de fomentar as necessidades de organização e uso da informação em um domínio específico, neste caso nas áreas de hematologia e hemoterapia. Para tanto, a teoria proposta é construída de forma a assegurar o compromisso ontológico de realizar uma representação aproximada da realidade do domínio sob estudo, tal como realizado na elaboração de teorias científicas.

Afim de clarificar essa abordagem adotada, optou-se por estruturar o presente capítulo nas seguintes seções: na seção 2.1, apresentam-se definições sobre o conceito de teoria científica identificados na literatura da área; na seção 2.2, é feita uma breve revisão histórica sobre o papel das teorias no âmbito científico; e, na seção 2.3, apresenta-se uma visão histórica das teorias e dos modelos científicos mais relevantes para aquisição e representação do conhecimento no contexto da Ciência da Informação, tendo como marco inicial as teorias sobre terminologia da década de 1930 e como marco final as teorias ontológicas para organização e uso da informação de domínios diversos.

2.1 O que são teorias?

A definição do que venha ser teoria e seu papel na aquisição do conhecimento científico não é consensual e gera discussões filosóficas sobre esse tema. Para melhor entender o que são teorias, apresentam-se, nesta seção, três concepções básicas de

teorias científicas, de acordo com Reynolds (2007): (i) a *teoria na forma de um conjunto de leis*, (ii) a *teoria na forma axiomática* e (iii) a *teoria na forma de processo causal*.

Na primeira concepção – a *teoria na forma de um conjunto de leis* – aceita-se como teoria científica apenas aquela que inclua declarações que podem ser consideradas leis. Leis são conjuntos de generalizações empíricas bem fundamentadas. Como leis são diretamente suportadas por pesquisa empírica, todos os conceitos usados em leis devem ter definições que permitam sua identificação em situações concretas.

A segunda concepção – a *teoria na forma axiomática* – consiste de um conjunto inter-relacionado de *definições* (incluindo conceitos teóricos) e de *declarações* de dois tipos: (i) *existenciais*, que fazem afirmações sobre a existência e descrevem situações nas quais a teoria pode ser aplicada; e (ii) *relacionais*, que descrevem o relacionamento entre dois conceitos. As declarações relacionais englobam os *axiomas*, conjuntos de declarações das quais outras declarações na teoria podem ser derivadas, e as *proposições*, que são todas as outras declarações na teoria, derivadas de combinações de axiomas e proposições.

Essa concepção axiomática de teorias é, geralmente, conhecida como *visão sintática* (CARNAP, 1938; HEMPEL, 1965) e, como produto da lógica positivista, considera que para termos uma teoria, ela deve ser enunciada tendo comobase formalismos, que incluem os axiomas e as proposições, do contrário, não teremos uma teoria científica.

Essa abordagem a teorias foi adotada por número significativo de pesquisadores, como por exemplo, Carnap (1938), Reichenbach (1938), Popper (1963), dentre outros, que buscavam definir teorias científicas a partir da lógica. Tal corrente filosófica definia teorias científicas como teorias axiomáticas formuladas na linguagem da lógica matemática L, satisfazendo às condições de: (i) *logicismo clássico*: teoria formulada na lógica de primeira ordem; (ii) *distinção teórico-observacional*: um vocabulário lógico (constantes lógicas), um vocabulário observacional V_0 (termos de observação) e vocabulário teórico V_t (termos teóricos); (iii) *fisicalismo*: os termos de V_0 interpretados como se referindo a objetos físicos ou atributos destes; (iv) *leis teóricas*: conjunto de postulados teóricos T; e (v) *regras de correspondência*, que especificam as aplicações admissíveis de T a fenômenos empíricos.

A terceira e última concepção de teoria – a *teoria na forma de processo causal* – faz parte da *visão semântica* de teorias (VAN FRAASSEN, 1980; GIÉRE, 1988; SUPPES, 2002) e, além de fornecer explicação e predição científicas, como as duas anteriores, também provê o senso de entendimento.

Diferentemente da visão sintática, a visão semântica de teorias (VAN FRAASSEN, 1980; GIÉRE, 1988; SUPPES, 2002), dispensa a perspectiva formal e

enxerga teoria como uma família de modelos científicos. Segundo essa abordagem, as teorias científicas são entidades extra-linguísticas que não podem ser identificadas apenas através de suas formulações linguísticas, tais como os formalismos lógicos.

A teoria na forma de processo causal é um conjunto inter-relacionado de *definições* e *declarações* de dois tipos: (i) *existenciais*, que descrevem situações nas quais se espera que um ou mais processos causais ocorram; e (ii) *causais*, que descrevem um ou mais processos causais ou um mecanismo causal que identifica o efeito de variáveis independentes sobre variáveis dependentes.

A principal diferença entre a teoria na forma causal e a teoria na forma axiomática é que na primeira todas as declarações são consideradas de igual importância, não sendo classificadas em axiomas e proposições. Um conjunto de declarações na forma axiomática pode ser reorganizado em uma descrição de um processo causal (REYNOLDS, 2007).

2.2 Teorias no âmbito da Ciência

A abordagem a teorias científicas é discutida, principalmente, no âmbito da filosofia da ciência, que busca compreender e explicar os avanços científicos, além de mostrar o melhor caminho para que a ciência possa evoluir. Nesta seção, adota-se uma visão histórica para descrever o papel recente das teorias na ciência.

Nosso breve relato histórico inicia-se no século XIX, no qual dois posicionamentos filosóficos distintos se destacavam mais na explicação sobre o conhecimento científico: de um lado os adeptos de Hegel e seus sucessores *neo-hegelianos*, que procuravam explicar a realidade em termos de entidades metafísicas abstratas que não podiam ser especificadas empiricamente, como o chamado Absoluto; e de outro lado, o movimento do *positivismo lógico*, que defendia o empirismo e combatia os excessos metafísicos de Hegel e seus sucessores (SUPPE et al., 1977).

Ainda segundo este mesmo autor, dentre as variações existentes no positivismo lógico, destacava-se o *materialismo mecanicista*, que era uma mistura do positivismo comtista (Auguste Comte), do materialismo e do mecanicismo. Seu principal porta-voz era o alemão Ludwig Büchner, que rejeitava o idealismo e o “super-naturalismo”, em defesa de um método científico que forneceria conhecimento objetivo à luz da investigação empírica e sem recurso à especulação filosófica.

Mesmo depois de tanto tempo, o mecanicismo ainda exerce suas influências na forma de pensar em alguns domínios científicos, entre eles o domínio médico. Rios et. al (2007) destaca a influência do paradigma cartesiano, reducionista e mecanicista sobre o pensamento médico. Tais autores citam o fato de que o corpo humano é considerado uma

máquina que pode ser analisada peça a peça; a doença é vista como um mau funcionamento dos mecanismos biológicos e o papel dos profissionais de saúde é intervir, física ou quimicamente, para consertar o defeito no funcionamento de um específico mecanismo enguiçado. Rios et. al (2007) acrescenta ainda que a medicina moderna mantém esse paradigma dominante, porém centrado numa abordagem hospitalocêntrica, curativista e verticalizada.

O final do século XIX registrou uma crise do materialismo mecanicista, que passou a ser questionado como resultado dos avanços na psicologia e na fisiologia. Hermann Von Helmholtz (1863) apud Suppe et al. (1977) salientou a importância da mediação dos sentidos e da atividade pensante do sujeito no crescimento do conhecimento científico. Surgiu então, a *filosofia da ciência neo-kantiana*, que compreendia o objetivo da ciência como o de descobrir as estruturas ou formas gerais das sensações, que constituem teias de relações lógicas entre sensações (COHEN 1871 apud EDGAR, 2011).

Na visão neo-kantiana, a ciência descobriria as estruturas dos fenômenos e não das coisas em si. Nesse contexto, as teorias científicas passam a ter um papel fundamental e, na posição neo-kantiana, todas elas devem conter um elemento a priori de caráter puramente formal.

E foram as próprias teorias científicas mais importantes da época – a teoria da relatividade e a teoria quântica – que fizeram emergir uma crise nas filosofias da ciência daquela época, devido principalmente a duas questões: (i) Como incorporar as revoluções da física? (ii) Qual é a natureza da investigação científica? Nenhuma das três escolas de filosofia da ciência da época – o *materialismo mecanicista* de Büchner, o *neo-kantismo* de Cohen e o *neo-positivismo machiano* de Ernst Mach – não conseguiam dar conta dos novos avanços científicos (SUPPE et al., 1977).

Diante deste cenário, no início do século XX, Carnap e Schlick estabeleceram um influente grupo de discussão sobre o conhecimento e os avanços científicos, fundamentalmente, baseados no positivismo lógico, que ficou conhecido como *Círculo de Viena* (PESSOA JR., 1993) – fato marcante na época. Esse grupo de filósofos, que se reuniu até 1936 – ano do assassinato de Schlick -, fundou as bases para a criação da *visão ortodoxa (ou visão recebida) de teorias*, que buscava explicar as teorias científicas a partir da lógica e de uma postura empirista e positivista. A *visão ortodoxa* consolidou-se como uma filosofia ciência na década de 50, quando alguns de seus adeptos empiristas (Carnap, Hempel, Feigl, Reichenbach) se instalaram nos Estados Unidos, conforme destacado por esse mesmo autor.

Essa consolidação não significou consenso entre os filósofos da época. Duras críticas foram desferidas contra a visão ortodoxa, como, por exemplo, por Karl Popper. Popper (2008) rejeita o critério verificacionista de significância cognitiva, defendendo a idéia de que teorias científicas não podem ser verificadas por qualquer acúmulo de evidência observacional. Ainda, sustenta que o problema central da filosofia da ciência é o crescimento do conhecimento científico, e isso não poderia ser reduzido a um estudo de linguagens artificiais, como sugerido pela visão ortodoxa.

No estudo das teorias científicas, Popper é conhecido por seu critério da falseabilidade e por advogar o *falibilismo* (POPPER, 1963). De acordo com o critério da falseabilidade, hipóteses científicas são falsificáveis e, dessa forma, os cientistas estão aptos a estabelecer quais achados empíricos tornam as hipóteses falsas. O falibilismo é a visão de que nenhum conhecimento presumido, nem mesmo conhecimento científico, é absolutamente certo. Como o conhecimento humano é incompleto, provável e conjectural, o cientista deve buscar a verdade, mas esperar *verossimilhança*. Verossimilhança é uma medida qualitativa de como uma teoria pode ser mais ou menos próxima da verdade (BHASKAR, 1978).

O escopo de críticas à visão ortodoxa não se restringe aos enunciados de Popper e é bastante amplo (SUPPE et al., 1977). Apesar de haver uma discordância sobre quais dos aspectos de tal visão são inadequados, esse mesmo autor afirma que de fato há um consenso entre os filósofos da ciência que a visão ortodoxa é inadequada para explicar o avanço das teorias científicas, mesmo em sua versão final desenvolvida por Carnap e Hempel.

Uma abordagem alternativa à visão ortodoxa surgiu no final da década de 50 e viria a se constituir em uma nova filosofia da ciência, que não se limitava apenas aos aspectos lógicos da ciência (PESSOA JR., 1993). Essa abordagem ficou conhecida como *teoria dos globalistas*, da qual faziam parte Willard Quine, Paul Feyerabend, Norwood Hanson, Stephen Toulmin e Thomas Kuhn. Alguns pontos ressaltados pelos globalistas (BACHELARD, 1934; KUHN, 1962) são: (i) rejeita-se que uma teoria científica se assenta em bases sólidas fornecidas por dados observacionais; (ii) o contexto social e histórico tornam-se relevantes para entender porque uma teoria é preferida em relação a outra; (iii) a transição de uma teoria para outra não é mais vista como uma ampliação cumulativa de conhecimento, mas como uma ruptura.

Nos anos 60, Thomas Kuhn é o principal ícone dentre os teóricos globalistas e sua obra *A Estrutura das Revoluções Científicas* ressalta o fato de que a transição entre teorias se dá através de revoluções, e que entre estes períodos de transição tem-se a chamada *ciência normal*.

Kuhn (1962) advoga que a evolução científica não emerge diretamente da acumulação de fatos, mas através de “revoluções” surgidas de um conjunto de circunstâncias e possibilidades desafiadoras. Tais circunstâncias geram um novo *paradigma*, caracterizado por: uma nova e radical conceitualização dos fenômenos, a necessidade de uma nova estratégia de pesquisa para obtenção de evidências de suporte às novas teorias, o surgimento de questões e problemas que as teorias anteriores não estão aptas a explicar.

Na abordagem kuhniana, os períodos de ciência não-revolucionária, chamados de *ciência normal*, são conduzidos por comunidades científicas que compartilham uma *matriz disciplinar*, que sendo compreendidas como um tipo de paradigma, não podem ser explicitadas de maneira completa, mas adquiridas de forma implícita no processo educacional, especialmente através da assimilação de exemplares.

Nas décadas de 70 e 80, outros autores como Lakatos (1970), Feyerabend (1975), Van Fraassen (1980) e Suppes (2002), apresentam suas metateorias debatendo as idéias de Kuhn e dos globalistas. Lakatos (1970), por exemplo, propõe sua metateoria com base numa versão do falseacionismo de Popper unificada com as idéias de Kuhn, enquanto Van Fraassen (1980); Giere (1988); Suppes (2002) criam suas metateorias tendo como base a *visão semântica de teorias*, que enxerga as teorias científicas como entidades extra-linguísticas, muito além de um conjunto de formalismos.

Por fim, o panorama da filosofia da ciência, nas décadas de 80 e 90 aos dias atuais, caracteriza-se por uma pluralidade de abordagens, enraizadas tanto nos globalistas quanto na visão ortodoxa (PESSOA JR., 1993). Nesse cenário, destacam-se: (i) os seguidores da visão semântica, especialmente popular na Alemanha; (ii) a chamada *epistemologia evolucionária* (Campbell, 1960; Hull, 1988), que procura entender a dinâmica científica a partir da teoria da evolução em biologia¹; e (iii) o embate entre o realismo científico e o anti-realismo provocado, principalmente, pela obra de autores anti-realistas, tais como Van Fraassen (1980).

Sobre este último item citado, é importante ressaltar a diferença central nos argumentos apresentados por realistas e anti-realistas: o realismo científico (POPPER, 1963; BUNGE, 1974; CHURCHLAND, 1985; SMITH e CEUSTERS, 2010) defende o argumento de que as teorias e os modelos científicos possuam valor ontológico, isto é, eles devem ser construídos em direção à verdade, buscando captar, numa maior

¹ Segundo a teoria da evolução em biologia (Campbell, 1960; Hull, 1988), grupos de pesquisa seriam análogos às espécies, e as idéias científicas funcionariam como traços genéticos, passando por um processo de seleção ao serem compartilhados por outros grupos de pesquisa.

aproximação possível, “o que realmente existe” no mundo; por outro lado, o anti-realismo (CARTWRIGHT, 1983; NEWTON-SMITH, 1985; VAN FRAASSEN, 1980; PUTNAM, 1992) não julga como necessário o comprometimento com entidades sob o ponto de vista ontológico. Sob essa perspectiva, as hipóteses acerca do mundo são apenas construções mentais, que se impõem não por seu caráter referencial, mas em função de sua capacidade explicativa (SILVA, 1998), não sendo a verdade o objetivo central das teorias científicas.

Apesar de adotarem visões antagônicas, realistas e anti-realistas concordam em um ponto sobre a questão da explicação científica. Para ambos, as teorias e os modelos “dão conta” dos fenômenos do mundo, já que eles são capazes de explicar de maneira bem sucedida tais fenômenos. Nesse sentido, modelos e teorias são além de instrumentos de descrição, instrumentos de explicação científica e, assim, fundamentais para aquisição do conhecimento científico acerca do mundo a nossa volta.

2.3 Das teorias terminológicas às teorias ontológicas

Ao longo de sua história, a Ciência da Informação, pela sua própria natureza ampla e interdisciplinar, teve obrigatoriamente de utilizar-se de teorias e modelos de outras áreas para mapear toda a sua realidade (SAYÃO, 2001). Modelos das áreas de Comunicação, Ciência da Computação, Linguística, Economia e Marketing foram tomados emprestados pela Ciência da Informação ao longo de sua trajetória científica.

Nesta seção, são citados exemplos de teorias e modelos usados pela Ciência da Informação na aquisição e representação do conhecimento de domínios diversos e, desta forma, possibilitar a organização e uso da informação em tais domínios. A estruturação da presente seção ocorre por meio de uma perspectiva histórica, que se inicia com as teorias sobre terminologia, na década de 1930, e se encerra com a proposta de teorias ontológicas para a organização da informação nos dias de hoje.

Antes de apresentar um breve histórico sobre teorias e modelos no âmbito da Ciência da Informação é importante esclarecer, de forma resumida, o uso dos termos “modelo” e “teoria” na literatura científica, tal como explicitado em Mendonça e Almeida (2012).

Segundo esses autores, a palavra “teoria” é utilizada num sentido mais abrangente do que os modelos propriamente ditos, afim de indicar a ampla utilização e aplicabilidade de um determinado modelo. Ainda sim, os dois termos muitas vezes são tratados como sinônimos, embora existam diferenças entre eles.

Afim de esclarecer as diferenças entre esses termos, Mendonça e Almeida (2012) definem-os da seguinte forma: *modelos* são representações de uma parte

específica da realidade e também interpretações ou realizações de uma determinada teoria e *teorias* consistem de conjunto de declarações logicamente organizadas com poder explanatório sobre eventos genéricos em um dado domínio de conhecimento.

Retornando, rapidamente, aos conceitos de filosofia da ciência temos duas perspectivas diferentes sobre a relação modelo-teoria: (i) a *visão ortodoxa ou sintática* (CARNAP, 1938; HEMPEL, 1965), considera que um modelo científico é uma interpretação alternativa de certa teoria e, assim, é visto como irrelevante para a ciência; e (ii) a *visão semântica* (VAN FRAASSEN, 1980; GIERE, 1988; SUPPES, 2002) reconhece os modelos como componentes essenciais e integrais das teorias, possibilitando testá-las e fornecer suas explicações.

Apesar de existirem essas duas diferentes visões sobre o valor dos modelos na ciência, geralmente se aceita que modelos são unidades centrais na criação de teorias científicas e, portanto, são bastante úteis para aquisição do conhecimento (LE MOIGNE, 1977).

Realizados os esclarecimentos acima, passa-se à revisão histórica, que na presente pesquisa se inicia nos anos 30. A partir da década de 1930 surgiram as teorias sobre terminologia mais conhecidas em Ciência da Informação: (i) a *teoria da classificação facetada* (RANGANATHAN, 1967); (ii) a *teoria geral da terminologia* (WÜESTER, 1979); e (iii) a *teoria do conceito* (DAHLBERG, 1978).

A *teoria da classificação facetada*, desenvolvida por Ranganathan a partir da *Colon Classification*², se diferenciou dos demais esquemas de classificação bibliográfica da época pelo fato de não trabalhar com categorias pré-estabelecidas (CAMPOS, 2001). No esquema facetado de Ranganathan, a classificação dos livros era criada somente no momento que um livro fosse analisado de acordo com os elementos constituintes do assunto. A idéia deste pesquisador representou uma revolução do pensamento sobre classificação bibliográfica da época.

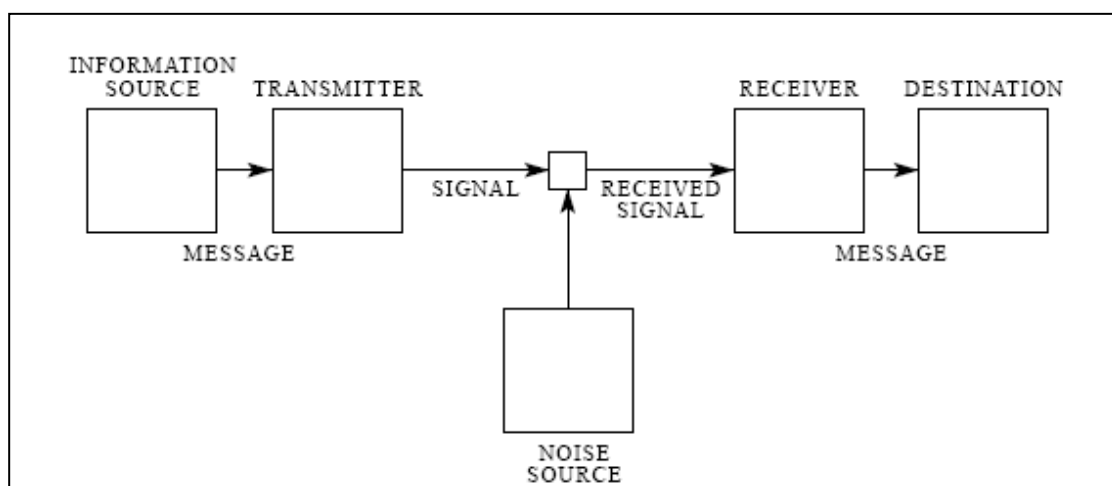
Por sua vez, a *teoria geral da terminologia*, desenvolvida pelo engenheiro Wüester, teve como objetivo inicial garantir comunicação precisa na área de Eletrotécnica, mas depois foi bastante usada pela Ciência da Informação na organização da informação. Já a *teoria do conceito*, desenvolvida por Dahlberg, possibilitou uma base mais sólida para a determinação e o entendimento do que consideramos como conceito, para fins de representação e recuperação da informação (CAMPOS, 2001).

² Colon Classification é uma tabela de classificação elaborada para a organização do acervo da Biblioteca da Universidade de Masdras, na Índia (CAMPOS, 2001).

Sobre essas teorias terminológicas citadas é importante ressaltar que elas foram utilizadas como base metodológica para a criação dos esquemas de classificação bibliográfica, surgidos ao longo do século XIX, para descrever o estado do conhecimento da época. A Classificação Decimal de Dewey (CDD), a Classificação Decimal Universal (CDU) e a *Library of Congress Classification* são alguns exemplos de esquemas de classificação bibliográfica, até hoje utilizados nas bibliotecas do mundo inteiro, que se fundamentam em tais teorias.

Já na década de 40, os pesquisadores americanos Shannon e Weaver propuseram um modelo matemático para explicar a comunicação humana entre dois pólos, denominados de *emissor* e *receptor* (SHANNON e WEAVER, 1949). Tal modelo conhecido como *teoria da comunicação* (veja Figura 1) representou a fundamentação do que rapidamente se transformou na *teoria da informação*, como destaca Eco (1997).

Figura 1 - Diagrama esquemático de um sistema geral de comunicação.



Fonte: adaptado de Shannon (1948).

Sobre o processo de comunicação de Shannon (1948), apresentado na Figura 1, Eco (1997) afirma que os instrumentos fornecidos pela *teoria da comunicação* são úteis “não porque os fenômenos de comunicação mais complexos sejam redutíveis à passagem de um sinal de uma máquina a outra, mas porque é útil individuar a relação comunicacional, na sua dinâmica essencial, sugerindo-nos a construção de um modelo exemplar.”

No contexto da Ciência da Informação a *teoria da comunicação* de Shannon e Weaver é um dos modelos de maior sucesso e ampla utilização (SAYÃO, 2001). “Tal modelo, criticado, adaptado, modificado, ainda hoje está sendo amplamente utilizado, na medida em que, de modo preciso, simples e preditivo, propicia uma boa idéia de como se

dá a comunicação humana. É, em essência, um modelo matemático, da mesma forma que as leis de Zipf, Bradford, Ortega, 80/20 e outras amplamente utilizadas na bibliometria, mas é também, na sua concepção geral, um modelo sistêmico interligando o emissor ao receptor” (SOUZA, MENDONÇA e DODEBEI, 1993 apud SAYÃO, 2001).

Analisando o papel das teorias até então apresentadas para a organização e uso da informação pode-se afirmar que: as teorias sobre terminologia, usadas nos esquemas de classificação bibliográfica, têm por objetivo possibilitar à organização dos materiais informacionais em bibliotecas ou centros de informação, de modo a serem úteis na organização, localização e consulta de conteúdo (SATIJA, 2000); enquanto a teoria da comunicação está preocupada em apresentar uma forma adequada para a transmissão da informação entre seus usuários (emissor e receptor).

Considerando o objetivo de tais teorias, tornou-se necessário à Ciência da Informação, ao longo de sua trajetória científica, recorrer a outros modelos e teorias para desempenhar um outro papel que lhe é atribuído: a representação da informação de domínios diversos. Para essa tarefa, a Ciência da Informação tomou emprestado teorias e modelos de outras áreas, como, por exemplo, os *modelos de dados* oriundos da Ciência da Computação.

Modelos de dados são criados nas organizações com o objetivo de representar o que deve ser codificado e processado em sistemas de informação. Os sistemas de informação organizacionais têm papel relevante na consolidação de práticas administrativas relacionadas às necessidades das pessoas no desempenho de suas funções. O desenvolvimento de sistemas de informação envolve a criação de modelos para representar atividades que tem lugar nos mais diversos tipos de organizações. Um modelo de dados organizacionais é uma representação explícita da estrutura, atividades, processos, fluxos, metas e restrições de uma organização (GANDON, 2002).

Nesse sentido, pode-se afirmar que o desenvolvimento de sistemas de informação é uma aplicação evidente de modelos e teorias em um nível mais específico: primeiro é preciso entender os fenômenos de interesse, para depois representá-los de uma forma que se possa tirar vantagem de seu caráter explicativo.

No desenvolvimento de sistemas de informação, *modelagem conceitual* é o estágio em que os modelos são criados para representar a compreensão humana. *Modelos conceituais* são criados a partir de abstrações da realidade de acordo com a perspectiva de um indivíduo ou grupo de indivíduos. Abstrações são meios de especificar entidades e as relações entre entidades dentro do domínio de um campo do conhecimento que é de interesse para os usuários do sistema.

As primeiras iniciativas para a especificação de modelos de dados datam do final dos anos 1950 (BOSAK et al., 1962; YOUNG e KENT, 1958). Tais iniciativas foram conduzidas de forma a criar modelos que atendessem aos requisitos de estruturas de dados computacionais.

Nos anos de 1960, a pesquisa em bancos de dados gerou três tipos principais de modelos de dados: o *modelo hierárquico*, o *modelo em rede* e o *modelo relacional*. Esses modelos são conhecidos como modelos lógicos, uma vez que eles não se referem a aspectos físicos da implementação do sistema. Entretanto, modelos lógicos causaram muitos problemas o que acabou limitando sua utilização na modelagem conceitual (MYLOPOULOS, 1998).

Os primeiros *modelos semânticos* usados em modelagem conceitual apareceram nos anos 1970 no contexto do trabalho do comitê ANSI/X3/SPARC para padronização dos sistemas de gerenciamento de bancos de dados. Dentre estes, os mais conhecidos são o *modelo de dados semântico* (ABRIAL, 1974), a *arquitetura em três camadas* (JARDINE, 1976), o *modelo entidade relacionamento* (CODD, 1979), dentre outros. As principais características dos modelos semânticos, em comparação com os anteriores, é a facilidade de entendimento. O modelo entidade relacionamento, por exemplo, fornece termos adicionais para uso como primitivas de modelagem.

Nos anos 1990, propostas para modelagem orientada a objetos se tornaram populares. Os *modelos orientados a objetos* têm características adicionais em relação aos modelos de dados, mas também mantém similaridades (MILTON, 2000). A Linguagem de Modelagem Conceitual (popularmente conhecida como UML) foi uma tentativa de padronizar notações orientadas à objeto que reuniu outras iniciativas: o método de Booch (BOOCH, 1993), a técnica de modelagem de objetos (RUMBAUGH et al, 1991), a engenharia de software orientada a objetos (JACOBSON et al, 1992), dentre outras.

Ao longo dos anos a criação de modelos conceituais tem sido motivada pela busca por melhores formas de representar a realidade. De fato, a modelagem conceitual é a atividade de descrever formalmente alguns aspectos do mundo físico e social com propósito de compreensão e comunicação (MYLOPOULOS, 1992). De acordo com Smith e Welty (2001), a inconsistência na modelagem durante os primeiros anos da modelagem conceitual tem sido a causa dos atuais problemas de interoperabilidade entre sistemas. Sobre tais problemas, Fonseca e Martin (2007) acrescentam o fato de que a maioria dos esquemas conceituais, resultantes da atual prática de modelagem conceitual, são construídos para fins específicos de um dado sistema de informação em desenvolvimento. Assim, o propósito prático de um esquema conceitual é definir, restringir e limitar o que

deve ser registrado e manipulado por esse sistema de informação, sem manter o compromisso com o que de fato exista na realidade de um dado domínio.

Outro problema relacionado aos sistemas de informação atuais é que grande parte do conhecimento sobre o domínio modelado e implementado encontra-se implicitamente armazenado, algumas vezes de forma obscura, nas linguagens de codificação dos programas de aplicação destes sistemas (GUARINO, 1998). Assim, essa parte do conhecimento, armazenada de forma implícita, é perdida, por não ser possível decifrá-la (torná-la informação explícita).

Como alternativa para solucionar esses problemas da modelagem conceitual, muitos pesquisadores defendem e argumentam em favor dos *modelos baseados em ontologias*. Alguns desses pesquisadores – Guarino (1998); Smith (2003); Wand e Weber (2004); Fonseca e Martin (2007) - apontam que o uso de ontologias no campo de modelagem representa uma evolução na área e uma alternativa para solucionar os problemas gerados pela prática atual da modelagem de sistemas de informação. As ontologias ajudam tornar mais explícito o conhecimento adquirido de um certo domínio, promovem o compartilhamento do conhecimento e possibilitam a integração da informação entre diferentes instrumentos de representação, tais como os sistemas de informação.

Nos *modelos baseados em ontologias*, a ontologia equivale-se a uma teoria usada para representar os principais fatos e regras que governam certa parte da realidade, porque ao se criar ontologias está se avaliando fatos reais, identificando a estrutura sobre a qual se organizam, fazendo generalizações e abstrações. Por isso, como já citado, a presente pesquisa defende o argumento de que ontologias fundamentadas na realidade correspondem às teorias científicas.

3 Organização da informação

Tratar os problemas relativos à organização da informação nas mais diversas áreas do conhecimento é uma tarefa, na maioria das vezes, complexa e dispendiosa, que, normalmente, exige conhecimentos de diversas áreas e instrumentos de suporte para tratar tais problemas.

Uma das atividades necessárias para proceder com a organização da informação em um determinado domínio do conhecimento é buscar uma forma adequada de representá-la. Quando falamos de representação da informação e de conhecimento, também estamos tratando de um processo complexo, já que envolve, necessariamente, a reprodução de conceitos e significados construídos pela mente humana em algum suporte, como em um dispositivo eletrônico ou no papel, por meio de símbolos e linguagens.

Pesquisadores de diferentes áreas do conhecimento corroboram com este pensamento e conduzem suas pesquisas sobre este tema. Mendes (1998, p. 2), por exemplo, afirma que “capturar o conhecimento humano e torná-lo explícito não é uma tarefa simples”. Campos (2004, p. 24) acrescenta o fato de que a representação perfeita de um determinado objeto é, em geral, impossível. A única representação completamente precisa de um objeto é o objeto em si mesmo.

Diante do exposto, torna-se imprescindível para a representação da informação buscar a maior aproximação possível entre o símbolo representante do objeto e o próprio objeto no mundo real. Essa correspondência é, comumente, chamada de semântica da representação.

No âmbito da Ciência da Informação, os problemas de representação surgem à medida que o volume de documentos a processar e ordenar cresce ao ponto dos usuários não mais se contentarem com sua organização em grandes categorias, exigindo informações mais precisas, como destaca Robredo (1986). Atualmente, esses problemas se agravaram, estando em níveis ainda mais críticos, devido à explosão informacional que se sucedeu após o advento das novas tecnologias da informação e comunicação, especialmente a *web*. Hoje, em sua grande maioria, as informações encontram-se distribuídas aleatoriamente em diferentes meios eletrônicos.

Para buscar soluções para os problemas de representação e organização da informação, a Ciência da Informação desenvolveu, ao longo de sua trajetória, técnicas variadas que permitissem às pessoas buscar, encontrar e classificar a informação. Dentre essas técnicas, merecem destaque os instrumentos terminológicos de organização e representação da informação, que se baseiam na categorização dos objetos e “coisas” da realidade.

Neste capítulo da tese, o enfoque será exatamente abordar esses instrumentos de organização da informação comumente utilizados em Ciência da Informação para realizar tal tarefa. Além de uma breve revisão geral dos instrumentos usados na Ciência da Informação, discorre-se também sobre aqueles voltados para a área biomédica, tais como terminologias e ontologias biomédicas, uma vez que o domínio de estudo desta pesquisa refere-se às áreas de hematologia e hemoterapia.

Nesse sentido, optou-se por dividir o presente capítulo em duas partes: na seção 3.1, são apresentados alguns dos instrumentos de organização da informação mais relevantes no âmbito da Ciência da Informação, incluindo vocabulários controlados, terminologias e ontologias; e, na seção 3.2, são destacados, dentre esses instrumentos, aqueles comumente usados na área médica, que englobam as terminologias e ontologias biomédicas.

3.1 Instrumentos de organização da informação

A organização e representação do conhecimento humano sempre foram objetos de estudo para as pesquisas em Ciência da Informação, buscando encontrar soluções para disseminar o conhecimento entre as pessoas.

Argumentando sobre o papel da Ciência da Informação na tarefa de possibilitar ao usuário acessar o conhecimento requerido, Saracevic (1996, p.47) define a Ciência da Informação como “um campo dedicado às pesquisas científicas e à prática profissional voltadas para os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos”.

A Ciência da Informação desenvolveu, ao longo desses anos, um arcabouço teórico e metodológico para atingir esse objetivo, que incluem, por exemplo: (i) a Classificação Decimal de Dewey (CDD) e a Classificação Decimal Universal (CDU) para a organização dos grandes repertórios bibliográficos em bibliotecas do mundo inteiro; (ii) obras como as de Ranganathan (1967), Dahlberg (1978) e do *Classification Research Group* (CRG) (WILSON, 1972 apud CAMPOS, 2001). Essas obras citadas, por exemplo, influenciaram e permitiram também a criação de instrumentos de controle terminológico, tais como: vocabulários controlados, tesouros e, mais recentemente, as ontologias.

É claro que muita coisa mudou desde a época dos primeiros estudos sobre a organização do conhecimento nos repertórios bibliográficos, até os dias de hoje, cujo desafio principal é encontrar soluções para organizar o grande volume de informação disponível em diferentes meios eletrônicos. No entanto, esse “olhar para o passado” ajuda a compreender os objetivos pretendidos com o uso dos atuais instrumentos de controle terminológico para organização e representação da informação.

Nas subseções seguintes (de 3.1.1 à 3.1.3), são descritos os instrumentos de organização da informação que serviram como referencial teórico para a presente pesquisa. Serão abordados os vocabulários controlados, incluindo os tesouros e as taxonomias, além das teorias terminológicas e das ontologias.

3.1.1 Vocabulários controlados

Segundo definição de Lancaster (1993), **vocabulário controlado** é uma lista de termos autorizados, em que o indexador apenas pode atribuir a um documento termos que existam na lista adotada pela unidade de informação em que trabalha. São funções do vocabulário controlado: (i) o controle de sinônimos, através de um termo padrão; (ii) a diferenciação entre os homógrafos³; e (iii) o agrupamento de termos em que os significados apresentem um relação mais estreita entre si, como, por exemplo, relações hierárquicas ou não hierárquicas.

No âmbito da Ciência da Informação, é possível identificar alguns instrumentos que possibilitam a criação de vocabulários controlados categorizados, que incluem uma análise semântica das palavras. Alguns desses instrumentos são: os **esquemas de classificação bibliográfica**, as **listas de cabeçalhos de assunto**, os **tesouros** e as **taxonomias**. Tais instrumentos podem ser vistos como diferentes tipos de vocabulários controlados, como destaca Lancaster (1993).

De maneira geral, o propósito principal de todos esses tipos de vocabulários controlados é fornecer um significado para a organização da informação, que inclua tradução, consistência, relacionamentos, visualização e recuperação perante à informação, de acordo com a ANSI/NISO (2005).

Os **esquemas de classificação bibliográfica** foram criados com o propósito de descrever o estado do conhecimento do século XIX. Satija (2000) afirma que seus objetivos remetem-se à organização dos materiais informacionais em bibliotecas ou centros de informação, de modo a serem úteis na organização, localização e consulta de conteúdo. A Classificação Decimal de Dewey (CDD), a Classificação Decimal Universal (CDU) e a *Library of Congress Classification* são alguns dos esquemas de classificação bibliográfica mais conhecidos e utilizados, até hoje, nas bibliotecas do mundo inteiro.

Enquanto os esquemas de classificação foram ou são usados para organizar os livros nas estantes das bibliotecas, as **listas de cabeçalhos de assunto** foram utilizadas, a partir da segunda metade do século XIX, para organizar as informações dos

³ Vocabulário que tem a mesma grafia de outro, porém sentido diferente (ANSI/NISO, 2005).

documentos e do catálogo alfabético de assunto, devido à existência das fichas ou catálogos impressos nesta época, como afirmam Campos et al. (2002).

Os cabeçalhos de assunto são formados de uma palavra ou um conjunto de palavras que representam o assunto do documento. Como os elementos que integram estes cabeçalhos adquirem significado quando formam um contexto - uso da linguagem natural -, Campos et al. (2002) afirmam que estes instrumentos não são representações eficazes para atender às necessidades do usuário em busca da informação. Em virtude disso, surge a necessidade de utilização de instrumentos que tratem como unidade de representação temática de um documento não mais o assunto e sim a reunião de conceitos em um sistema pós-coordenado, cujos termos são combinados no momento de seu uso.

Os **tesauros** surgem, então, como instrumentos terminológicos para a organização dos conteúdos dos documentos, de forma que os termos estejam organizados de acordo com as palavras que os compõem. Segundo Dodebei (2002), o tesouro documentário começou a ser usado na Ciência da Informação a partir de 1940 com o propósito de recuperar informação em ambientes com uma diversidade de documentos especializados.

Em Unisist (UNESCO, 1973, p. 6) tesouro é definido “como um vocabulário controlado dinâmico de termos relacionados semântica e genericamente, cobrindo um domínio específico do conhecimento”. Complementar a essa definição, é aquela encontrada no *National Information Standards Organization* (ANSI/NISO 1993) que conceitua tesouro como “um vocabulário organizado em uma ordem conhecida, na qual as relações de equivalência, hierárquicas e associativas entre os termos são claramente mostradas e identificadas através de indicadores de relação padrão”.

De fato, uma das mais importantes contribuições dos tesauros para a organização e representação da informação é introduzir tipos de relações entre os termos de um dado domínio. A partir disso, outros instrumentos para organização da informação, como as ontologias, especificam e estendem tipos de relações entre os termos de um domínio. No caso dos tesauros, pode-se identificar três tipos de relações: (i) as hierárquicas, que englobam as relações gênero/espécie, todo/parte e enumerativa; (ii) as associativas; e (iii) as relações de equivalência. Exemplos e maiores detalhes sobre tais relações podem ser encontrados em Cintra et al. (2002).

Por fim, destacamos as **taxonomias** como outro tipo de vocabulário controlado. A norma ANSI/NISO (2005) define taxonomias como “coleções de termos classificados em uma estrutura hierárquica, na qual se empregam apenas relacionamentos de generalização e especialização.”

No âmbito da Ciência da Informação, as taxonomias têm sido utilizadas em aplicações que organizam a informação em ambientes eletrônicos e virtuais, como os portais corporativos, as bibliotecas digitais e outras aplicações *web*.

3.1.2 Terminologias

Foram apresentados na subseção anterior (3.1.1), os instrumentos mais relevantes para organização e representação da informação de um dado domínio. Estes instrumentos são chamados, na literatura, de linguagens documentárias e são utilizados no contexto dos Sistemas de Recuperação de Informação (SRI's)⁴ para auxiliarem o usuário na busca e acesso à informação.

Segundo Campos (2001), as linguagens documentárias, tais como os esquemas de classificação e os tesouros, são os instrumentos responsáveis por possibilitar a comunicação entre os usuários de um sistema de informação no chamado espaço informacional. A autora destaca ainda que para a criação destes instrumentos é, obrigatoriamente, necessária uma estrutura terminológica, que será buscada em um sistema terminológico.

Diante dessa constatação, para que as chamadas linguagens documentárias possam ser de fato instrumentos que permitam a efetiva comunicação do conhecimento humano é preciso que estejam apoiadas em teorias terminológicas.

Nesta parte do trabalho serão brevemente abordadas as teorias terminológicas mais relevantes no contexto da Ciência da Informação: (i) a *teoria da classificação facetada*; (ii) a *teoria geral da terminologia*; e (iii) a *teoria do conceito*.

Antes de descrever tais teorias é importante registrar os significados que são atribuídos a palavra "terminologia" na literatura. Segundo Wüester (1979) podemos identificar, na literatura, três significados distintos para terminologia: (i) uma lista de termos e seus significados, que nos remete ao campo dos dicionários técnicos, dos vocabulários e dos léxicos; (ii) os termos de uma área de especialidade, que se refere ao estudo científico dos termos de uma área particular do conhecimento em uma certa língua; e (iii) um conjunto de princípios teóricos, que compreende a terminologia como uma disciplina científica que propicia princípios metodológicos para elaboração de terminologias estruturadas para diversos domínios – ciência da terminologia.

⁴ Sistemas de Recuperação da Informação são a interface entre uma coleção de recursos da informação, em meio impresso ou não, e uma população de usuários; e desempenham as tarefas de aquisição, armazenamento, organização, controle e distribuição de documentos aos usuários (LANCASTER e WARNER, 1993).

De maneira geral, essas três definições podem ser aplicadas para explicar o uso que é feito da palavra “terminologia” nesta pesquisa. No entanto, é preferível compreender “terminologia” a partir do terceiro significado, uma vez que foi a partir dele que a área recebeu o status de ciência e área do conhecimento.

A primeira teoria terminológica, aqui destacada, é a Teoria da Classificação Facetada, desenvolvida por Shiyali Ramamrita Ranganathan, na década de 30, a partir da *Colon Classification*. Tal teoria também foi publicada em uma segunda edição (1957) e terceira edição (1967), ambas com o nome de Teoria Dinâmica da Classificação Bibliográfica, como afirma Dahlberg (1976).

O sistema terminológico de Ranganathan se diferenciou dos demais esquemas de classificação bibliográfica pelo fato de não trabalhar com classes pré-estabelecidas, característica da teoria descritiva, que exigia que os assuntos tratados nos documentos tivessem de ser adequados à estrutura classificatória existente nos esquemas.

No esquema facetado de Ranganathan, a classificação dos livros era criada somente no momento que um livro fosse analisado de acordo com elementos constituintes do seu assunto. Para esse matemático e bibliotecário indiano, o conhecimento era visto como algo dinâmico, sendo necessário desenvolver teorias que fossem capazes de superar as barreiras apresentadas nos esquemas de classificação vigentes na época.

Five Laws of Library Science (1931), *Colon Classification* (1933), *Prolegomena to Library Classification* (1937) e *Philosophy of Book Classification* (1951) são as obras clássicas onde a Teoria de Ranganathan foi apresentada, conforme citado por Campos (2001). Nesta Teoria de Classificação Facetada, um assunto é dividido por seus múltiplos aspectos ou facetas, que podem ser compreendidas, atualmente, como um grupo de classes reunidas por um mesmo princípio de divisão. Ou seja, os princípios usados por Ranganathan são similares aqueles usados hoje em ontologias no âmbito da Ciência da Informação.

Por fim, vale ressaltar que ao introduzir cinco idéias ou categorias fundamentais (Personalidade, Matéria, Energia, Espaço e Tempo – Ranganathan (1985)) a serem utilizadas na análise dos assuntos contidos nos documentos, Ranganathan baseou-se na noção de categorização para representar contextos, “coisas” ou objetos do mundo real, introduzido por Aristóteles em seus trabalhos, embora o pesquisador indiano nunca tenha feito referência, em suas obras, ao trabalho do filósofo grego.

Outra teoria relevante é a Teoria Geral da Terminologia, desenvolvida na década de 30 pelo engenheiro austríaco Eugen Wüester, originalmente com o objetivo de garantir comunicação precisa na área de Eletrotécnica. Segundo essa teoria, a

Terminologia se ocupa dos conceitos de uma língua técnica ou especial, os quais se relacionam entre si como um sistema de conceitos (CAMPOS, 2001).

As idéias difundidas na Teoria Geral da Terminologia de Wüester foram de extrema importância para os trabalhos desenvolvidos no âmbito da chamada Ciência da Terminologia e se apresentam refletidas em três escolas clássicas: a Escola de Viena, a Escola de Praga e a Escola Soviética de Terminologia, como destaca Felber (1981, p. 47) apud Campos (2001).

O objetivo do trabalho terminológico de Wüester é a fixação de conceitos e o estabelecimento de princípios para a criação de novos termos, possibilitando, assim, comunicação mais precisa entre os especialistas dos diversos domínios do conhecimento. A partir dessa definição, pode-se verificar que a Teoria Geral da Terminologia é a origem ou o cerne de todos os trabalhos atualmente realizados para a organização e representação da informação de domínios específicos, como ocorre com as ontologias para a organização de domínios biomédicos.

Segundo a Teoria Geral da Terminologia, o trabalho terminológico inicia com o conceito, que possui uma unidade de denominação que é o termo. Assim, um termo designa um conceito que, por sua vez, é o significado do termo (WÜESTER, 1979). Para tanto, é necessário garantir uma unificação de conceitos e termos no trabalho terminológico, buscando atingir uma correspondência exata entre conceitos e termos. No entanto, esse enunciado da Teoria Geral da Terminologia nem sempre é possível na prática, uma vez que a representação exata de um objeto só pode ser objeto em si mesmo.

Apesar da imprecisão dessa definição, os princípios da Teoria de Wüester são fundamentais para a compreensão dos atuais trabalhos com terminologias nas mais diversas áreas do conhecimento. Esses princípios envolvem: (i) a identificação do objeto concreto ou abstrato; (ii) a análise das características intrínsecas e extrínsecas do objeto, de forma a definir o conceito e as relações entre conceitos; (iii) construção do sistema de conceitos, a partir do princípio de contextualização ou universo de conhecimento, em que as definições dos conceitos e seus posicionamentos no sistema estão diretamente ligados à área de conhecimento na qual a terminologia está sendo construída; e (iv) definição dos termos como forma de expressar as unidades de pensamento ou conhecimento (conceitos) presentes no sistema, conforme descritos por Wüester (1979).

As definições sobre as relações são umas das partes mais importantes na Teoria Geral da Terminologia. Segundo Wüester (1979), quando as relações são tratadas em um nível conceitual, elas passam a ser consideradas relações lógicas e ontológicas, havendo uma diferença considerável entre elas. Enquanto as relações lógicas, também chamadas de relações de abstração, resultam forçosamente da própria compreensão dos

conceitos, as relações ontológicas nascem do fato de elevarem-se a um nível de abstração as relações que existem na realidade entre os indivíduos. Assim, pode-se dizer que entre conceitos existam somente relações lógicas e que as relações ontológicas se dão entre o conceito e a realidade.

Wüester (1979) apresenta um quadro com uma síntese da classificação das relações, segundo a Teoria Geral da Terminologia, que pode ser vista na Tabela 1, a seguir.

Tabela 1 - Classificação das relações segundo a Teoria Geral da Terminologia

Relações conceituais									
Sistema de conceitos (ordenação de conceitos)									
Relações lógicas (abstração, semelhança)		Relações Ontológicas							
		Relações de contato (de contigüidade)			Relações de causalidade (em particular, relações de parentesco)				
		Relações de coordenação (em particular, relações parte-todo)	Relações de encadeamento (em particular, relações de sucessão)		Geral	filogênico	ontogênico	substâncias	
1	2	3	4	5	6	7	8	9	10
Membros das relações	Genérica	>	Conceito de inclusão		Predecessor	Ascendente		Ex: larva de ovo	
	Específica	<	Conceito de parte		Sucessor	Descendente			
	Conceito expandido								
	Conceito restrito								
	Conceito associado								
Combinações	Determinação								
	Conjunto de conceitos								
	Disjunção de conceitos		Integração						

Fonte: adaptado de Wüester (1979).

Por fim, destaca-se a Teoria do Conceito, desenvolvida pela pesquisadora alemã Ingetraut Dahlberg na década de 70, que, segundo Campos (2001), possibilitou uma

base mais sólida para a determinação e o entendimento do que consideramos como conceito, para fins de representação e recuperação da informação.

Segundo Dahlberg (1978), “conceito é uma unidade de conhecimento, cuja formação se dá pela reunião e compilação de enunciados verdadeiros a respeito de determinado objeto”. Ainda em seu trabalho, a pesquisadora apresenta três passos para a formação de conceitos: (i) o passo referencial, que consiste na seleção do referente num determinado domínio do conhecimento; (ii) o passo predicacional, no qual há uma atribuição das características relevantes ao referente; e (iii) o passo representacional, que refere-se à representação do conceito por meio do termo, em função das características atribuídas ao referente.

Em outros trabalhos, a pesquisadora Dahlberg demonstrou ainda a possibilidade de adotar os princípios para a elaboração de terminologias no âmbito das Ciências Sociais: primeiramente, em Dahlberg (1978), onde é apresentada a ligação entre a Teoria do Conceito e a Teoria da Classificação e, em Dahlberg (1980) apud Campos (2001), onde os princípios da Teoria do Conceito são empregados para a elaboração de tesouros, o que é possível ainda hoje.

Um último registro sobre essas teorias terminológicas citadas é que, como já citado, elas foram utilizadas como base metodológica para a criação dos esquemas de classificação bibliográfica, surgidos ao longo do século XIX, para descrever o estado do conhecimento da época. A Classificação Decimal de Dewey (CDD), a Classificação Decimal Universal (CDU) e a *Library of Congress Classification* são alguns exemplos de esquemas de classificação bibliográfica, até hoje utilizados nas bibliotecas do mundo inteiro, que se fundamentam em tais teorias.

Além das influências exercidas nos esquemas de classificação bibliográfica, as teorias sobre terminologia da década de 30 continuam influenciando os atuais instrumentos de controle terminológico usados para a organização e uso da informação, que incluem os vocabulários controlados, os tesouros, as redes semânticas e até mesmo as ontologias.

A teoria da classificação facetada, por exemplo, foi aplicada no âmbito das chamadas linguagens documentárias, como cita Dahlberg (1978), e seus princípios podem ser vistos, hoje, na organização das classes presentes numa ontologia. Já a teoria do conceito, ao descrever sobre enunciados verdadeiros a respeito de um objeto, contribui (relaciona-se) com os axiomas de uma ontologia, que são sentenças verdadeiras sobre uma realidade (SILVA, 2008).

3.1.3 Ontologias

Nesta seção será feita uma breve descrição sobre **ontologias** e sua utilização para a organização e representação da informação, no âmbito da Ciência da Informação. O objeto, aqui pretendido, é apenas apresentar esse tema, que é recorrente ao longo deste trabalho.

Ontologia é um tema já estudado a muito tempo pela Filosofia, a qual considera-a como “um ramo da metafísica que se preocupa com as coisas que existem” (BLACKBURN, 1996). Os filósofos, habitualmente, referem-se ao termo com “O” maiúsculo e compreendem “Ontologia” como uma descrição básica das coisas que verdadeiramente existem no mundo. Por outro lado, os cientistas da informação e da computação compreendem “ontologia” como um artefato de engenharia e um componente importante de sistemas de informação, referindo-se ao termo como “o” minúsculo, afim de diferenciá-lo da Filosofia.

Na presente pesquisa, a compreensão do conceito de “ontologia” vai além dessa visão simplista, geralmente adotada no mundo de sistemas de informação. A visão, aqui adotada, compreende a ontologia como um instrumento para a organização e representação da informação, capaz de capturar, mapear e compartilhar o conhecimento de domínios diversos, de forma a “estabelecer a verdade sobre a realidade, encontrando respostas para a questão ‘o que existe’” (SMITH, 2004, p. 22). Nesse sentido, compreende-se ontologias muito além do que um simples artefato da engenharia de software.

O estudo de ontologias na organização e representação do conhecimento surgiu como um ramo de pesquisa no final dos anos 80, propondo ser uma alternativa aos demais instrumentos usados para representar o conhecimento. Usadas como instrumentos para a modelização do conhecimento de domínio diversos, as ontologias apresentam uma natural proximidade com as técnicas metodológicas usadas pela Ciência da Informação para organização do conhecimento humano, estando também inseridas no arcabouço metodológico desta ciência.

No entanto, apesar da proximidade com o assunto, o interesse sobre ontologias na Ciência da Informação de fato acontece na década de 90, a partir dos trabalhos de Soergel (1997) e de Vickery (1997), que é considerado o precursor de ontologias nesta área, ao pesquisar sua aplicação no campo da biblioteconomia.

Desde então, o interesse na utilização de ontologias em Ciência da Informação aumentou consideravelmente e o seu uso já está difundindo em diversos ramos do conhecimento da área. Vickery (1997) e Smith (2004) relatam algumas aplicações na área:

- Para dar suporte à extração de informação, por exemplo, das grandes bibliotecas médicas ou da literatura científica sobre a área de medicina, que, em grande parte, encontra-se disponível em diversas fontes na *web*.
- Na tradução de linguagem natural, na qual as ontologias podem ser úteis para o processo de *parsing* e para resolver problemas de ambiguidade.
- Para a integração automática de um conjunto de vocabulários padronizados ou dicionários de dados relacionados a um domínio específico. A ontologia, neste caso, pode auxiliar na construção de um único vocabulário padrão, resolvendo um problema comum em terminologias, conhecido como fusão.
- Ontologias podem alimentar um banco de dados com informações sobre categorias (conceitos) existentes no mundo/no domínio, além das propriedades referentes a esses conceitos, bem como as relações existentes entre eles.
- Para a integração de bancos de dados, de softwares ou de modelos de negócio.

Atualmente, as potencialidades de uso das ontologias em Ciência da Informação e em outras áreas correlacionadas estão reforçadas pelo grande volume e diversidade de dados manipulados na *web* e, em especial, pelo desenvolvimento e avanço das tecnologias semânticas para uso no ambiente *web*, as quais se inserem no contexto da *web semântica*.

Na literatura da área são encontradas muitas definições para o termo “ontologia” (SOWA, 2000; GRUBER, 1993; GUARINO e GIARETTA, 1995; SOERGEL, 1997; VICKERY, 1997; SMITH, 2004), as quais são concordantes em alguns pontos e discordantes em outros. Muitas das diferenças entre tais definições deve-se ao fato de que algumas delas aproximam-se mais de seu conceito na Filosofia, enquanto outras estão mais próximas de sua definição no contexto dos sistemas de informação. Na presente pesquisa, defende-se o argumento de que as ontologias devem ser fundamentadas tanto numa abordagem filosófica quanto no seu papel como um artefato da engenharia de software, não sendo ambos excludentes.

A definição de ontologia como “uma especificação explícita de uma conceitualização” (GRUBER, 1993) é uma das mais citadas na literatura. Borst (1997) acrescenta a essa definição os termos “formal” e “compartilhada”, afirmando que ontologia é “uma especificação explícita e formal de uma conceitualização compartilhada”, onde “especificação explícita” refere-se a conceitos, atributos, relações e axiomas explicitamente

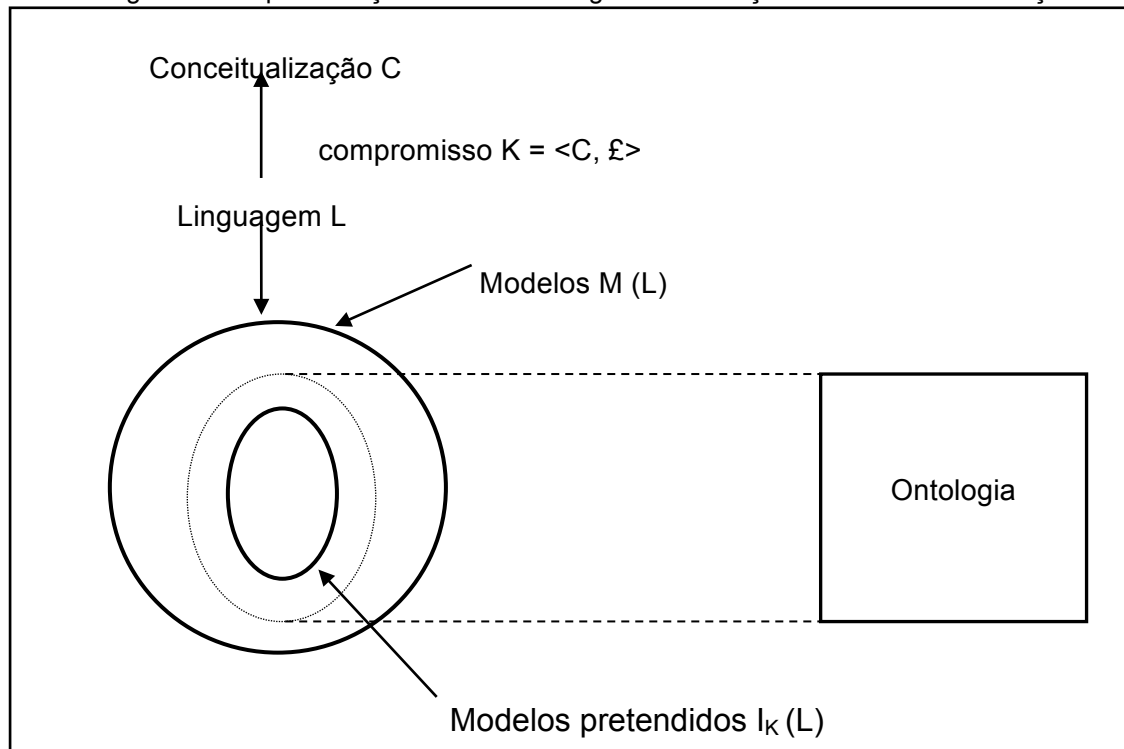
definidos, “formal” significa legível por máquinas (computadores), “conceitualização” pode ser entendido como um modelo abstrato de algum fenômeno do mundo real e “compartilhada” significa conhecimento consensual.

O problema de tais definições está relacionado com o uso do termo “conceitualização”. Segundo Guarino e Giaretta (1995) seria incorreto considerar a noção extensional ao definir “conceitualização”. Para esses autores, a noção extensional está relacionada a um estado particular das coisas e não representa a estrutura semântica de certa realidade ou modelo pretendido. Uma melhor definição para “conceitualização”, a partir de uma noção intensional, é a de um conjunto de modelos semânticos, os quais representam certa realidade.

Pela interpretação intensional, a conceitualização precisa se comprometer em usar um vocabulário comum ou compartilhado que seja coerente com a linguagem do modelo pretendido, que é compreendido como compromisso ontológico. Assim, um conjunto de modelos compõem um conceitualização e eles são construídos por meio de uma linguagem de modelagem especificada pela ontologia.

Essa interpretação faz parte do trabalho de Guarino (1998) e é, normalmente, bem aceita na literatura da área. O autor ilustra essas relações entre linguagem, conceitualização, compromisso ontológico e ontologia, como pode ser visto na Figura 2.

Figura 2 – Representação de uma ontologia e sua relação com a conceitualização



Fonte: adaptado de Guarino (1998).

Na Figura 2, é possível perceber que para uma ontologia se aproximar de um modelo pretendido deve-se obter uma conceitualização a partir de uma boa axiomatização – regras formais – um conjunto formal de conceitos e suas relações relevantes. Nesse sentido, Guarino (1998) aprimora a definição de Gruber (1993), descrevendo a ontologia constituída de um vocabulário intensional utilizado para descrever certa realidade, em conjunto com pressupostos explícitos na forma de lógica de primeira ordem, representando conceitos e relações entre eles.

Ainda considerando a relação de ontologia e conceitualização, pode-se afirmar que a conceitualização é uma das mais importantes etapas do processo de construção de ontologias, porque é nela em que se realizam atividades como identificação, análise e negociação dos conceitos de um determinado domínio, que poderão se tornar classes ou relações da ontologia (PEREIRA e SOARES 2008). Nesse sentido, em qualquer projeto que envolva o desenvolvimento de ontologias é necessário dar uma atenção especial à fase de conceitualização, cujos resultados – modelos conceituais do domínio – irão influenciar fortemente a construção dos elementos da ontologia.

A conceitualização é em geral intermediada por um responsável pela modelagem do domínio (analistas de sistemas, analistas de negócio, desenvolvedores de ontologias, etc.), mas realizado por especialistas (médicos, engenheiros, advogados, etc.) em domínios do conhecimento específicos que são de interesse para o problema analisado. Os especialistas fornecem a expertise para o entendimento e modelagem das entidades e dos processos que serão representados na ontologia do domínio. Entretanto, os diversos especialistas tem diferentes visões de mundo, o que resulta em diferentes conceitualizações concorrentes e ditas corretas (BERZEL, 2010). Dessa forma, *conceitualizações colaborativas* têm sido adotadas e incentivadas no ambiente web, por exemplo através de ontologias desenvolvidas colaborativamente (TUDORACHE et al., 2008), como forma de fomentar a discussão e proporcionar a produção de sistemas realmente representativos das necessidades de uma comunidade. Nesse sentido, a *colaboração* no processo de conceitualização é uma estratégia bastante indicada para o desenvolvimento de ontologias e que deve ser levada em consideração em projetos atuais que envolvam o ambiente web e o uso de ontologias.

Outra característica essencial de ontologias que deve ser mencionada são seus componentes básicos, os quais formam o conteúdo ontológico e serão tratados recorrentemente ao longo da presente pesquisa. De acordo com Gruber (1993) os componentes básicos de uma ontologia são: (i) *classes*: organização dos conceitos de um domínio, dispostas através de uma taxonomia; (ii) *relações*: representam o tipo de interação entre as classes do domínio; (iii) *axiomas*: regras formais usadas para restringir a

interpretação e o uso dos termos envolvidos, geralmente, especificadas em uma linguagem formal da lógica; e (iv) *instâncias*: exemplos ou casos particulares das classes, usadas para representar objetos específicos.

Além desses elementos citados, outra importante informação que deve estar contida em uma ontologia são as chamadas *questões de competência*. Segundo Uschold e Gruninger (1996), as questões de competência são utilizadas para especificar tarefas e os problemas que uma ontologia pode solucionar, antes mesmo de sua construção. Trata-se, então, de uma forma de auxiliar na definição do escopo e das características da ontologia.

As questões de competência também são importantes para uma ontologia no sentido de prover condições para avaliação da eficácia e completude da mesma. Nesse sentido, Kim, Fox e Gruninger (1999) afirmam que a competência de uma ontologia é determinada pelos seguintes passos:

- *Declaração de cenário*: consiste em uma narrativa sobre questões do negócio e sobre problemas que os sistemas baseados na ontologia devem ser capazes de referenciar;
- *Declaração de escopo*: consiste na criação de hipóteses sobre um domínio, de forma a esclarecer sua abrangência; a partir dessas hipóteses, estimam-se os objetos, as relações e os atributos que devem compor a ontologia;
- *Declaração de problema*: consiste em estabelecer o problema geral que justifica a construção da ontologia; trata-se de uma questão que, originada nos cenários e limitada pelo escopo, serve como base para outras questões de competência;
- *Declaração de questões de competência do usuário*: consiste em questões específicas de competência, motivadas pela declaração de cenário e apresentadas na forma de uma declaração de problemas; são elaboradas de acordo com as necessidades de um usuário.

Outro importante conceito que deve ser mencionado sobre esse tema refere-se aos tipos de ontologias existentes. Na literatura da área é possível encontrar diferentes classificações para as ontologias, baseadas em critérios diversos, tais como: grau de formalidade, tipo de estrutura, assunto da conceitualização, função específica, grau de rigorosidade, entre outros.

Para a presente pesquisa interessa, particularmente, três critérios de classificação das ontologias: (i) tipo de estrutura; (ii) grau de rigorosidade; e (iii) propósito ou papel da ontologia.

Quanto ao tipo de estrutura de uma ontologia, isto é, os tipos de classes presentes nela, os autores (HAAV e LUBI, 2001; GUARINO, 1998) classificam-nas em:

- *Ontologias de alto nível*, que descrevem conceitos gerais como espaço, tempo, matéria, objeto, evento, ação, etc., os quais são independentes do problema ou domínio;
- *Ontologias de domínio*, que descrevem o vocabulário de um domínio, por exemplo, medicina ou automóveis;
- *Ontologias de tarefa*, que descrevem uma tarefa ou atividade, por exemplo, diagnósticos ou compras, através da inserção de termos especializados.

A *Basic Formal Ontology* (BFO) e a *Unified Formal Ontology* (UFO) são exemplos de **ontologias de alto nível** (na literatura da área, também chamadas de **ontologias de fundamentação**⁵, **ontologias fundacionais** ou ainda **meta-ontologias**), enquanto as ontologias sobre o sangue humano, aqui desenvolvidas, são exemplos de **ontologias de domínio**.

Geralmente, as ontologias de fundamentação são usadas como base ou ponto de partida para a construção de ontologias sobre domínios específicos e, como citado anteriormente, essa é uma prática científica bem aceita na literatura da área.

Já quanto ao grau de formalidade ou grau de rigorosidade, alguns autores (DING e ENGELS, 2001; GÓMEZ-PÉREZ, 2004) classificam as ontologias em:

- *Ontologias leves (light-weight)*, que são ontologias com pouco rigor formal, geralmente, composta de classes facilmente compreensíveis e de relações mais comuns entre estas classes, não incluindo relações especiais – por exemplo, relações lógicas - entre as classes e outros tipos de primitivas de representação – por exemplo, axiomas lógicos.
- *Ontologias pesadas (heavy-weight)*, que se referem a ontologias com alto rigor formal, incluindo além das classes e relações comuns, relações especiais e alto grau de axiomatização. Em geral, essas ontologias são especificadas com um rigoroso grau matemático e formal, baseado nas linguagens lógicas, tal como a lógica de primeira ordem.

⁵ Na presente pesquisa, optou-se por utilizar o termo “ontologias de fundamentação”, preferencialmente, em relação aos termos “ontologias de alto nível”, “ontologias fundacionais” ou “meta-ontologias”.

Por fim, as ontologias são classificadas por alguns autores (GUARINO, 1998; FONSECA, 2007) quanto ao seu papel para o desenvolvimento de sistemas de informação, ou seja, com qual propósito são criadas. Nesse caso, temos dois tipos:

- *Ontologias de sistemas de informação*: ontologias criadas para auxiliar a validação de ferramentas usadas na criação de modelos conceituais.
- *Ontology-driven information systems* ou *ontologias para sistemas de informação*: são criadas para explicar um dado domínio de conhecimento e, por se basearem em princípios filosóficos, tem como propósito serem usadas para geração ou validação de componentes dos sistemas de informação, tais como os esquemas conceituais.

3.2 Organização da informação médica

Alguns dos instrumentos de representação da informação mais comuns em Ciência da Informação, apresentados na seção anterior (3.1), são também empregados para organização da informação em domínios biomédicos. A presente seção tem por objetivo descrever tais instrumentos, mostrando como são utilizados nesta área do conhecimento.

Como citado anteriormente, Bodenreider (2008) nos alerta para o aumento no número de artefatos criados para a representação de entidades biomédicas, seus termos e suas relações, verificado na última década. O autor mostra que o número de citações sobre ontologias e vocabulários controlados na base de dados MEDLINE/PubMed⁶ - uma das maiores e mais relevantes bases da área médica - vêm crescendo de 600% a 1200% a cada ano.

Apesar desse aumento registrado, muitos problemas na organização e uso da informação médica continuam em aberto, conforme já apresentado na problemática desta pesquisa. Nesse sentido, torna-se importante uma revisão de literatura da área, apontando os pontos fortes e fracos de alguns dos instrumentos de organização da informação médica.

A presente seção versará, basicamente, sobre dois tipos de instrumentos mais usados em domínios biomédicos: as terminologias médicas e as ontologias biomédicas, que são apresentadas na subseção 3.2.1. Por sua vez, na subseção 3.2.2, discorre-se sobre alguns repositórios de ontologias biomédicas, os quais objetivam promover a

⁶ Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/>. Acesso em: 04 de Maio de 2012.

integração e a padronização da informação médica através da construção de ontologias na área.

3.2.1 Terminologias e ontologias biomédicas

O domínio biomédico, historicamente, quase sempre contou com artefatos para representação e organização dos termos específicos da área, devido à vasta gama de termos técnicos usados.

Desde os primórdios da história da medicina, os gregos (fundadores da medicina moderna) e os romanos (difusores dos termos médicos pelo mundo) já utilizavam alguns artefatos para a organização dos termos da área. Os registros históricos relatam que o primeiro dicionário médico surge na década de 1830, compostos de termos de origem latina – idioma usado por gregos e romanos nesta época (HENDERSON e DORSEY, 2009).

No que se refere, propriamente, às **terminologias**, compreendidas como artefatos linguísticos que unem os diversos sentidos ou significados das entidades linguísticas, seu primeiro uso na medicina é registrado durante o século XVII. Neste período, autoridades de saúde de Londres, na Inglaterra, utilizaram uma lista padronizada de aproximadamente 200 causas de morte – a *London Bills of Mortality* - para tabular as mortalidades ocorridas na cidade. Mais tarde, a *London Bills of Mortality* foi incorporada a um tradicional sistema de classificação médica, conhecido como Classificação Internacional de Doenças (CID) (GERSENOVIC, 1995).

A partir de então, diversas terminologias na área médica tem sido criadas com o propósito de coletar nomes de substâncias, qualidades, estruturas e processos utilizados tanto no ambiente de pesquisa como na prática clínica, na recuperação de documentos, no registro de estatísticas de mortalidade e morbidade, entre outros.

De maneira geral, pode-se afirmar que uma **terminologia médica** é um instrumento de organização da informação composto de termos que descrevem a anatomia e a fisiologia humana (órgãos do corpo, sistemas e suas funções), partes do corpo humano, doenças, procedimentos clínicos, diagnósticos, entre outros.

Nesta seção serão apresentadas as terminologias médicas mais conhecidas e relevantes para a padronização de termos da área, que incluem a Classificação Internacional de Doenças (CID), o *Medical Subject Headings* (MeSH), o vocabulário Descritores em Ciências da Saúde (DeCS), o *National Center Institute's (NIC) Thesaurus* e o *Systematized Nomenclature of Medicine* (SNOMED). Por fim, será feita uma breve descrição do *Unified Medical Language System* (UMLS), que congrega mais de vinte terminologias médicas especializadas organizadas em uma estrutura única. O UMLS

engloba, por exemplo, as terminologias anteriormente citadas: CID, MeSH, SNOMED e DeCS.

A **Classificação Internacional de Doenças (CID)** foi criada em 1880 pela Organização Mundial de Saúde (OMS) baseada na *London Bills of Mortality* (OMS, 2012). Atualmente em sua versão 10 – a CID-10 – engloba aproximadamente 13000 classes para a classificação de doenças e formas de contratação. A CID-10 é mantida pela OMS e, segundo Freitas e Schulz (2009), representa, atualmente, o sistema de codificação de doenças mais amplamente utilizado em todo o mundo, servindo de base comum para as estatísticas de saúde.

A CID-10 está dividida em 22 capítulos que denotam classes de doenças e problemas relacionados, o que significa que cada doença específica se encaixa em uma categoria com um código único. A hierarquia *is-a* é a única relação de construção hierárquica presente na CID e que contém ao todo cinco níveis hierárquicos. Além disso, a CID garante que nenhuma classe tenha mais que uma classe matriz (ou nó pai), com o objetivo de impedir que uma doença seja contada mais de uma vez.

O **Medical Subject Headings (MeSH)** é um vocabulário controlado criado pela *National Library of Medicine* (NLM) para indexação do *Index Medicus*⁷. Atualmente é utilizado na indexação de documentos da área de saúde, principalmente resumos literários da base de dados biológica MEDLINE (NELSON e SCHULMAN, 2007).

Com relação à sua estrutura, o MeSH consiste de um conjunto de termos, denominados “descritores”, disposto em uma estrutura hierárquica – ao todo são 16 árvores ou ramificações – que permite a busca em vários níveis de especificidade. Os nós das árvores do MeSH são chamados de *cabeçalhos* e denotam um significado padronizado de um grupo de termos médicos (FREITAS e SCHULZ, 2009). Por exemplo, o cabeçalho *MeSH Leishmaniose* é parte da hierarquia *Doenças Parasitárias* e também da hierarquia *Doenças da Pele e do Tecido Conjuntivo*, conforme apresentado em Freitas e Schulz (2009). Desta forma, o MeSH dispõe seus cabeçalhos em hierarquias múltiplas, em contraste ao CID, com cada cabeçalho possuindo um identificador único e um “número de árvore” a que pertence.

Sobre o MeSH, Hersh (2003) afirma ainda que este vocabulário possui três tipos de relacionamentos entre seus termos: (i) o hierárquico, (ii) os sinônimos e (iii) os

⁷ *Index medicus* foi um índice criado em 1879 por John Shaw Billings para ajudar profissionais médicos a encontrar artigos relevantes em jornais da época. Os artigos de jornais eram indexados por nome de autor e assunto do título e, então, eram colocados em volumes separados para um pesquisador encontrar com maior facilidade um assunto específico (HERSH et al., 2001).

relacionados; que são usados para auxiliar na busca de documentos que contenham referências a uma doença, por exemplo.

Em 1986, a Biblioteca Regional de Medicina (BIREME⁸) traduziu e adaptou o MeSH criando o vocabulário controlado **Descritores em Ciências da Saúde (DeCS)**, para uso na indexação de artigos de revistas científicas, livros, anais de congressos, relatórios técnicos e também para ser usado na pesquisa e recuperação de assuntos da literatura científica nas bases de dados da Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS), SciELO e a própria MEDLINE, como afirma Pellizzon (2004).

Desta forma, o DeCS caracteriza-se como um vocabulário trilingue (português, espanhol e inglês) com o propósito de facilitar o acesso à informação biomédica em tais idiomas. Segundo a BIREME (2012), o DECS engloba aproximadamente 26851 descritores, sendo 3656 referentes à Saúde Pública e outros 1950 relacionados com Homeopatia.

Criado pela comunidade de pesquisa do câncer, o **National Center Institute's (NIC) Thesaurus** é um vocabulário controlado de domínio público, baseado no UMLS MetaThesaurus. Seu propósito principal é integrar informações clínicas e moleculares relacionados ao câncer e, para tanto, contém uma estrutura taxonômica de conceitos clínicos e básicos usados nas pesquisas de câncer, além das relações entre esses conceitos. Além disso, é baseado em lógica descritiva. (NCI, 2012 – 81, 82).

Considerado por seus autores (GOLBECK et al. 2003) como um vocabulário mais amplo e mais profundo que os demais vocabulários clínicos controlados, porque contém maior riqueza de interrelacionamento semântico entre seus nós, o tesouro NCI sofre dos mesmos problemas das terminologias médicas (CEUSTERS, SMITH e GOLDBERG (2005); KUMAR e SMITH (2005)): (i) erros e inconsistências nos princípios de formação dos termos, que incluem uso de verbos e definições inapropriadas; (ii) sinonímia e (iii) problemas com as relações especificadas.

Apesar das falhas identificadas do tesouro NCI, Kumar e Smith (2005) consideram que tal terminologia, caso passe por processos de refinamento, constitui-se num ótimo ponto de partida para o trabalho ontológico no domínio do câncer, podendo ser, futuramente, uma ontologia de referência neste domínio.

O **Systematized Nomenclature of Medicine (SNOMED)** foi publicado pela primeira vez, em 1975, como sucessor do *Standard Nomenclature Pathology (SNOP)*, desenvolvido pelo *College of American Pathologists* para englobar termos clínicos de todas

⁸ BIREME é o nome original do atual Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde, que é um centro especializado da Organização Pan-Americana da Saúde / Organização Mundial da Saúde (OPAS/OMS), orientado à cooperação técnica em informação científica em saúde (BIREME, 2012).

as especialidades de medicina (IHTSDO-SNOMED, 2012). Atualmente em uma versão conhecida como SNOMED-CT - *Systematized Nomenclature of Medicine-Clinical Terms* – o SNOMED é uma terminologia abrangente que já passou por várias revisões publicadas em versões diferentes: SNOMED-2 (1979), SNOMED-3 (1993), SNOMED-3.5 (1996) e SNOMED-RT (*Reference Terminology*) (1997).

O SNOMED-CT, resultado da fusão do SNOMED-RT com a versão 3 do *UK Clinical Terms*, cobre o registro do paciente por inteiro, além de abordar estruturas corporais, procedimentos e aspectos relevantes relacionados à saúde (SPACKMAN, 1997 apud FREITAS e SCHULZ, 2009). Desde abril de 2007 é mantido e distribuído pela *International Health Terminology Standards Development Organization* (IHTSDO).

Quanto à sua estrutura, o SNOMED-CT possui 310.000 nós distribuídos em múltiplas hierarquias *is-a*. Alguns desses nós são chamados de *conceitos* e denotam classes de entidades como doenças, procedimentos, medicamentos e até entidades geográficas, como afirmam Freitas e Schulz (2009). Tais conceitos incluem sinônimos e, em alguns casos, definições em texto livre.

Duas novidades do SNOMED-CT em relação às terminologias, anteriormente citadas, é a inclusão de *qualificadores*, que oferecem refinamentos opcionais para os conceitos, e também 50 tipos de links chamados de *conceitos de ligação*, que podem ser expressos em lógica descritiva para representar relações entre os conceitos. A partir disso, pode-se afirmar que o SNOMED-CT aproxima-se de uma representação ontológica simplificada.

Nesta breve revisão de literatura das terminologias médicas é possível notar a proliferação destes instrumentos, nos últimos anos, visando à organização da informação biomédica. Com o propósito de buscar um alinhamento de tais terminologias e vocabulários controlados, surgiram projetos e estruturas únicas para unificar essas propostas, tais como o ***Unified Medical Language System (UMLS)*** e a ***Open Biological Ontologies (OBO)***.

O *Unified Medical Language System* (UMLS) foi criado em 1986, pela *National Library of Medicine* (NLM), com o propósito de integrar informações de diversas fontes terminológicas incompatíveis. O UMLS pode ser definido como um repositório de vocabulários biomédicos que integra, atualmente, cerca de 2 milhões de termos médicos em 900.000 conceitos de mais de 60 famílias de vocabulários biomédicos, além de 12 milhões de relações entre esses conceitos (BODENREIDER, 2004).

Apesar do seu propósito de integração, o projeto UMLS enfrenta problemas por não possuir uma arquitetura comum e terminologias separadas (CEUSTERS et al., 2004). Além disso, os mapeamentos UMLS não conseguem fundir/unificar seus termos em um

único sistema (CAMPBELL, OLIVER e SHORTLIFFE, 1998), já que tais mapeamentos são realizados em termos contidos em terminologias separadas – não integradas em uma mesma arquitetura.

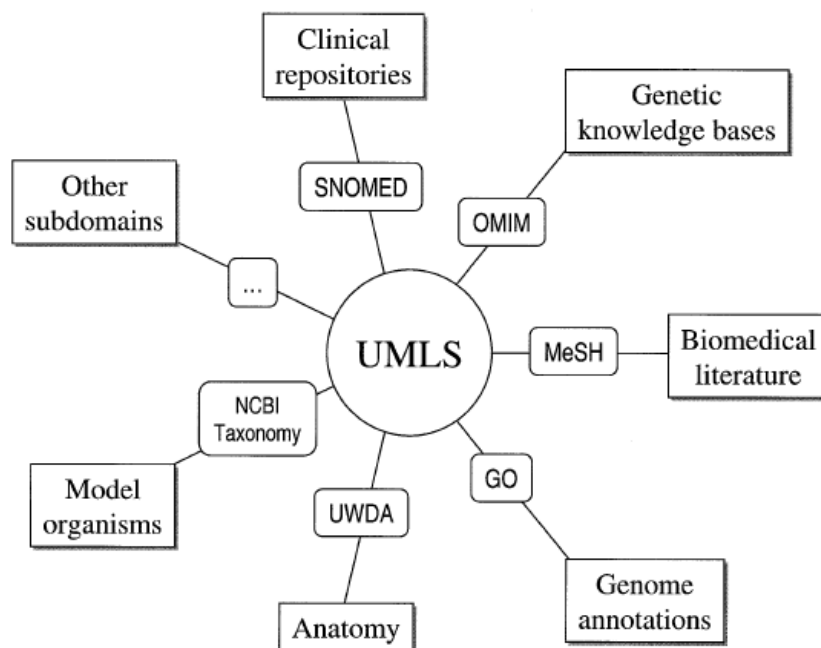
Mesmo com esses problemas citados, o UMLS é ainda muito utilizado, atualmente, por uma gama enorme de pesquisadores, profissionais e cientistas, caracterizando-se como um dos principais recursos para indexação e recuperação de documentos da literatura biomédica.

Seus três componentes básicos são: (i) o *Metathesaurus* UMLS, um repositório de conceitos biomédicos inter-relacionados, que representa seu principal componente; (ii) a Rede Semântica UMLS, que provê categorias básicas ou de alto nível usadas para categorizar cada conceito do *Metathesaurus*; e (iii) recursos léxicos, que incluem o léxico SPECIALIST e programas para geração de variações léxicas dos termos biomédicos (BODENREIDER, 2004).

Ao nomearem o tesouro UMLS como um metatesouro, seus criadores introduzem a idéia de terem desenvolvido um tesouro dos tesouros, que engloba os demais tesouros da área biomédica, preservando seus nomes, idéias, contextos hierárquicos, atributos de interrelacionamento e adicionando a eles informação básica para cada conceito, além de estabelecerem novos relacionamentos entre os termos destes diferentes vocabulários.

Nelson, Powell e Humphreys (2006) afirmam que a estrutura unificada do *Metathesaurus* UMLS constitui a mais rica fonte atual de terminologias, tesouros, sistemas de classificação e ontologias biomédicas, cobrindo diversos subdomínios da área biomédica. Esses subdomínios cobertos pela UMLS são apresentados por Bodenreider (2004), conforme Figura 3, a seguir.

Figura 3 - Os vários subdomínios biomédicos integrados ao UMLS.



Fonte: adaptado de Bodenreider (2004).

A Figura 3 ilustra alguns dos subdomínios biomédicos cobertos pela UMLS a partir das terminologias, vocabulários e ontologias biomédicas usados em tal estrutura. Assim, temos: (i) a taxonomia NCBI para representar os organismos; (ii) o *Digital Anatomist Symbolic Knowledge Base* (UWDA) para representar o domínio de anatomia; (iii) a *Gene Ontology* (GO), usada para anotação dos produtos genéticos dos vários modelos de organismos; (iv) o vocabulário MeSH, para cobrir a literatura médica na base MEDLINE; (v) o *Online Mendelian Inheritance in Man* (OMIM), como uma de suas bases de representação do conhecimento genético; e (vi) a terminologia SNOMED Internacional (incluindo a SNOMED-CT), para representar os componentes clínicos da biomedicina, além de outros subdomínios específicos não apresentados em tal figura.

Sobre a estrutura do *MetaThesaurus* UMLS, Nelson, Powell e Humphreys (2006) descrevem que cada conceito possui uma identificação única - *Concept Unique Identifier (CUI)* – que mapeia um ou mais termos das diferentes terminologias da UMLS. Esses termos também possuem um identificador único - *Common Term Identifier (LUI)* – e mapeiam uma ou mais variações morfológicas dos termos, representadas por strings com um identificador único - *String Unique Identifier (SUI)*. O quadro da Figura 4 ilustra essa estrutura do UMLS, apresentando exemplos de cada mapeamento.

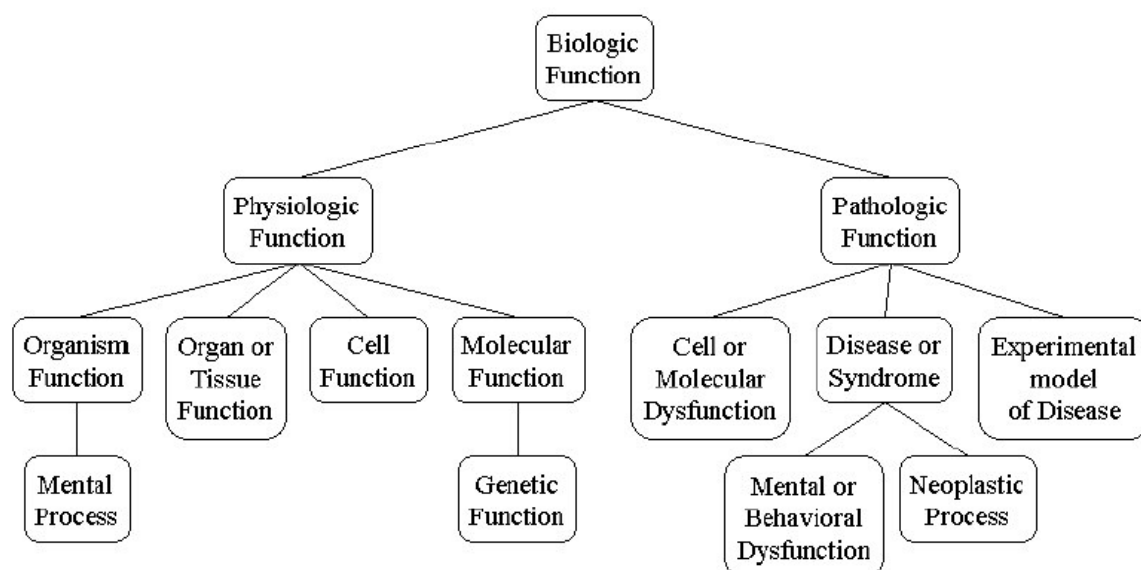
Figura 4 - Exemplos de identificadores no MetaThesaurus UMLS.

CUI	LUI	SUI
C0004238 Atrial Fibrillation Atrial Fibrillations Auricular Fibrillation Auricular Fibrillations	L0004238 Atrial Fibrillation Atrial Fibrillations	S0016668 Atrial Fibrillation
		S0016669 Atrial Fibrillations
	L0004327 Auricular Fibrillations Auricular Fibrillations	S0016899 Auricular Fibrillation
		S0016900 Auricular Fibrillations

Fonte: adaptado de NLM-UMLS (2012).

Já a Rede Semântica do UMLS provê uma categorização consistente dos conceitos existentes no *MetaThesaurus*, estruturando-os em tipos semânticos ou categorias básicas. Uma árvore de 135 tipos semânticos, ligados por relações *is-a*, forma o suporte principal desta rede, que ainda inclui uma hierarquia de 53 relações associativas (por exemplo, *localização_de*, *trata*), que são utilizadas para formar 612 trios (por exemplo, *Tecido*, *Procedimento Diagnóstico*), dos quais 6252 trios podem ser inferidos (FREITAS e SCHULZ, 2009). Uma parte desta Rede Semântica da UMLS é apresentada na Figura 5, a seguir.

Figura 5 - Extrato da rede semântica UMLS.



Fonte: adaptado de NLM-UMLS (2012).

Por fim, temos os recursos léxicos do UMLS, que incluem o já citado SPECIALIST e outros programas, tais como o LVG, usado para geração das variações léxicas dos termos biomédicos e o *Metamap*, que pode ser utilizado para a extração de conceitos UMLS de um conjunto de textos como fonte de dados (BODENREIDER, 2004). Tais recursos têm sido sendo desenvolvido para trabalhar com Sistemas de Recuperação de Informação que implementem Processamento de Linguagem Natural (PLN) usando a língua inglesa. No caso do léxico desenvolvido, ele registra para cada palavra a informação sintática, morfológica e ortográfica necessária para um sistema de PLN especialista.

As terminologias médicas, normalmente, são construídas com fins definidos, como a recuperação de documentos ou registro de estatísticas, e seus termos são definidos utilizando expressões da linguagem humana, assim como as relações entre esses termos expressam relações informais próximas da linguagem humana. Resumidamente, pode-se dizer que as terminologias médicas não utilizam descrições formais e bem definidas, como é possível com o uso de ontologias.

Como alternativa às terminologias médicas, as **ontologias biomédicas** têm sido utilizadas como forma de solucionar os problemas mencionados. Suas vantagens em relação às terminologias relacionam-se, principalmente, aos formalismos baseados em lógica, que são usados para mapear o conhecimento de um determinado domínio. Desta forma, é possível representar com mais expressividade as classes (conceitos), suas relações, suas ocorrências (ou instâncias) e as restrições, tratadas como axiomas. Todas essas características permitem que as ontologias (não apenas biomédicas) possam ser processadas pelos computadores e, assim, verificar se uma sentença está ou não correta frente à semântica do domínio e inferir novas sentenças a partir daquelas já existentes.

Projetos que reúnem diferentes fontes de conhecimento biomédico, tal como o UMLS, passaram a incluir além das terminologias médicas em sua estrutura, as ontologias biomédicas. No UMLS, por exemplo, houve, inicialmente, a incorporação da *Gene Ontology* à sua estrutura e, posteriormente, à inclusão do *Digital Anatomist Foundational Model of Anatomy* (ROSSE e MEJINO, 2003), mais conhecido na literatura como *Foundational Model of Anatomy* (FMA).

Como forma de ilustrar exemplos de ontologias biomédicas, apresenta-se aqui uma breve introdução de algumas das mais conhecidas e importantes, trazendo uma breve descrição delas e o domínio médico onde são usadas.

A Tabela 2, a seguir, apresenta tais ontologias biomédicas, caracterizando-as a partir a partir do seu nome, que tipo de ontologia são, o domínio de aplicação e uma breve descrição.

Tabela 2 - Ontologias biomédicas.

Ontologia	Tipo	Domínios	Descrição
<i>Basic Formal Ontology (BFO)</i>	Ontologia de fundamentação	Domínios biomédicos em geral	Ontologia de orientação realista, que fornece suporte aos desenvolvedores de ontologias de domínio, orientando-os em suas tomadas de decisões ontológicas. (GRENON e SMITH, 2004)
<i>Biological Top-Level (BIOTOP)</i>	Ontologia de fundamentação	Domínios biomédicos em geral	Ontologia desenvolvida para prover uma camada ontológica para a ligação e integração de diversas ontologias de domínios específicos em ciências da vida (BEISSWANGER et al., 2007).
<i>Cell Ontology (CL)</i>	Ontologia de domínio	Anatomia (Células)	Ontologia que representa e descreve os tipos de células. Nela estão inclusos tipos de células que vão desde as procariotas até as células mamárias, cobrindo assim, diferentes tipos de organismos (OBO, 2014).
<i>Foundational Model Anatomy (FMA)</i>	Ontologia de domínio	Anatomia	Ontologia de domínio que representa o conhecimento declarativo e explícito sobre a anatomia humana. (ROSSE e MEJINO, 2003).
<i>Generalized Architecture for Language, Encyclopaedias and Nomenclatures (GALEN)</i>	Ontologia de domínio	Anatomia, fisiologia humana, patologia e sintomatologia	Uma ontologia clínica de fonte aberta e um modelo de referência comum. (RECTOR et al., 2003).
<i>Gene Ontology (GO)</i>	Ontologia de domínio	Genético e biológico	Uma ontologia, às vezes considerada apenas um vocabulário controlado, usada para descrever os produtos genéticos e suas funções em qualquer organismo (SMITH, WILLIAMS e SCHULZE-KREMER, 2003).
<i>Protein Ontology (PRO)</i>	Ontologia de domínio	Proteínas	Ontologia que descreve relacionamentos entre as proteínas e as classes evolucionárias de proteínas, além das múltiplas formas de proteínas dos genes. (OBO, 2014)

Fonte: elaborado pelo autor.

Evidentemente, existe ainda uma grande quantidade de ontologias biomédicas, não citadas na Tabela 2, muito em função da complexidade inerente ao domínio médico.

Maiores detalhes sobre as ontologias biomédicas, aqui citadas, e outros exemplos são destacados no capítulo 4 desta tese, cujo assunto principal são as aplicações destas ontologias em domínios do conhecimento.

3.2.2 Repositórios de ontologias

Desde a proliferação das terminologias médicas e das ontologias biomédicas, disponibilizadas, em sua maioria, através da *web*, alguns projetos e iniciativas de integração e padronização da informação médica foram desenvolvidos com o propósito de possibilitarem uma melhor organização e uso da informação na área. A idéia geral de tais projetos concentra-se em tentativas de integrar o conteúdo semântico das terminologias médicas e ontologias biomédicas, atualmente disponíveis e em desenvolvimento.

Diante da importância desses projetos para a organização da informação médica, a presente seção cita alguns exemplos, destacando, principalmente, os repositórios de ontologias, tal como a *Open Biomedical Ontologies* (OBO), que visam à integração e padronização das ontologias biomédicas.

Uma das primeiras iniciativas para integração da informação médica, especialmente no que se refere às terminologias, foi o projeto UMLS, descrito anteriormente. Criado no ano de 1986, o UMLS é um repositório de vocabulários biomédicos contendo mais de 60 famílias destes vocabulários e cerca de 2 milhões de termos médicos.

Outros dois projetos mais recentes de organização e representação da informação biomédica são: (i) o *National Cancer Institute's Biomedical Informatics Grid*⁹ (caBIG) e (ii) o *HL7's Reference Information Model*¹⁰ (RIM). O caBIG tem por objetivo integrar todos os dados relativos à pesquisa sobre câncer numa ciber-infraestrutura em comum, de modo a padronizar a aquisição, formatação, processamento e armazenamento dos dados. De forma similar, o HL7 RIM oferece um padrão de intercâmbio, gerenciamento e integração de toda informação relevante no cuidado à saúde, desde a genética clínica até o faturamento hospitalar (SMITH et al., 2007).

Esses dois últimos projetos estão interessados, especificamente, em como os dados e a informação deve ser representada num computador e não se preocupam com a representação conceitual da informação biomédica, independente de seu processamento computacional.

⁹ Disponível em: <https://cabig.nci.nih.gov>. Acesso em: 30 de Abril de 2012.

¹⁰ Disponível em: <http://hl7.org>. Acesso em: 30 de Abril de 2012.

Foi considerando o contexto apresentado, que, no ano de 2001, Ashburner e Lewis criaram a plataforma **Open Biomedical Ontologies (OBO)**, que evoluiu como uma fonte de ontologias biomédicas *online* de domínio público. A partir de então, Smith et al. (2007) destaca que a *OBO Foundry* desenvolveu um conjunto de princípios compartilhados e reguladores que tem por objetivo coordenar os desenvolvedores de ontologias das ciências da vida. Tais princípios encontram disponíveis no BioPortal¹¹ da *NIH Roadmap National Center for Biomedical Ontology (NCBO)*.

Disponíveis, através do BioPortal da NCBO, as ontologias OBO, atualmente, cobrem diversos domínios biomédicos, incluindo ontologias anatômicas, como o FMA, ontologias especializadas da bioquímica, como a *Chemical Entities of Biological Interest (ChEBI)*, ontologias sobre os tipos de células, como a *Cell Ontology (CL)*, além da *Gene Ontology (GO)* sobre produtos genéticos. Atualmente, a *OBO Foundry* compreende aproximadamente 300 ontologias biomédicas (OBO, 2014). A tabela 3, a seguir, lista uma relação de ontologias biomédicas que fazem parte do escopo da OBO, atualmente.

Tabela 3 - Ontologias biomédicas integradas a OBO.

Ontologia	Domínio	Escopo	URL
<i>Gene Ontology</i> ¹² (GO)	Genética	Atributos de produtos genéticos de todos os organismos.	http://www.geneontology.org
<i>Cell Ontology (CL)</i>	Anatomia (Células)	Tipos de células, das procariontas até as mamárias.	http://cellontology.org/
<i>Chemical entities of biological interest (CHEBI)</i>	Bioquímica	Entidades moleculares que são produzidas de produtos sintéticos ou naturais usados para intervir nos processos dos organismos vivos.	http://www.ebi.ac.uk/chebi
<i>Foundational Model of Anatomy (FMA)</i>	Anatomia	Estrutura dos mamíferos e, em particular, do corpo humano.	http://fma.biostr.washington.edu
<i>Phenotypic quality (PATO)</i>	Fenótipos	Qualidades das entidades biomédicas.	http://www.phenotypeontology.org
<i>Protein Ontology (PRO)</i>	Proteínas	Tipos de proteínas e modificações classificadas com base nos relacionamentos evolucionários.	http://pir.georgetown.edu/pro
<i>Zebrafish anatomy and development (ZFA)</i>	Anatomia	Estruturas anatômicas em <i>Danio rerio</i>	http://zfin.org/zf_info/anatomy/dict/sum.HTML

Fonte: adaptado de Smith et al. (2007).

Como projeto de domínio público, a *OBO Foundry* conta com a colaboração dos desenvolvedores de ontologias biomédicas para ampliar seu escopo e, ao mesmo

¹¹ Disponível em <http://obofoundry.org>. Acesso em: 23 de Abril de 2012.

¹² Na estrutura da OBO, a GO é representada pelos seus três vocabulários controlados: (i) processos biológicos; (ii) componentes celulares – domínio anatomia; e (iii) funções moleculares, conforme listados no site da OBO (<http://obofoundry.org>). Acesso em: 04 de Maio de 2012).

tempo, proporciona benefícios a eles ao regular o processo de desenvolvimento e oferecer-lhes outras ontologias que podem ser integradas ao projeto conduzido. No entanto, para que uma ontologia biomédica possa ser integrada a OBO, ela passa por uma avaliação e deve seguir alguns princípios básicos, conforme apresentado por Smith et al. (2007):

- A ontologia deve ser desenvolvida através de um esforço colaborativo;
- Devem ser usadas relações comuns que sejam definidas sem ambiguidade;
- A ontologia deve prever procedimentos para um *feedback* de seus usuários e para identificações de sucessivas versões;
- A ontologia deve ser claramente limitada ao seu sujeito principal, ou seja, todos os termos e relações definidos devem fazer parte do escopo do domínio que a ontologia cobre.

Sobre a estrutura das ontologias OBO, Freitas e Schulz (2009) afirmam que, assim como na GO, os nós de tais ontologias denotam classes de entidades do mundo real e as relações (ou links) entre as classes são interpretadas como links quantificados existencialmente. Assim, por exemplo, a relação *A part_of B* significa que toda ocorrência de A é parte de alguma ocorrência de B, mas não o contrário.

A OBO provê um conjunto expressivo de relações entre as classes de uma ontologia e como forma de regular as relações usadas pelos desenvolvedores em suas ontologias foi desenvolvido a *OBO Relation Ontology* (OBO-RO) que, segundo Smith et al. (2007), contém diretrizes para que os desenvolvedores possam fazer formulações consistentes de suas asserções relacionais. Um exemplo de diretriz é aquela que formaliza a representação de uma mudança anatômica.

Nesta pesquisa, a RO servirá como base de fundamentação para a criação de relações na teoria ontológica do sangue e, diante dessa importância, esta ontologia será descrita com maiores detalhes numa seção seguinte (vide seção 3.2).

Outras duas áreas nas quais a *OBO Foundry* provê diretrizes são: (i) as convenções de nomeação e (ii) os caminhos de representação (SMITH et al., 2007). Na formulação de definições, por exemplo, é utilizado o modelo de boas práticas da FMA, que especifica que a representação de tipos entidades anatômicas deve ser construída em torno de duas hierarquias de relações: a *is_a* e a *whole_part*.

Sobre os formalismos de representação, a *OBO Foundry* dissemina o uso de duas linguagens representativas: (i) um formato específico e patenteado pela OBO, gerado

através de um software especializado em ontologias biomédicas – o OBO-EDIT; e (ii) a linguagem *Ontology Web Language*¹³ *Description Logic*¹⁴ (OWL-DL). Segundo Smith et al. (2007), a maioria das ontologias OBO estão no formato específico gerado pelo OBO-EDIT, enquanto a ontologia sobre técnicas de investigação biomédica – a *Ontology for Biomedical Investigations*¹⁵ (OBI) usa o formato OWL-DL.

¹³ A *Ontology Web Language* (OWL) é a linguagem oficial da W3C para a representação formal de ontologias, capaz de possibilitar que o conteúdo informacional das ontologias possa ser representado e processado pelas máquinas. Sua documentação completa está disponível em: <http://www.w3.org/TR/owl-features/>. Acesso em: 30 de setembro de 2012.

¹⁴ A *Ontology Web Language Description Logic* (OWL-DL) é uma das três sublinguagens que compõem a OWL. Além dela, temos a OWL Lite e a OWL Full.

¹⁵ A *Ontology for Biomedical Investigations* (OBI) é um projeto desenvolvido para integração de ontologias que representam o conhecimento relativo às ciências da vida e às investigações clínicas. Disponível em: <http://obi-ontology.org/>. Acesso em: 30 de setembro de 2012.

4 Ontologias: fundamentos e aplicações

No capítulo anterior foram apresentados os instrumentos de representação e organização da informação, mais relevantes e conhecidos no âmbito da Ciência da Informação. Dentre os instrumentos apresentados, as ontologias são o foco principal do presente estudo.

Desta forma, este capítulo traz um aprofundamento maior sobre o tema “ontologia”, discorrendo sobre dois assuntos importantes: (i) os fundamentos básicos para a compreensão e aplicação de ontologias na resolução de problemas de representação da informação; (ii) a descrição de algumas das ontologias mais usadas e conhecidas atualmente, evidenciando seus fundamentos, princípios e regras utilizados para a organização e representação da informação em diferentes domínios do conhecimento.

Optou-se por estruturar o presente capítulo nas seguintes seções: na seção 4.1, apresentam-se fundamentos ontológicos básicos; na seção 4.2, são descritas ontologias de fundamentação usadas para organizar e representar o conhecimento de domínios genéricos; na seção 4.3, discorre-se sobre ontologias de fundamentação desenvolvidas para aplicação em domínios biomédicos; e na seção 4.4, são descritas ontologias que representam o conhecimento de domínios específicos da biomedicina, tais como anatomia, bioquímica, fisiologia, entre outras, que são relevantes para esta pesquisa.

4.1 Fundamentos ontológicos básicos

Alguns fundamentos ontológicos básicos são necessários para uma adequada compreensão do uso de ontologias na resolução de problemas de representação da informação. Nesse sentido, esta seção da tese trata deste assunto, focando na explicação dos conceitos sobre as categorias ontológicas aristotélicas (quadrado e sexteto ontológico), sobre universais e particulares, além de continuantes e ocorrentes.

O uso de ontologias para representação do conhecimento no contexto da Ciência da Informação e em áreas afins é relativamente recente (SOWA, 1999) (LOWE, 2002), entretanto, os fundamentos ontológicos são antigos e tem origem na ontologia como disciplina filosófica de Aristóteles, que estabelece 10 categorias básicas para descrever os objetos do mundo: *substância (entidades fundamentais), quantidade, qualidade, relação, lugar, tempo, posição, estado, ação, efeito (causa)*.

Uma visão resumida e comum na literatura sobre as categorias ontológicas aristotélicas é dada pelo conhecido quadrado ontológico (veja Tabela 4), cuja referência é importante pois ele representa o esquema categórico frequentemente

usado pelas ontologias de fundamentação para representar a realidade e suas entidades (LOWE, 2002).

Tabela 4 - Quadrado ontológico Aristotélico.

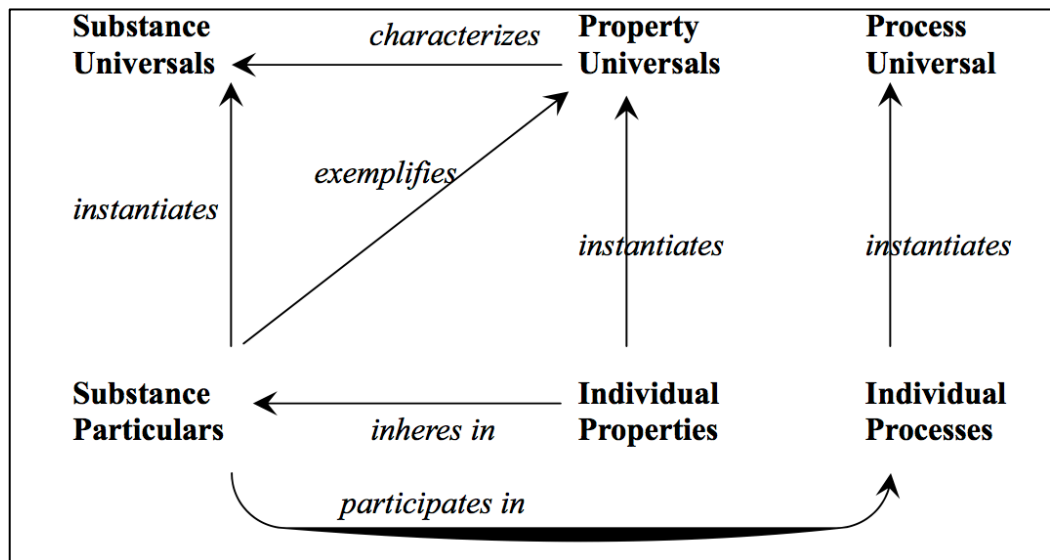
	Substanciais	Acidentais (não substanciais)
Universais	Tipos (por exemplo, Ser humano)	Atributos (por exemplo, Sabedoria)
Particulares (Individuais)	Substâncias (por exemplo, Sócrates)	Modos ou acidentes (por exemplo, Sabedoria de Sócrates)

A principal distinção do quadrado aristotélico dá-se entre entidades que são *universais* e outras que são *particulares*. *Universais* referem-se a tipos de coisas que existem no mundo real e que possuem características em comum, tal que são instanciadas ou exemplificadas por mais de um particular. *Particulares* são também chamados de *instâncias*, *tokens* ou *individuais* e referem-se a um objeto específico ou um indivíduo do mundo real. Uma diferença essencial entre tais tipos de entidades refere-se à predicação em uma declaração lógica: particulares, como Sócrates ou minha altura, não podem ser predicados de um sujeito, já universais, por exemplo “ser humano”, podem aparecer tanto como sujeito (“Ser humano é um animal vertebrado”) ou predicado (“Sócrates é um ser humano”) (MUNN e SMITH, 2008). As ontologias de fundamentação atuais variam na sua forma de incluir ou não esses dois tipos de entidades, por exemplo: a DOLCE (MASOLO et al., 2003) é considerada uma ontologia de particulares, a *Unified Foundational Ontology* (UFO) (GUIZZARDI e WAGNER, 2009) uma ontologia de universais e a *Basic Formal Ontology* (BFO) (GRENON e SMITH, 2004) uma ontologia tanto de universais quanto de particulares.

Outra importante distinção ontológica derivada da teoria aristotélica, embora não tão explícita quanta às entidades contidas no quadrado ontológico, é entre os tipos *continuanes* e *ocorrentes*. *Continuanes* são entidades que persistem ao longo do tempo mantendo sua identidade e que não possuem partes temporais; e *ocorrentes* são entidades que se revelam, se manifestam, ou se desenvolvem ao longo do tempo e possuem partes temporais (GRENON e SMITH, 2004) (SMITH et al., 2005). De maneira geral, ontologias de fundamentação incluem essa distinção entre *continuanes* e *ocorrentes* para descrever o conhecimento da realidade de um dado domínio, tal como faz a BFO. A diferença entre tais ontologias nesse tópico são os nomes usados para descrever esses tipos de entidades: os *continuanes* da BFO são sinônimos de **endurantes** das ontologias DOLCE e UFO e **presenciais** e **persistentes** na General Formal Ontology (GFO) (HERRE, 2010). Já os **ocorrentes** da BFO equivalem aos **perdurantes** da DOLCE e aos **eventos** da UFO.

A partir do quadrado aristotélico é possível deduzir relações ontológicas entre seus campos, tais como: *Acidentes inhere in Substâncias; Atributos characterize Tipos; Substâncias instantiate Tipos*; entre outras (MUNN e SMITH, 2008). Essas relações deduzidas são extremamente importantes para representação do conhecimento, entretanto, apenas com o quadrado aristotélico não é possível representar outras importantes relações ontológicas que ocorrem entre continuantes e ocorrentes. Para tanto, Smith (2005) sugere a extensão do quadrado para um sexteto ontológico (veja Figura 6 abaixo) para abrigar particulares e universais, continuantes e ocorrentes, e as relações entre tais entidades. O sexteto ontológico abrange seis tipos de entidades: *universais substanciais ou tipos, particulares substanciais ou substâncias, universais-propriedade ou atributos, propriedades individuais ou acidentes, universais processos e processos individuais* relacionadas por: *inherence, participation, instantiation, exemplification, characterization*, conforme apresentado na Figura 6. A referência a essas relações do sexteto ontológico é importante, pois foi a partir delas que outras relações foram sendo criadas e incorporadas nas ontologias de fundamentação atuais.

Figura 6 - O sexteto ontológico e suas relações formais-ontológicas.



4.2 Ontologias de fundamentação

Nos últimos anos, é possível perceber o surgimento de um número considerável de ontologias de fundamentação ou de alto nível (*CYC Ontology* (REED e LENAT, 2002); *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE) (MASOLO et al., 2003); *Knowledge Representation Ontology* (SOWA, 1999); *Standard Upper Merged Ontology* (SUMO) (IEEE, 1995); *Basic Formal Ontology* (BFO) (GRENON e SMITH, 2004); *General Formal Ontology* (GFO) (HERRE et al., 2006) e *Unified Formal Ontology* (UFO) (GUIZZARDI e WAGNER, 2009)) com o propósito de orientar os desenvolvedores de ontologias em domínios específicos. Tais ontologias funcionam como meta-ontologias, que podem ser usadas pelos desenvolvedores como o ponto de partida para a criação de ontologias de domínio.

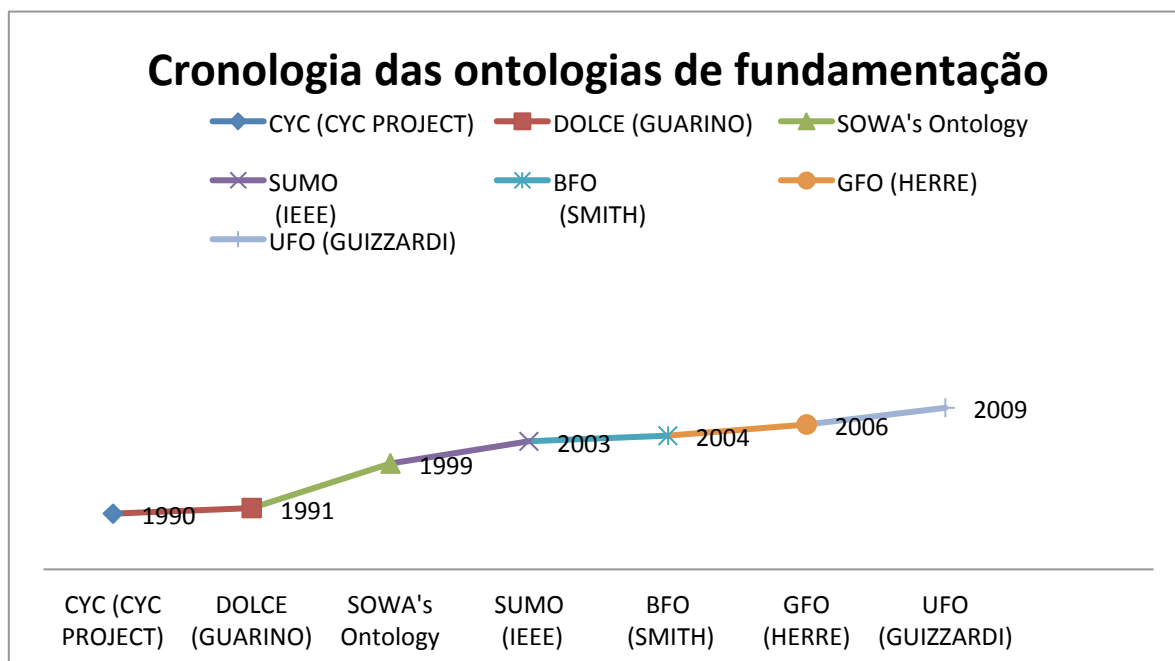
A favor do uso de ontologias de fundamentação, alguns pesquisadores (DEGEN et al., 2001; MASOLO et al., 2003; GRENON e SMITH, 2004; HERRE, 2010) afirmam que a construção de uma ontologia de domínio, tendo como base a estrutura de uma ontologia de fundamentação, é benéfico, se não for necessário. Degen et al. (2001), por exemplo, afirmam que “toda ontologia de domínio deve usar como referência alguma ontologia de nível superior”, de forma que a ontologia de domínio se comprometa a aderir às teorias e princípios propostas na ontologia de nível superior.

Seguindo essa linha de pensamento, torna-se imprescindível, para a presente pesquisa, o estudo das ontologias de fundamentação, principalmente por se tratar de uma pesquisa no contexto biomédico, para o qual o uso de tais ontologias é ainda mais crucial, como citado em Grenon e Smith (2004).

Apesar do objetivo geral das ontologias de fundamentação ser o mesmo, elas discordam em alguns princípios ontológicos, o que as diferem. Cada uma engloba um conjunto de entidades, meta-propriedades, relações e axiomas que servem para especificar ou mesmo limitar as decisões ontológicas do desenvolvedor de uma ontologia de domínio.

Nesta seção apresentam-se algumas das ontologias de fundamentação mais citadas na literatura da área, descrevendo seus princípios e atributos principais. A Figura 7, a seguir, ilustra tais ontologias, apresentadas em ordem cronológica de sua criação.

Figura 7 - Cronologia das ontologias de fundamentação pesquisadas.



Fonte: elaborado pelo autor.

Antes de passar à descrição de tais ontologias de fundamentação, uma observação a ser feita é que além dessas ontologias mencionadas, a revisão de literatura da presente pesquisa inclui uma outra ontologia nessa parte do referencial teórico, que é a *Information Artifact Ontology* (IAO) (IAO, 2012), por considerá-la importante na área de engenharia ontológica. A IAO não foi incluída no gráfico da cronologia das ontologias de fundamentação (Figura 7) pois é considerada uma ontologia de nível médio (*mid-level ontology*). Entretanto, mesmo que a IAO não seja propriamente uma ontologia de fundamentação, ela tem como propósito ser uma ontologia de orientação e suporte aos desenvolvedores da área para a representação de entidades de informação realizáveis, bastante necessário em diversos domínios do conhecimento. Assim, pelos motivos mencionados, incluiu-se tal ontologia nesta parte do referencial teórico.

Passando à descrição das ontologias, destaca-se, inicialmente, a **CYC Ontology**, que surgiu no ano de 1990 e foi desenvolvida como uma representação formal do conhecimento humano, incluindo fatos, regras e heurísticas para deduções sobre objetos e eventos do dia a dia (CYC PROJECT, 2012). Seu objetivo principal é a integração de aplicações computacionais que possuem dados com variações estruturais.

Além desse enfoque computacional, a CYC esforçou-se também em ser uma ontologia unificada que pudesse descrever os objetos do mundo. Para tanto, Reed e Lenat (2002) destacam que muitas ontologias de complexidades variadas foram integradas ou

mapeadas para a CYC. Similar ao propósito da SUMO, o projeto da CYC destaca-se mais como um esforço de unificação e integração de algumas ontologias do que propriamente como uma ontologia de fundamentação, que propõe princípios formais. A diferença principal da CYC para a SUMO é a incorporação de terminologias médicas em sua estrutura. A CYC inclui MeSH, SNOMED, UMLS, além de um tesouro farmacêutico referente ao domínio médico, mas também boa parte da *WordNet*, do *Central Intelligence Agency (CIA) World Factbook*¹⁶ e *Federal Information Processing Standards (FIPS)*¹⁷ 10-4 (REED e LENAT, 2002).

A representação do conhecimento da CYC é operacionalizada por uma linguagem formal denominada de *Cyc Language (CycL)*, um vocabulário de termos e um conjunto de declarações que relacionam os termos. Ao todo, possui 47.000 termos e 306.000 declarações, numa linguagem formal de representação, as quais relacionam esses termos e restringem o seu significado. Essas declarações são agrupadas sem microteorias que representam um domínio particular de conhecimento, um nível específico de detalhes, um intervalo de tempo, etc.

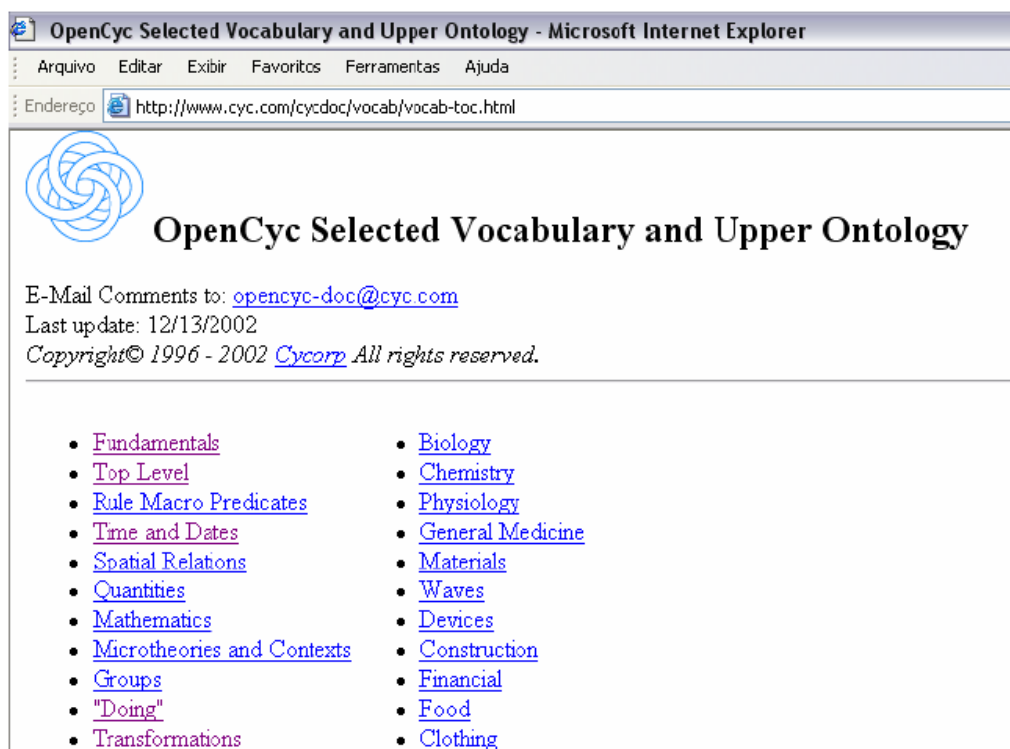
A CYC também possui uma versão de código aberto conhecido como projeto *Open-Cyc*. A partir do site do *Open-Cyc*¹⁸ obtém-se um *software* para instalação do servidor *Cyc*, o qual dá acesso a uma página de vocabulário e a uma interface de busca por termos e relações. A página do vocabulário de termos da *Open-Cyc* é ilustrada na Figura 8, a seguir.

¹⁶ O *World Factbook* provê informações sobre a história, pessoas, governos, economia, geografia, comunicações e transportes de cerca de 267 entidades mundiais, representadas através de mapas físicos, políticos, entre outros (CIA, 2012).

¹⁷ FIPS é um padrão público desenvolvido pelo governo federal dos Estados Unidos para uso em computadores por todas as agências governamentais não militares e por empresas contratadas por esse governo. Muitos padrões FIPS são versões modificadas dos padrões ANSI, IEEE e ISO. Especificamente o FIPS 10-4 é padrão que inclui uma lista de códigos (duas letras) dos países que são usados pelo governo norte-americano para o processamento de dados geográficos em diversas publicações (NIST, 1995).

¹⁸ Disponível em <http://www.opencyc.org>. Acesso em 25 de Abril de 2012.

Figura 8 - Página de vocabulários da Open-CYC.



OpenCyc Selected Vocabulary and Upper Ontology

E-Mail Comments to: opencyc-doc@cyc.com
 Last update: 12/13/2002
 Copyright© 1996 - 2002 Cycorp All rights reserved.

- [Fundamentals](#)
- [Top Level](#)
- [Rule Macro Predicates](#)
- [Time and Dates](#)
- [Spatial Relations](#)
- [Quantities](#)
- [Mathematics](#)
- [Microtheories and Contexts](#)
- [Groups](#)
- ["Doing"](#)
- [Transformations](#)
- [Biology](#)
- [Chemistry](#)
- [Physiology](#)
- [General Medicine](#)
- [Materials](#)
- [Waves](#)
- [Devices](#)
- [Construction](#)
- [Financial](#)
- [Food](#)
- [Clothing](#)

Fonte: site do Open-CYC.

A Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE), criada em 1991, é a ontologia de fundamentação originada do projeto WonderWeb¹⁹, que tem por objetivo principal o desenvolvimento de uma biblioteca de ontologias de fundamentação, denominada de *WonderWeb Foundational Ontologies Library* (WFOL). Nesse projeto, a DOLCE exerce um papel crucial, cujo propósito não é ser uma ontologia padrão “universal”, mas sim um módulo de referência a ser adotado como ponto de partida para comparação e elucidação dos relacionamentos com outros módulos futuros dessa biblioteca, conforme destaca MASOLO et al. (2003).

Quando os autores mencionados utilizam o termo “módulos”, eles querem transmitir a idéia de que a partir da arquitetura WFOL é possível integrar um conjunto (módulo) de ontologias à essa estrutura, fazendo um link através da ontologia DOLCE. Para tanto, a DOLCE está disponível em domínio público, desde sua primeira versão, como forma de estimular outros grupos de pesquisa que trabalham com ontologias formais

¹⁹ WonderWeb é um projeto que se originou na Victoria University of Manchester com o propósito de desenvolver uma biblioteca de ontologias de fundamentação, sistematicamente relacionadas umas com as outras, e capazes de justificar as preferências que fundamentam as escolhas ontológicas, tornando-as mais explícitas possível (MASOLO et al., 2003).

a incorporarem suas ontologias na biblioteca WFOL como módulo independente (MASOLO et al., 2003). A *Object-Centered High-level Reference Ontology*²⁰ (OCHRE) e a BFO são duas conhecidas ontologias incorporadas como módulos externos da arquitetura WFOL.

A idéia de uma arquitetura única de ontologias de fundamentação visa compreender e comparar as diferentes escolhas ontológicas no processo de desenvolvimento, além de harmonizar a documentação sobre elas. No caso das ontologias da WFOL é possível identificar algumas similaridades das escolhas adotadas por elas em razão da linha filosófica que seguem: (i) a distinção entre universais e particulares; (ii) a distinção entre coisas (ou objetos espaciais) e eventos (ou objetos temporais); (iii) a distinção entre entidades abstratas e entidades concretas; (iv) a distinção entre espaço tridimensional e espaço tetradimensional e (v) a distinção entre endurantes (também chamados de continuantes) e perdurantes (também chamadas de ocorrentes).

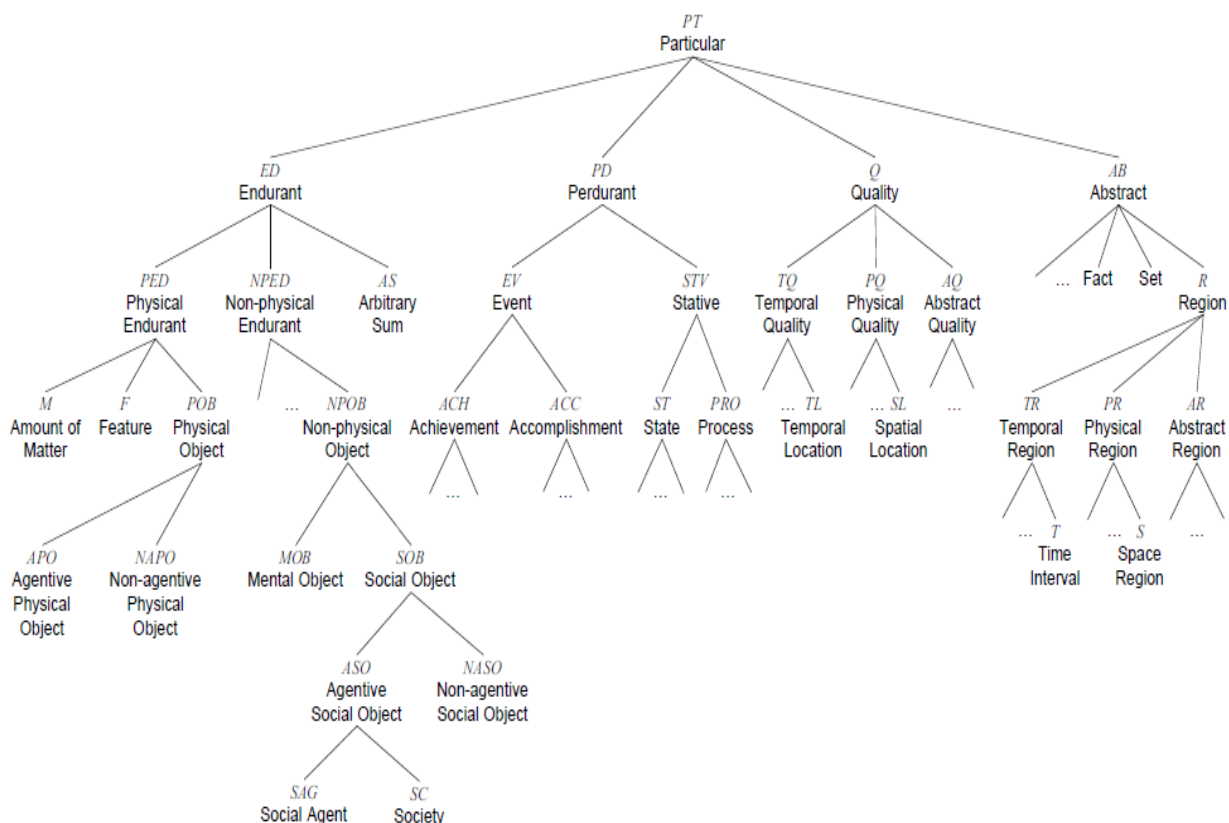
A respeito da ontologia DOLCE, especificamente, ela baseia-se no chamado *viés cognitivo* que, segundo Masolo et al. (2003), que refere-se a uma forma de capturar categorias ontológicas fundamentadas na linguagem natural e no senso comum humano. Desta forma, a DOLCE compreende o termo “categorias”, como artefatos cognitivos estritamente dependentes da percepção humana, inspirado na noção de Searle sobre “background profundo”.

Na DOLCE, a distinção ontológica fundamental entre *universais* e *particulares* é realizada tomando a relação de *instanciação* como uma primitiva. Assim, particulares são entidades que não possuem instâncias e os universais as possuem. Propriedades e relações (correspondentes aos predicados da linguagem lógica) são consideradas universais.

A DOLCE é considerada uma ontologia de particulares e que possui uma rica axiomatização, incluindo ao todo 37 categorias básicas, 7 relações básicas, 80 axiomas, 100 definições e 20 teoremas. A Figura 9, a seguir, apresenta a taxonomia das categorias mais básicas de particulares contidos na DOLCE.

²⁰ A OCHRE é considerada o segundo módulo de ontologias da WFOL. Ela foi desenvolvida por Luc Schneider do Departamento de Filosofia da Universidade de Gênova e difere-se da DOLCE devido ao seu diferente conjunto de suposições, que dizem respeito a uma visão revisionária da noção padrão de senso comum (MASOLO et al., 2003).

Figura 9 - Taxonomia das categorias mais básicas da DOLCE.



Fonte: Masolo et al. (2003).

Na Figura 9, as categorias básicas estão representadas como nós da taxonomia, incluindo seu nome e um acrônimo para descrevê-la, como, por exemplo, *Agentive Physical Object* (APO). Para uma melhor compreensão de tais categorias, a Tabela 5, a seguir, relaciona algumas instâncias para cada *categoria folha* da taxonomia apresentada.

Tabela 5 - Exemplos (instâncias) de categorias básicas da DOLCE

Categorias básicas DOLCE	Exemplos (instâncias)
Abstract Region	O valor convencional de 1 Real.
Agentive Physical Object	Uma pessoa física (ao contrário de uma pessoa jurídica).
Feature	Um buraco, um abismo, uma abertura, um limite.
Non-agentive Physical Object	Uma casa, um computador, um corpo humano.
Non-agentive Social Object	Uma lei, um sistema econômico, uma moeda.
Physical Quality	O peso de uma caneta, a cor de uma maçã.
Process	Correr, escrever, estudar.
State	Ser feliz, ser vermelho, ser aberto.
Temporal Region	O eixo do tempo, 22 de Junho de 2012, um segundo.

Fonte: Masolo et al. (2003).

A DOLCE também inclui um conjunto de relações primitivas básicas, adequadas para caracterizar os compromissos ontológicos da forma mais neutra possível (MASOLO et al., 2003). Essas relações estão formalizadas na DOLCE, utilizando a lógica de primeira ordem, conforme apresentado na Figura 10, a seguir:

Figura 10 - Relações primitivas básicas da DOLCE representadas em FOL.

Parthood: “ <i>x is part of y</i> ” $P(x,y) \rightarrow (AB(x) \vee PD(x)) \wedge (AB(y) \vee PD(y))$
Temporary Parthood: “ <i>x is part of y during t</i> ” $P(x,y,t) \rightarrow (ED(x) \wedge ED(y) \wedge T(t))$
Constitution: “ <i>x constitutes y during t</i> ” $K(x,y,t) \rightarrow ((ED(x) \vee PD(x)) \wedge (ED(y) \vee PD(y)) \wedge T(t))$
Participation: “ <i>x participates in y during t</i> ” $PC(x,y,t) \rightarrow (ED(x) \wedge PD(y) \wedge T(t))$
Quality: “ <i>x is a quality of y</i> ” $qt(x,y) \rightarrow (Q(x) \wedge (Q(y) \vee ED(y) \vee PD(y)))$
Quale: “ <i>x is the quale of y (during t)</i> ” $ql(x,y) \rightarrow (TR(x) \wedge TQ(y))$ $ql(x,y,t) \rightarrow ((PR(x) \vee AR(x)) \wedge (PQ(y) \vee AQ(y)) \wedge T(t))$

Fonte: adaptado de Masolo et al. (2003).

As relações *Parthood* e *Temporary Parthood* são relações do tipo *whole-part*, sendo diferentes quanto à atribuição da relação no tempo: (a) *Parthood* é atemporal, ou seja, *x part_of y* em qualquer instante de tempo; (b) *Temporary Parthood* é indexada no tempo, isto significa que *x part_of y* apenas no instante de tempo *t*.

A relação *Constitution* é derivada da literatura filosófica, onde Doepke (1986) apud Simons (1987) define que “*x constitui y no tempo t se e somente se pode ser um substrato da destruição de y*” e Simons (1987) define que “quando *x constitui y*, há certas propriedades de *x* as quais são acidentais a *x*, mas essenciais para *y*”. Já a relação *Participation* ocorre quando endurantes estão envolvidos (participam) de uma ocorrência num determinado tempo *t*, podendo ser uma *Temporary Participation* ou uma *Constant Participation*. Essa relação é necessária numa ontologia, porque endurantes participantes não são partes de ocorrentes, apenas ocorrentes são partes de ocorrentes.

Por fim, têm-se as relações *Quality* e *Quale* para expressar que uma dada entidade (endurante ou ocorrente) é uma qualidade (ou atributo) de outra entidade. A diferença entre elas está também na questão temporal: *Quality* (atemporal) pode ser compreendida como uma qualidade inerente à entidade durante toda sua existência,

enquanto *Quale* corresponde a um valor de qualidade, que ocorre durante um determinado tempo *t*.

Sobre os formalismos de representação, o conteúdo da ontologia DOLCE foi documentado utilizando tanto a lógica de primeira ordem (FOL-*First Order Logical*), que garante maior expressividade, quanto à linguagem *Knowledge InterchangeFormat* (KIF²¹), que foi usada com o propósito da ontologia DOLCE ser representada e lida por computadores.

Publicada no ano de 1999, a ***Knowledge Representation Ontology*** ou simplesmente ***ontologia de Sowa*** é composta por *categorias básicas* e *distinções* derivadas de fontes variadas da lógica, da filosofia e da inteligência artificial (SOWA, 1999). O termo *distinção* é empregado por Sowa no sentido do termo aristotélico “*differentiae*”, que corresponde ao conjunto de propriedades, de características ou de atributos que distinguem um tipo de outros tipos que têm um supertipo comum (ou gênero). A ontologia de Sowa não é baseada em uma estrutura fixa de categorias, mas em um conjunto de distinções a partir das quais uma hierarquia pode ser gerada automaticamente.

A ontologia de Sowa foi fundamentada em trabalhos de filósofos, principalmente Charles Sanders Peirce e Alfred North Whitehead. Sowa (2000) afirma que o objetivo de sua ontologia é utilizar tais trabalhos para a criação de uma estrutura básica formada por definições e axiomas, que proporcione integração e expansão de bases de conhecimento e de bancos de dados.

Sowa é um dos primeiros pesquisadores sobre ontologias em Ciência da Computação a utilizar uma abordagem filosófica no processo de desenvolvimento das mesmas. Por isso, o registro histórico do seu trabalho é importante como influência para os demais pesquisadores desta área e outras afins, que seguiram essa linha de pensamento.

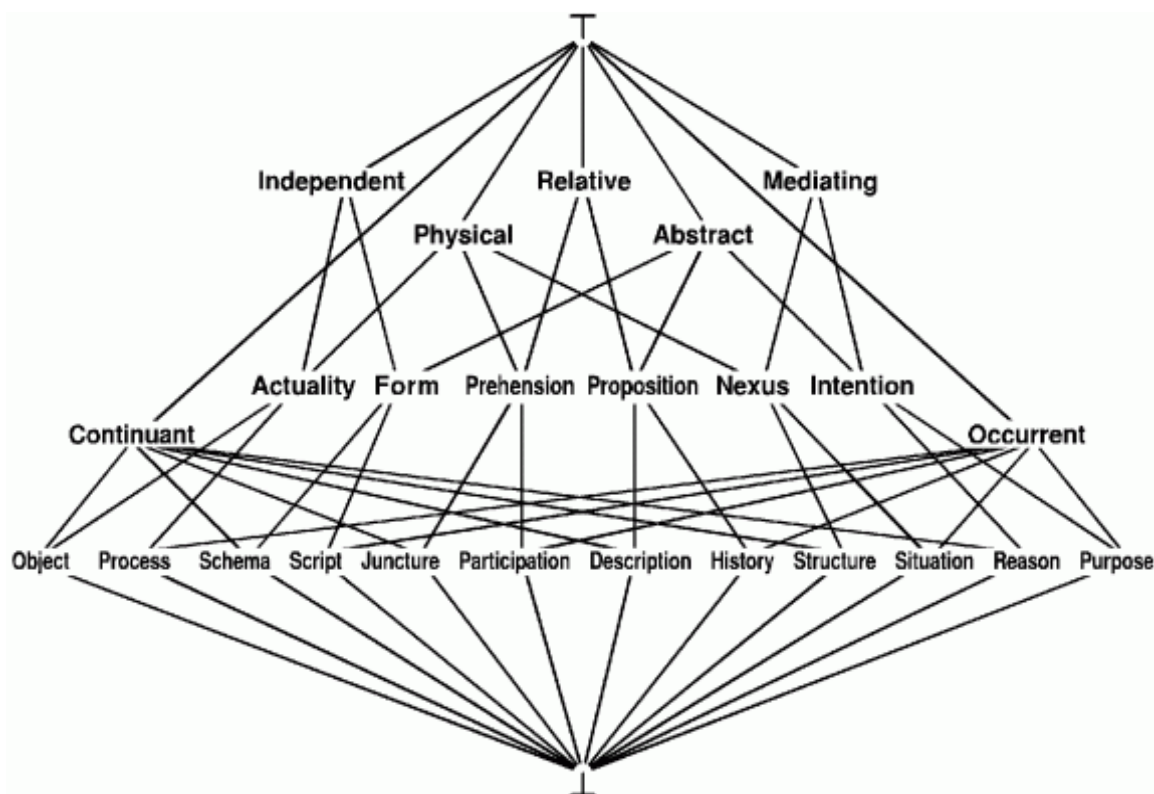
Apesar da origem do termo remeter-se a uma disciplina filosófica, o que se pode ver na prática é a criação de ontologias sem um embasamento filosófico. No entanto, a adoção de princípios formais, advindos da Filosofia, são muito benéficos à criação de ontologias, já que eles funcionam muito bem para distinguir diferenças que são muitos sutis nas decisões ontológicas a serem tomadas no processo de desenvolvimento.

Quanto à estrutura da ontologia de Sowa, ela é composta por níveis denominados *alto-nível*, *processos*, *relações*, *causalidade*, *agentes* e *papéis temáticos*. Cada um dos níveis desta ontologia é formado por um conjunto de termos específicos, usados para a representação e descrição dos objetos de um domínio.

²¹ Disponível em <http://www-ksl.stanford.edu/knowledge-sharing/kif/>. Acesso: 03 de Maio de 2012.

A Figura 11, a seguir, apresenta a estrutura da hierarquia das categorias alto-nível da ontologia de Sowa.

Figura 11 - Hierarquia das categorias alto-nível da ontologia de Sowa.



Fonte: adaptado de Sowa (2000).

Sowa (2000) explica da seguinte forma tal hierarquia:

As categorias são derivadas das combinações dos três caminhos de particionamento (subdivisões) da categoria superior T: Physical or Abstract (P, A); Independent, Relative, or Mediating (I, R, M); Continuant or Occurrent (C, O). Cada uma das outras categorias é um sinônimo da combinação das categorias, das quais foi derivada: Object, por exemplo, pode ser representado pela abreviação PIC, referente à Physical Independent Continuant; e Purpose representado por AMO, referente à Abstract Mediating Occurrent. O símbolo \perp (absurd type), representa a conjunção contraditória de todas as categorias. Ele completa a hierarquia servindo como um subtipo de qualquer outro tipo.

Já a **Standard Upper Merged Ontology (SUMO)** é uma ontologia de fundamentação, que foi disponibilizada pela associação internacional *Institute of Electrical and Electronics Engineers*²² (IEEE), no ano de 2003, com o objetivo de servir como base

²² Disponível em: <http://www.ieee.org>. Acesso em: 25 de Abril de 2012.

para pesquisa sobre busca de informações, sobre linguística e sobre deduções automáticas.

A SUMO é composta por um conjunto de ontologias bem diversificado, que incluem a própria ontologia SUMO, a *MidLevel Ontology (MILO)*, além de ontologias de comunicação, de países e regiões, de economia e finanças, de engenharia, de geografia, de pessoas, de transportes, entre outras. Essa característica da SUMO provoca questionamentos justificados sobre sua natureza de ontologia de fundamentação, já que inclui ontologias de domínio. Mesmo sabendo de tal fato, optou-se por descrevê-la nessa parte do trabalho, por considerar que ela representou uma iniciativa de cobrir o maior número possível de domínios do conhecimento, através da unificação de diversas ontologias para descrever os objetos do mundo real. No entanto, sabe-se que a SUMO não pode ser considerada uma ontologia de fundamentação em sua essência.

Seu conteúdo engloba aproximadamente 20.000 termos e 60.000 axiomas, que são representados formalmente através de uma linguagem, derivada da *Knowledge Interchange Format (KIF)* (GENESERETH e FIKES, 1992), denominada *SUO-KIF*. Essa linguagem está integrada a uma ferramenta denominada *Sigma Knowledge Engineering Environment*, a qual permite buscas por termos da ontologia e buscas por termos no léxico *WordNet*²³ (MILLER, 1995). Ao se realizar a busca, a ferramenta faz a correspondência entre os termos da *SUMO* e os termos da língua inglesa (substantivos, verbos, adjetivos, pronomes, etc.), apresentando os resultados na tela.

Como forma de exemplificar a utilização da ontologia SUMO, são apresentadas, a seguir, as Figuras 12 e 13, que ilustram a interface de busca da SUMO. Na Figura 12 é realizada uma busca na ontologia SUMO pelo termo “record”, utilizando o campo “*Knowledge-Base Term (KB Term)*”, e na Figura 13, é feita uma busca ao termo “system”, qualificado como substantivo, no léxico *WordNet*, utilizando o campo “*English Word*”. O resultado dessa última busca, retorna as relações do termo com a ontologia SUMO.

²³ Um léxico para a língua inglesa, disponível na Internet em <http://wordnet.princeton.edu/>, em que são organizados setenta mil grupos de sinônimos, cada um representando um conceito. Os sinônimos são conectados por relações e o léxico é composto por substantivos, verbos, adjetivos e advérbios.

Figura 12 - Busca pelo termo "record" na interface da SUMO.

The screenshot shows a web browser window titled "Knowledge base Browser - Microsoft Internet Explorer". The address bar contains the URL: <http://sigma.ontologyportal.org:4010/sigma/Browse.jsp?kb=SUMO&lang=en&term=record>. The page header includes the Sigma logo and the text "Sigma knowledge engineering environment Browsing Interface". Navigation links include [Home | Ask/Tell | Graph | Prefs]. The KB is set to SUMO and the Language to en. The search interface shows "KB Term: record" and "English Word: Noun". Below the search results, the term "record" is displayed in large font, followed by a list of related terms: [RadiatingNuclear](#), [RadiatingSound](#), [Radio](#), [RadioBroadcasting](#), [RadioEmission](#), [RailroadCompany](#), [postalCode](#), [precondition](#), [prefers](#), [premise](#), [prevents](#), and [price](#).

Fonte: site da IEEE.

Figura 13 - Busca pelo termo "system" na Wordnet e suas relações com a SUMO.

The screenshot shows a web browser window titled "SUMO Search Tool - Microsoft Internet Explorer". The address bar contains the URL: <http://sigma.ontologyportal.org:4010/sigma/WordNet.jsp?word=system&POS=l>. The page header includes the Sigma logo and the text "Sigma knowledge engineering environment". A [Home] link is visible. The main heading is "SUMO Search Tool". Below the heading, a paragraph states: "This tool relates English terms to concepts from the [SUMO](#) ontology by means of mappings to [WordNet](#) synsets." The search interface shows "English Term: system" and "Noun". Below the search results, a paragraph states: "According to WordNet, the noun 'system' has 9 sense(s)." The first sense is described as "1. a group of independent but interrelated elements comprising a unified whole; 'a vast system of production and distribution and consumption keep the country going'." It lists a SUMO Mapping: [Collection](#) (subsuming mapping). The second sense is described as "2. a combination of interrelated interacting artifacts designed to work as a coherent entity; 'he bought a new stereo system' 'the unit consists of a motor and a small computer'." It lists a SUMO Mapping: [Device](#) (subsuming mapping).

Fonte: site da IEEE.

Outra ontologia incluída na cronologia das ontologias de fundamentação pesquisadas (vide figura 6) é a **Basic Formal Ontology** (GRENON e SMITH, 2004), cuja primeira versão (1.0) foi publicada em 2004. Já que esta ontologia foi desenvolvida para aplicação em domínios biomédicos, sua descrição, nesta pesquisa, é feita na seção seguinte (seção 4.2).

Por sua vez, a **General Formal Ontology (GFO)**, criada no ano de 2006, é um dos componentes do *Integrated System of Foundational Ontologies* (ISFO), que foi projetado para ser parte do *Integrated Framework for Development and Application of Ontologies* (IFDAO) e que é o predecessor do projeto GOL lançado, em 1999, pela Universidade de Leipzig (HERRE et al., 2006).

De forma resumida, a GFO é uma ontologia de fundamentação que integra objetos e processos. Segundo Herre (2010), a GFO possui uma arquitetura meta-ontológica formada por três camadas: (i) um nível básico, que inclui todas as categorias e relações GFO relevantes; (ii) um meta-nível, chamado de nível nuclear, que contém as meta-categorias²⁴ sobre o nível básico; e (iii) um nível superior, que inclui as chamadas meta-meta-categorias. Este último nível, também conhecido como *abstract top ontology* (ATO), contém duas meta-categorias denominadas de set e item, e é usado para expressar e modelar os níveis inferiores da GFO através de um conjunto teórico de expressões. Logo abaixo do ATO, tem-se um nível não formal (o meta-nível) que pode ser também chamado de nível filosófico, já que inclui assunções filosóficas básicas, tais como as noções de existência e dependência.

Originalmente desenvolvida após uma ampla revisão do projeto original GOL, pelo grupo de pesquisa *Ontologies em Medicina* (OntoMed), a GFO possui as seguintes características, segundo Herre (2010): (i) inclui objetos (objetos 3D) e processos (entidades 4D) integradas num framework coerente; (ii) possui uma abordagem multicategorial, admitindo universais, conceitos e suas inter-relações, além de símbolos estruturais; (iii) inclui diferentes tipos de realidade; (iv) contém um conjunto de axiomas formais, um módulo de funções e outro módulo de papéis e (v) foi desenvolvida para aplicações, principalmente, em áreas biomédicas, mas também em domínios da economia e da sociologia.

Segundo seus criadores (Herre et al., 2006), o propósito principal da GFO não é ser a ontologia de fundamentação de referência para o desenvolvimento de outras

²⁴ A noção de meta-categoria é uma generalização da noção de meta-conjunto ou meta-classe no sentido de um conjunto teórico (HERRE, 2010).

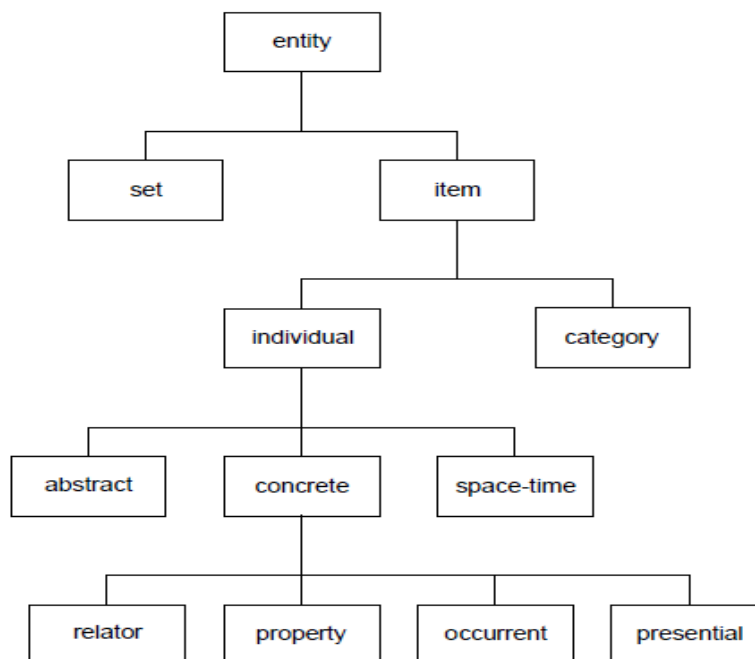
ontologias de domínio, mas sim ser parte de um sistema evolutivo de ontologias de fundamentação – o ISFO – capaz de integrar diferentes ontologias de fundamentação, com suas categorias, relações e axiomas específicos. Essa abordagem adotada pela GFO, baseia-se no fato de que, na opinião de seus criadores, diferentes visões do mundo podem ser sustentadas e acredita-se que um número reduzido delas seja realmente útil para representar o conhecimento humano.

Na realidade, a GFO distingue pelo menos três estratos ontológicos do mundo: (i) o material; (ii) o mental-fisiológico; e (iii) o estrato social, conforme destacado por Herre (2010). Segundo tal distinção, toda entidade do mundo participa de um certo estrato e nível, que é caracterizado por sistemas integrados de categorias. Um nível é especificado por uma categoria meta-nível e também inclui individuais (ou particulares).

Outra característica da GFO é que esta ontologia fundamenta-se no realismo como sua corrente filosófica, mas um tipo de realismo conhecido como realismo integrativo, que se difere do realismo adotado por Barry Smith e que fundamenta a BFO. Na visão de Herre et al. (2006), não existe a representação da realidade sem o uso de conceitos. Por isso, tais autores apresentam críticas ao realismo adotado por Smith e seus seguidores. Essa discussão filosófica ultrapassa o escopo deste trabalho e não será debatida aqui.

Para ilustrar as categorias GFO, foi extraído do trabalho de Herre et al. (2006) um diagrama das categorias mais básicas dessa ontologia (veja Figura 14). Ele representa o ponto de início para a modelagem de domínios do conhecimento utilizando a abordagem GFO.

Figura 14 - Diagrama das categorias mais básicas da GFO.



Fonte: adaptado de Herre et al. (2006)

Um registro final que merece ser destacado sobre a GFO é o trabalho realizado por seus autores (Herre et al., 2006) em comparar essa ontologia com duas outras ontologias de fundamentação (DOLCE e ontologia de Sowa), através de um mapeamento entre suas categorias.

A Tabela 6, a seguir, apresenta o mapeamento das categorias DOLCE para GFO. Tal mapeamento foi feito com base em especificações textuais e não pode ser considerado um mapeamento formal entre essas ontologias. Os parênteses na coluna direita indicam um nível inferior de adequação de um mapeamento particular.

Tabela 6 - Mapeamento das categorias DOLCE para GFO

DOLCE	GFO
Particular	(Individual)
Endurant	(Presential, Persistent)
Physical Endurant	Material Structure
Amount of Matter	Amount of Substrate
Feature	(Material Boundary)
Physical Object	Material Object
Agentive Physical Object	--
Non-agentive Physical Object	--
Non-physical Endurant	(Levels)

Non-physical Object	--
Mental Object	(Concept)
Social Object	--
Agentive Social Object	--
Social Agent	(Social Role)
Society	--
Non-agentive Physical Object	--
Perdurant	(Occurent)
Event	(Change)
Achievement	(Achievement)
Accomplishment	(Accoplishment)
Stative	Process
State	State
Process	--
Quality	Property
Temporal Quality	--
Temporal Location	--
Physical Quality	--
Spatial Location	--
Abstract Quality	--
Abstract	(Space-Time-Entity U Set U Fact)
Fact	Fact
Set	Set
Region	(Space-Time-Entity), (Measurement System)
Temporal Region	Time-Region
Time Interval	Chronoid
Physical Region	--
Space Region	Space Region
Abstract Region	--

Fonte: adaptado de Herre et al. (2006).

Já o mapeamento das categorias da ontologia de Sowa para a GFO é apresentado na Tabela 7, a seguir.

Tabela 7 - Mapeamento das categorias da ontologia de Sowa para a GFO

Sowa's Ontology	GFO
Entity	Entity
Independent	(Situoid)
Physical	Individual (intuitively): Presential (by axioms)

Relative	(Universal defined by a relational role universal)
Abstract	Category
Mediating	Relation
Continuant	(Presential U Persistent)
Occurent	Occurent
Actuality	(Material structure U Process)
Form	(Category of material structures or process)
Prehension	(Material structure in a role)
Proposition	(Instantiation)
Nexus	(Relator or foundation of a relator)
Intention	--
Object	Material structure U Persistent
Process	Process
Schema	Category of material structures
Script	Category of processes
Juncture	(Relational role)
Participation	(Processual role)
Description	Symbolic structure of material structures
History	Symbolic structure of histories
Situation (Structure)	(Material structure as a foundation of a relator)
Execution (Situation)	(Process as a foundation of a relator)
Reason	--
Purpose	--

Fonte: adaptado de Herre et al. (2006).

É importante destacar ainda que o mapeamento das ontologias (DOLCE para GFO; ontologia de Sowa para GFO), realizado por Herre et al. (2006), também inclui o caminho inverso do mapeamento (GFO para DOLCE; GFO para ontologia de Sowa) e outros detalhes comparativos, que extrapolam os objetivos pretendidos nesta pesquisa.

Outra ontologia de fundamentação, descrita nesta revisão de literatura, é a **Unified Foundational Ontology (UFO)**. Tal ontologia foi criada em 2009, como um projeto de unificação da GFO e da ontologia de alto nível de universais subjacente a OntoClean²⁵. A UFO foi desenvolvida com o propósito principal de prover uma teoria para resolver os problemas clássicos de modelagem conceitual. Para tanto, seu desenvolvimento baseou-

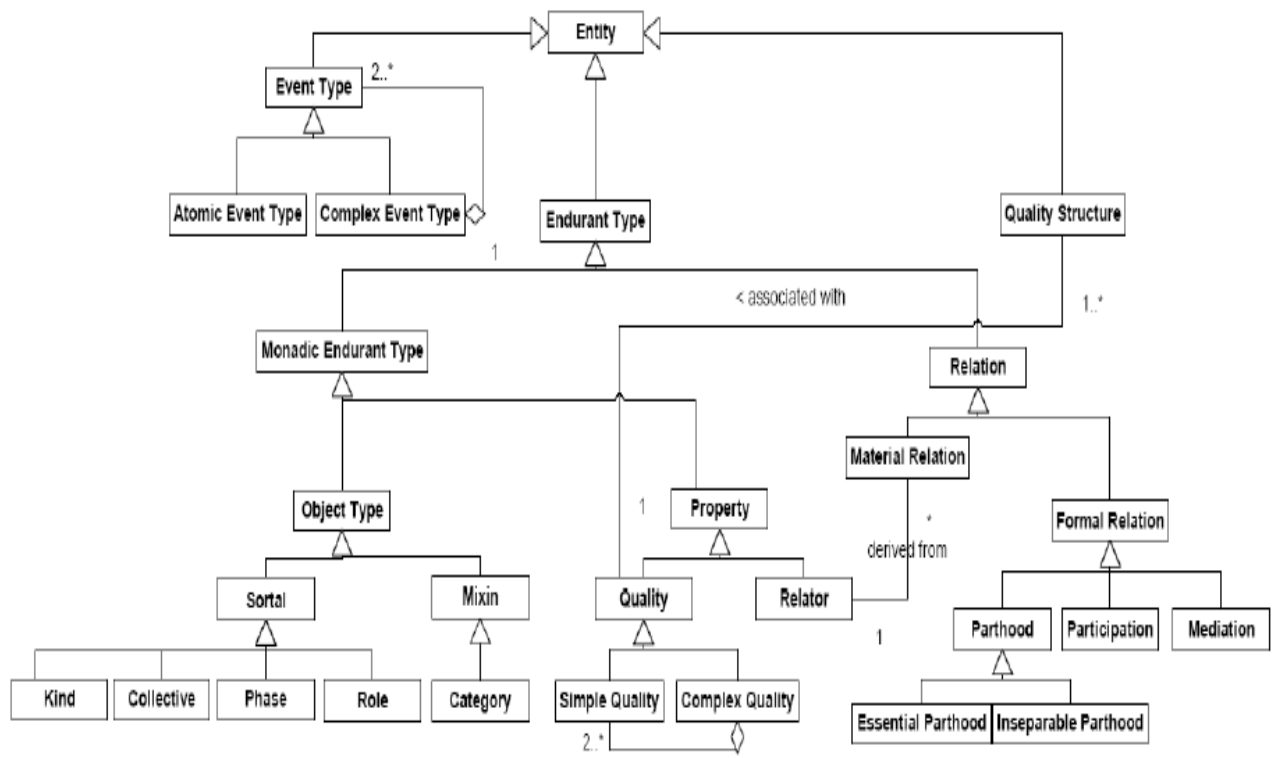
²⁵ OntoClean é uma metodologia de avaliação de ontologias, proposta por Guarino e Welty com o propósito de identificar e corrigir falhas na estrutura de ontologias, especialmente pela avaliação do “eixo principal” da taxonomia (*backbone taxonomy*), formada pelos relacionamentos de subsunção. Disponível em: <http://www.ontoclean.org>. Acesso em: 28 de Abril de 2012.

se numa série de teorias da Ontologia Formal, da Lógica Filosófica, da Filosofia de Linguagem, da Linguística e da Psicologia Cognitiva (GUIZZARDI e WAGNER, 2009).

A UFO inclui teorias sobre partes e todos, tipos, estruturas taxonômicas, relacionamentos, atributos e atributos de espaço, papéis e qualidades individuais, além de outras coisas (GUIZZARDI, 2005 apud GONÇALVES, 2009). Mais recentemente, a UFO foi expandida para tratar entidades dinâmicas como processos, eventos com entidades sociais indexadas no tempo como ações, agentes, intencionalidade, dependência social e delegação (GUIZZARDI, FALBO e GUIZZARDI, 2008).

Do ponto de vista filosófico, a UFO é uma ontologia de universais e não de particulares (GONÇALVES, 2009). Assim como em outras ontologias de fundamentação, a distinção básica da UFO é entre entidades *endurantes* ou entidades *perdurantes (eventos)*, que refletem a distinção do senso-comum entre *objetos* e *processos*. A Figura 15, a seguir, ilustra a estrutura taxonômica da UFO que inclui a coleção de categorias básicas de tal ontologia.

Figura 15 - Representação esquemática das categorias básicas da UFO.



Fonte: adaptado de Guizzardi, Falbo e Guizzardi (2008).

A Figura 15 mostra que existem dois tipos de entidades endurantes: as *entidades monádicas* e as *relações*. As *entidades monádicas*, por sua vez, podem ser tanto *objetos* quanto *propriedades* (entidades cuja existência depende da existência de

outra entidade, como, por exemplo, a cor de uma maçã ou o sintoma de um paciente). Entretanto, entidades como cores são distinguidas de entidades como sintomas pela rubrica de *qualidade* e *modo*. Em conformação com a DOLCE, a UFO distingue entre a cor de um particular (sua qualidade) e seu valor, por exemplo, uma tonalidade particular do vermelho. Este último exemplo é chamado de *quale*, assim como na DOLCE, e descreve a posição de uma qualidade individual com uma certa dimensão de qualidade, baseada na *teoria de espaços conceituais* do cientista cognitivo e filósofo sueco Peter Gardenfors, conforme afirma Gonçalves (2009).

Ainda com relação à estrutura taxonômica da UFO, mostrada na Figura 15, *momentos relacionais* ou *relatores* são tipos de propriedades que expressam uma conexão material entre entidades (exemplos: um casamento, uma ligação covalente, um emprego). Assim, os relatores fundamentam as *relações materiais*, que necessitam deles para serem estabelecidas. Exemplos de relações materiais são: *ser_casado_com*, *ser_conectado_a*, *trabalha_em*. Ao contrário das relações materiais, as *relações formais* ocorrem diretamente entre particulares. *Todo-parte*, *participação* e *mediação* (um tipo de dependência existencial) são tipos de relações formais. A relação formal de *participação*, por exemplo, é definida entre endurantes e eventos.

Na taxonomia da UFO, temos dois tipos básicos de objetos: *sortals* e *mixin*. Os tipos *sortals* incluem as substâncias que fornecem o princípio da identidade para suas instâncias, chamadas de *kind* (para objetos gerais) e *coletivo* (para coleções de entidades), além dos tipos de objetos que instanciam os objetos em algumas circunstâncias, mas não em outras, denominados de *fases* (quando instanciam o objeto em dado período de tempo e não necessariamente em todos os períodos, como, por exemplo, *Vivo* e *Falecido*) e *papéis* (que instancia um objeto apenas num dado contexto, como, por exemplo, *Médico* e *Estudante*). O tipo *mixin* engloba as *categorias* que, na definição da UFO, são usadas para classificar entidades que pertencem a tipos diferentes, mas que compartilham uma propriedade essencial comum como, por exemplo, *Entidade Racional*.

Ainda sobre a UFO, Guizzardi e Wagner (2009) destacam que suas categorias ontológicas têm sido usadas para o desenvolvimento de um perfil UML, ontologicamente bem fundamentado, chamado de OntoUML. As primitivas de modelagem da OntoUML representam diretamente as distinções ontológicas postuladas pela UFO (GUIZZARDI et al., 2004). Além disso, a axiomatização da UFO, que regula como as categorias ontológicas podem ser combinadas, é incorporada na linguagem deste meta-modelo, como restrições formais. O resultado dessa abordagem, é que se torna possível verificar se as especificações produzidas em OntoUML estão gramaticalmente corretas, segundo a axiomatização prescrita na UFO. Para facilitar esse trabalho, Benevides e Guizzardi

também desenvolveram uma ferramenta capaz de analisar os modelos conceituais especificados em OntoUML e apresentar ao usuário as falhas ontológicas desses modelos (BENEVIDES e GUIZZARDI, 2009 apud GONÇALVES, 2009). Resumidamente, pode-se dizer que o objetivo do uso da OntoUML para modelagem de domínios é restringir a construção de modelos conceituais àqueles ontologicamente consistentes e bem fundamentados.

Por fim, temos a ontologia **Information Artifact Ontology (IAO)**, a qual foi incluída nesta parte de revisão da literatura por representar uma ontologia capaz de orientar os desenvolvedores em suas decisões ontológicas, tais como as ontologias de fundamentação. A IAO é uma ontologia de nível médio (*mid-level ontology*) para representação de entidades de informação realizáveis, isto é, ela possibilita a representação das diferentes maneiras com as quais a informação relaciona-se com o mundo real.

Atualmente integrada ao repositório de ontologias da OBO, a IAO é uma ontologia de entidades de informação, originalmente desenvolvida para o trabalho com entidades digitais da *Ontology for Biomedical Investigations (OBI)*. A IAO é considerada uma descendente do ramo BRANCH, de entidades de informação realizáveis, do Projeto OBI (OBO, 2014).

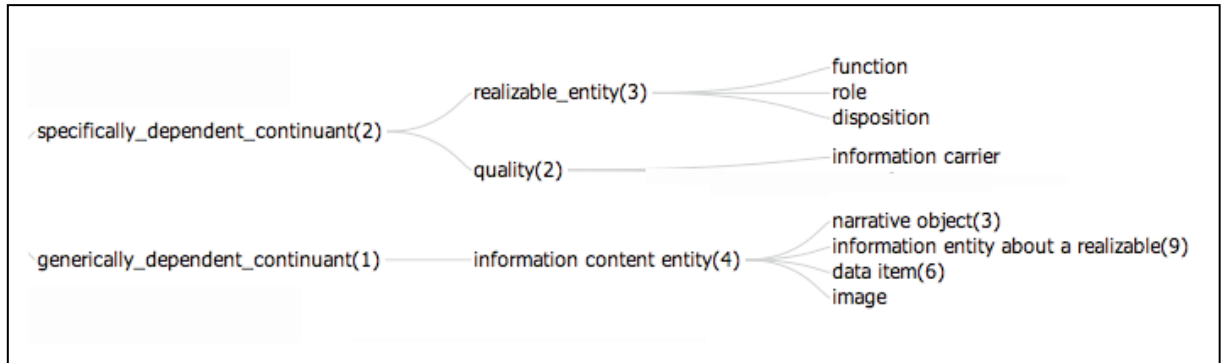
Historicamente, a IAO teve sua primeira apresentação à comunidade científica em um *workshop*, realizado no *MIT Stata Center* em Boston, Estados Unidos, em Junho de 2008 (IAO, 2012). O objetivo de seus criadores foi propor uma estrutura hierárquica para integração de dados em larga escala na *web* no domínio de “informação”. Nesse sentido, a IAO se propõe auxiliar os desenvolvedores de ontologias em suas escolhas de representação referentes às entidades de informação realizáveis.

Considerando o escopo coberto pela IAO, tal ontologia engloba os seguintes tipos de entidades: (i) entidades de conteúdo informacional, tais como: *report, journal article content, narrative object, specifications, serial numbers*; (ii) processos que consomem ou produzem entidades de conteúdo informacional, tais como: *writing, documenting, recording, measuring, encoding*; (iii) portadores materiais de informação, tais como: *books, bound journals, photographic prints, CDs*; e (iv) relações que estão relacionadas com as entidades de conteúdo informacional, que incluem: *is_about, denotes, is_measurement_of, encodes, is_topic_of, is_rendering_of*.

Como forma de ilustrar como tais entidades estão dispostas nas taxonomias da IAO, as Figuras 16 e 17, a seguir, apresentam partes das taxonomias da IAO, que correspondem ao ramo das entidades de informação realizáveis. Note que os números

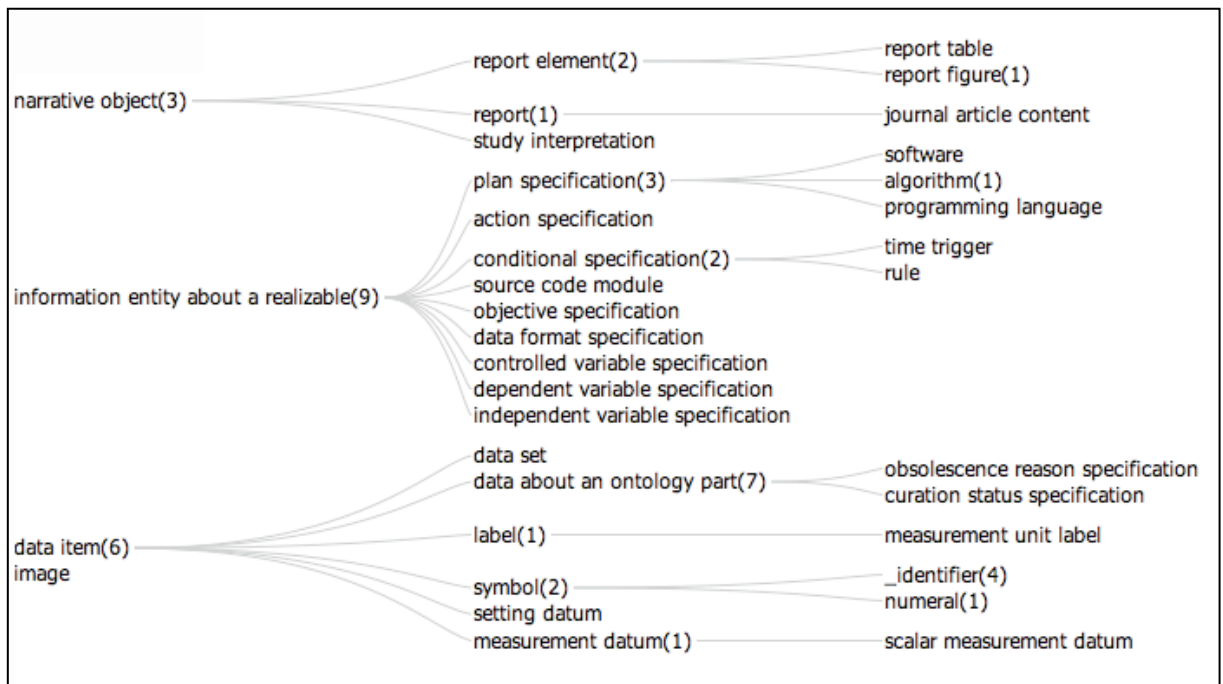
entre parênteses de cada entidade indicam o número de nós filhos que cada nó da taxonomia possui.

Figura 16 - Parte da taxonomia da IAO contendo entidades dependentes continuantes.



Fonte: adaptado de Ruttenberg (2009).

Figura 17 - Parte mais detalhada da taxonomia da IAO contendo dependentes continuantes.



Fonte: adaptado de Ruttenberg (2009).

O projeto da IAO não é uma exclusividade na literatura da área, na qual é possível identificar também outros esquemas/padrões para representação de informação bibliográfica, tais como o Medline, MARC, SWAN, Bibo e Cito. Entretanto, a IAO apresenta algumas diferenças em relação a esses outros esquemas, que podem ser vistas como vantagens desta ontologia (RUTTENBERG, 2009): (i) a IAO representa em sua estrutura

conexões entre as informações e as coisas que elas representam; (ii) a IAO está integrada ao framework da *OBO Foundry*, que reúne boa parte das ontologias biomédicas mais usadas atualmente; e (iii) a IAO pode ser capaz de prover integração de diferentes esquemas de informação bibliográfica, como, por exemplo, esquemas de codificação.

Quanto à aplicabilidade prática da IAO na área de engenharia ontológica, seu uso está relacionado, principalmente, com a representação de investigações biomédicas, possibilitando a representação das diferentes maneiras com as quais a informação médica relaciona-se com o mundo real, justificando, assim, sua integração ao repositório OBO. Alguns projetos conhecidos que fazem uso da ontologia IAO são: o projeto OBI; a *Oral Health and Disease Ontology*²⁶; a *Ontology for General Medical Science*²⁷ (OGMS); a *Ontology for Drug Discovery Investigations*²⁸ (DDI); o *Neural ElectroMagnetic Ontologies*²⁹ (NEMO), entre outros (IAO, 2012).

4.3 Ontologias de fundamentação para biomedicina

Apesar das ontologias biomédicas representarem poderosos recursos para a organização da informação médica, elas ainda carecem do uso de princípios formais no seu processo de desenvolvimento. Smith (2005); Bittner e Donnelly (2007) apontam para essa necessidade de uso de princípios formais, como forma de manter explícitas, o quanto possível, as definições das classes e relações contidas nas ontologias biomédicas.

Diante dessa constatação, torna-se imprescindível dar a devida importância às ontologias de fundamentação usadas como ponto de partida e referência no desenvolvimento de ontologias biomédicas de domínios especializados.

Na presente seção, apresentam-se algumas ontologias de fundamentação, atualmente muito utilizadas na representação de domínios biomédicos. Optou-se por apresentar ontologias de fundamentação consideradas relevantes para o domínio sob estudo e, conseqüentemente, para a pesquisa aqui conduzida. São elas: a *Basic Formal Ontology* (BFO); a *Relation Ontology* (RO) e a *Biological Top-Level Ontology* (BIOTOP).

Tratada como uma teoria que fundamenta o desenvolvimento da ontologia sobre o sangue humano proposta, a primeira ontologia de fundamentação para biomedicina apresentada é a ***Basic Formal Ontology (BFO)***.

O projeto da ***Basic Formal Ontology (BFO)*** iniciou-se no ano de 2002, desenvolvido pelo *Institute for Ontology and Medical Information Science* (IFOMIS), na Universidade de Leipzig, Alemanha (GRENON, SMITH e GOLDBERG, 2004). Desde então,

²⁶ Disponível em: <http://code.google.com/p/ohd-ontology/>. Acesso em: 27 de Setembro de 2012.

²⁷ Disponível em: <http://code.google.com/p/ogms/>. Acesso em: 27 de Setembro de 2012.

²⁸ Disponível em: <http://purl.org/ddi/home>. Acesso em: 27 de Setembro de 2012.

²⁹ Disponível em: <http://nemo.nic.uoregon.edu/>. Acesso em: 27 de Setembro de 2012.

a utilização desta ontologia vem aumentando em pesquisas científicas sobre domínios biomédicos, como pode ser percebido por um simples busca do termo “BFO” na *web*.

Com o projeto da BFO iniciado em 2002, a “ontologia em si” teve sua primeira publicação em 2004 na versão 1.0 e, nos anos seguintes, tivemos a incorporação da *Open Biomedical Ontologies-Relation Ontology* (OBO-RO) em sua estrutura (2005), e a publicação da versão BFO 1.1 (2006), de acordo com informações do portal onde é mantido a ontologia – o IFOMIS³⁰. Atualmente, o criador da teoria em que a BFO se baseia - Barry Smith – e seu grupo de pesquisa trabalham na versão 2.0, que está em fase de publicação.

A presente pesquisa fará uso da BFO 2.0 no desenvolvimento da teoria ontológica do sangue humano, utilizando-se de seus princípios de formalização e de suas entidades e relações.

Tendo sofrido influências dos trabalhos de descrição do conhecimento de Aristóteles (1998), por meio da metafísica, da lógica de Edmund Husserl (1969) e das ontologias DOLCE, GO e FMA; a BFO tem como base filosófica o realismo ontológico (GRENON e SMITH, 2004).

No entanto, apesar de ter nascido numa corrente filosófica que coincide com as ontologias DOLCE e SUMO, ao contrário destas, a BFO se propõe a ser uma ontologia estritamente formal ou de nível superior, com o propósito principal de dar suporte ao desenvolvimento de ontologias de domínio, especialmente na área biomédica. Nesse sentido, a BFO auxilia nas tomadas de decisão ontológicas do desenvolvedor e ajuda também a tornar o mais claro possível esse processo de escolha. Por ser considerada uma ontologia estritamente formal, termos específicos dos domínios físicos, químicos ou biológicos não estão na BFO, mas sim categorias fundamentais da realidade, sendo seu domínio é neutro.

Além de uma base filosófica realista, a BFO foi desenvolvida a partir de princípios filosóficos do *adequalismo*, *falibilismo* e *perspectivismo* (GRENON, SMITH e GOLDBERG (2004) e SPEAR (2006)). Essa orientação filosófica da BFO pode-se ser facilmente identificada em seus princípios, conforme apresentado por esses autores:

- Como toda ontologia de origem **realista**, tais como DOLCE e SUMO, a interpretação pretendida para as categorias fundamentais e as relações da ontologia é de que as divisões reais entre os tipos de entidades que existem no mundo são independentes da mente humana.

³⁰ Disponível em <http://www.ifomis.org/bfo>. Acesso em 17 de Março de 2012.

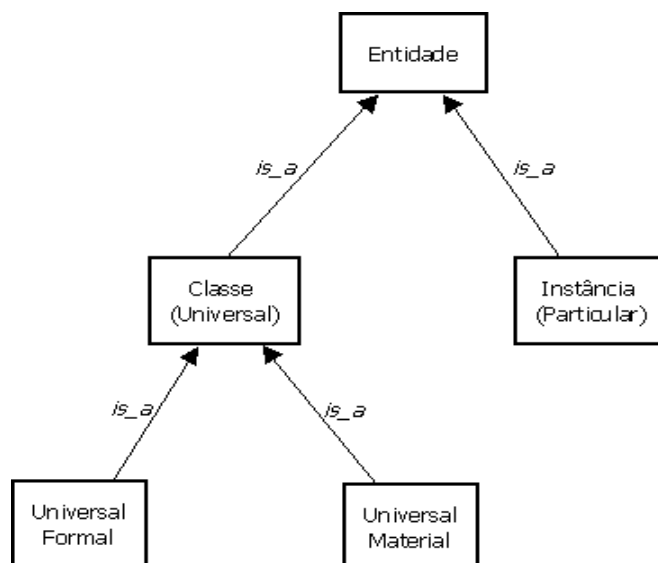
- **Adequalismo** refere-se ao fato de que a BFO foi desenvolvida para permitir a representação direta dos vários tipos de entidades da realidade, tais como as partículas fundamentais da física ou os procedimentos cirúrgicos.
- O princípio do **falibilismo** implica no reconhecimento de que alguns pressupostos básicos da ontologia podem ser revistos ao longo do tempo devido aos achados das pesquisas na área.
- Por fim, o **perspectivismo** defende a idéia de que nem todas as perspectivas ou formas de divisão da realidade são boas, porém há muitas boas maneiras de se representar a realidade usando perspectivas diferentes. Um exemplo simples, citado por Spear (2006), é analisar um organismo vivo, como o rinoceronte, sob a perspectiva de suas características genéticas – material e código genético.

Inserida nessa abordagem filosófica, a ontologia BFO inclui um conjunto de *categorias fundamentais*, também referenciadas como *universais*, e de *relações* entre essas categorias, que servem de suporte aos desenvolvedores de ontologias de domínio no processo de construção.

Categorias e relações somente são incluídas na BFO caso cumpram as restrições especificadas na teoria que fundamenta tal ontologia, cujo conteúdo completo pode ser encontrado em Grenon e Smith (2004). A especificação de tais restrições é realizada utilizando o *inglês semi-formalizado*, que se aproxima bastante da linguagem da *lógica de primeira ordem*. Essas especificações podem ser consideradas axiomas lógicos, que permitem restringir as meta-propriedades das categorias e suas relações, além de possibilitarem também o raciocínio automatizado através de inferências (GRENON e SMITH, 2004).

Sobre o seu conteúdo semântico, pode-se afirmar que a BFO é uma ontologia que usa tanto *instâncias* quanto *universais*. As entidades (instâncias e universais) representadas na BFO são aquelas estudadas pela ciência empírica e que afetam ou estão envolvidas em atividades humanas, tais como o processamento de dados, planejamento e organização (SMITH et al., 2012). Para a compreensão dos tipos de entidades mais básicas da BFO construiu-se uma taxonomia com tais tipos, apresentada na Figura 18 a seguir.

Figura 18 - Taxonomia das entidades mais básicas da BFO.



Fonte: elaborado pelo autor.

O nó raiz da taxonomia da Figura 18 é *Entidade (Entity)*, universal mais geral da BFO. Sua definição refere-se a qualquer coisa que existe na realidade. Na sequência, pode-se identificar que as *entidades* BFO são de dois tipos: (i) *Classe (Universal)*, como, por exemplo, *vida* e *sangue* e (ii) *Instância (Particular)*, como, por exemplo, *sua vida* ou *seu sangue*. É importante, considerar aqui, que a BFO, assim como outras terminologias e ontologias, utilizam a palavra *classe* como sinônimo de *universal* (UMLS e GO) e a palavra *instância* como sinônimo de *particular* (GFO e UFO). Por fim, os *universais* podem ser de dois tipos diferentes: (i) *Universal Formal (Formal Universal)* e (ii) *Universal Material (Material Universal)* (GRENON e SMITH, 2004; SPEAR, 2006). O que distingue esses dois tipos de universais são os domínios que eles descrevem. Enquanto, um *universal formal*, às vezes chamado de *categoria*, refere-se a uma entidade mais geral, independente de domínio, um *universal material* representa uma entidade de um domínio específico. Assim, o *BFO: object aggregate* (objeto agregado) é um exemplo de universal formal e *Célula* ou *Coração* são exemplos de universais materiais.

Uma última observação sobre a taxonomia da Figura 18 é que ela foi construída respeitando-se o *princípio da disjunção* ou *mono-hierarquia*, isto é, todo nó inferior (nó filho) da taxonomia pertence somente a um nó superior (nó pai), tal como ocorre na BFO. Assim, por exemplo, uma *entidade* deve ser sempre uma *instância* ou um *universal*, mas não ambos. Segundo Smith et al. (2012), o princípio da *mono-hierarquia* é uma boa prática da engenharia ontológica, em contraste ao uso de múltipla herança no

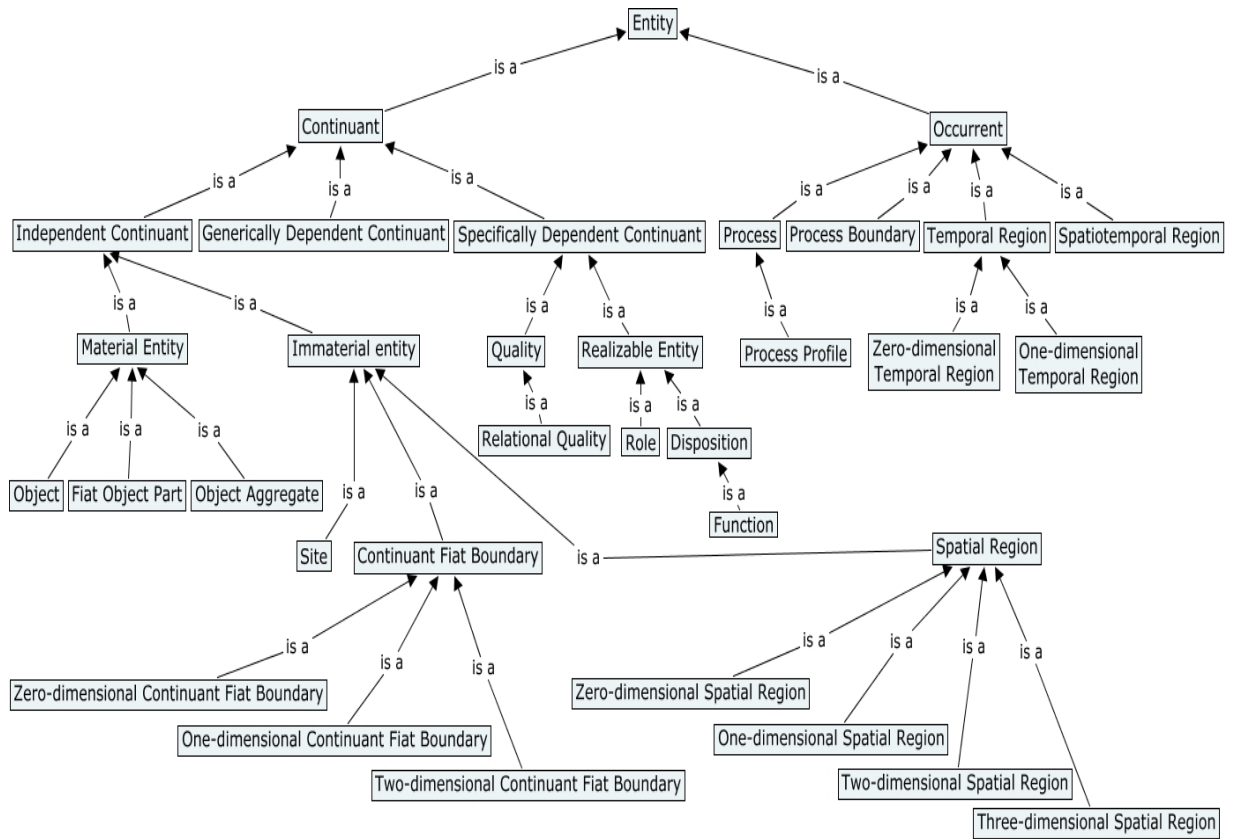
desenvolvimento de ontologias. Ainda assim, na prática nem sempre é possível obter uma ontologia sem herança múltipla.

Além desses tipos mais básicos é necessária a compreensão das demais entidades incluídas na BFO. Tal ontologia é dividada em dois grandes eixos, definidos da seguinte maneira: (i) *continuant* (continuentes), entidades que persistem ao longo do tempo mantendo sua identidade e que não possuem partes temporais; e (ii) *occurrent* (ocorrentes), entidades que se revelam, se manifestam, ou se desenvolvem ao longo do tempo e possuem partes temporais (GRENON e SMITH, 2004; SMITH, KUMAR e BITTNER, 2004).

Aqui é importante citar os sinônimos usados em outras terminologias e ontologias para se referir a continuentes e ocorrentes. **Continuantes** são referenciados como **endurantes**, por exemplo, nas ontologias DOLCE e UFO, e como **presenciais** e **persistentes** na GFO. Já os **ocorrentes** equivalem aos **perdurantes** da DOLCE e aos **eventos** da ontologia UFO.

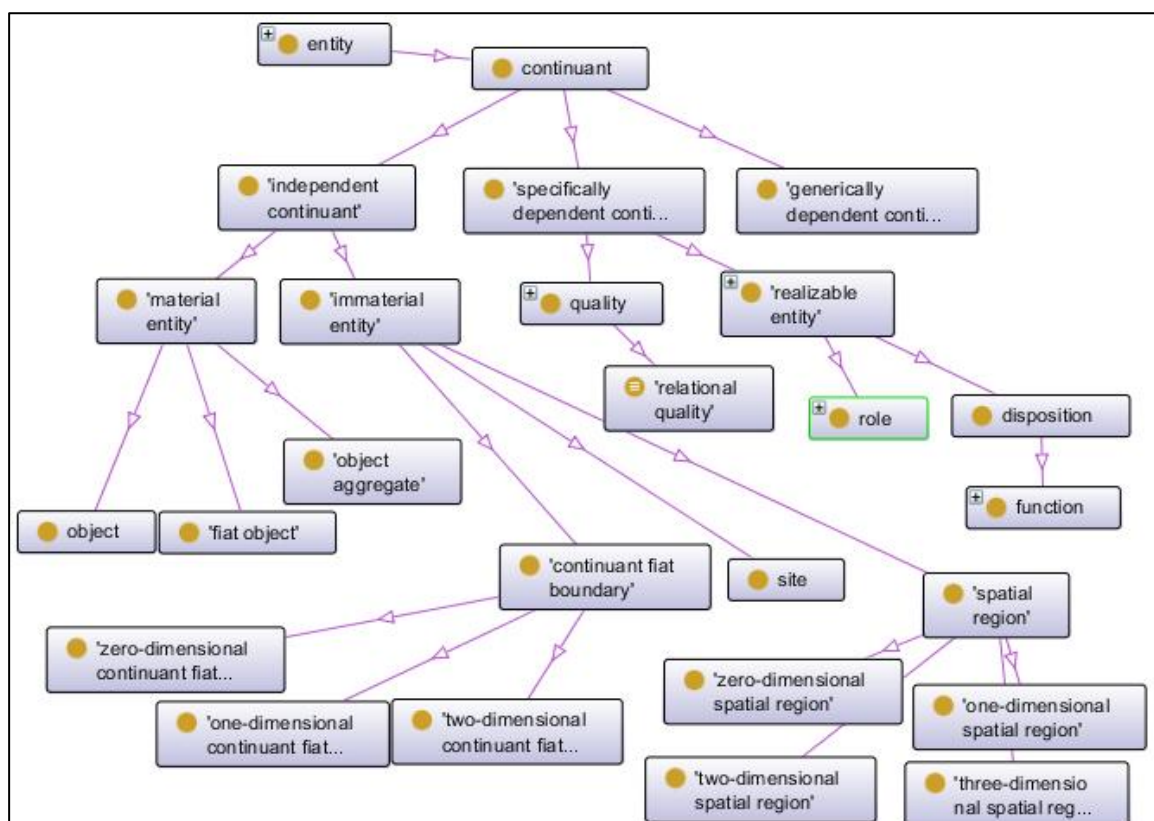
Considerando essa distinção entre entidades *continuentes* e *ocorrentes*, a BFO engloba duas subontologias: (i) a *ontologia SNAP*, composta de entidades *continuentes* e (ii) a *ontologia SPAN*, formada por entidades *ocorrentes* (GRENON e SMITH, 2004; GRENON, SMITH e GOLDBERG, 2004). As Figuras 19, 20 e 21, a seguir, apresentam tais entidades representadas através de taxonomias. A Figura 19 corresponde à taxonomia geral da BFO, a Figura 20 apresenta a ontologia *SNAP* (*entidades continuentes*) e a Figura 21 apresenta a ontologia *SPAN* (*entidades ocorrentes*).

Figura 19 – Taxonomia geral da BFO 2.0



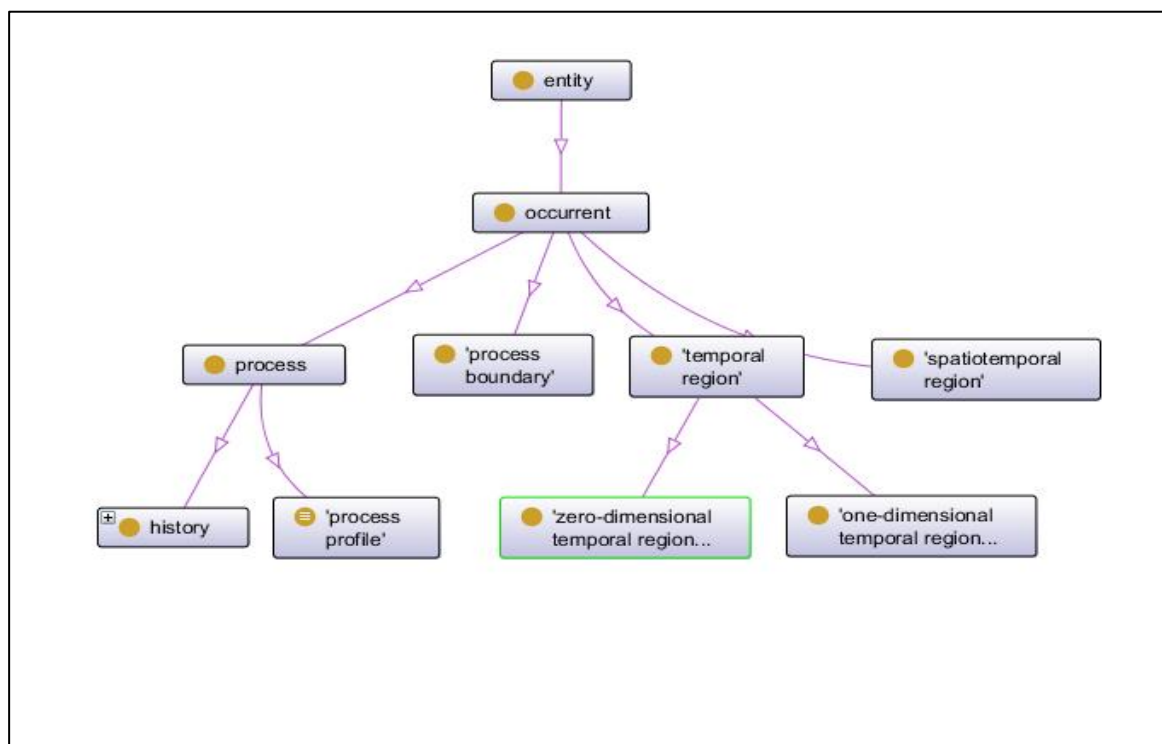
Fonte: adaptado de Smith et al. (2012).

Figura 20 – Ontologia SNAP da BFO 2.0



Fonte: elaborado pelo autor.

Figura 21 - Ontologia SPAN da BFO 2.0



Fonte: elaborado pelo autor.

Em ambas taxonomias apresentadas, as entidades BFO estão conectadas através da relação ontológica *is_a*, tal como definida formalmente na *Relation Ontology (RO)*. Na Figura 19, a relação *is_a* está explicitamente representada na taxonomia, enquanto nas Figuras 20 e 21, embora tal relação não esteja explícita, ela de fato conecta as entidades da taxonomia porque essas duas últimas representações foram obtidas através do editor de ontologias *Protégé 4.3*, que a utiliza na construção de taxonomias.

Ainda sobre a relação *is_a* é importante citar que ela ao lado da relação *part_of* correspondem às relações estritamente formais desta ontologia e, geralmente, são as mais comuns na representação das ontologias. Além destas, a BFO engloba um extenso conjunto de relações, mais específicas e menos formais, afim de possibilitar a representação de diversos domínios do conhecimento. Na prática, as relações BFO correspondem às relações contidas na ***Relation Ontology (RO)*** (SMITH et al., 2005), que foi incorporada à estrutura semântica da BFO a partir de 2005.

Sobre o desenvolvimento de ontologias de domínio a partir da taxonomia de entidades da BFO deve-se proceder da seguinte maneira: todo objeto real do domínio sob estudo deve ser associado, exatamente, a alguma entidade BFO da taxonomia, de acordo com seu nível de granularidade correspondente. A partir destas associações com o mundo real novas entidades (específicas do domínio representado) vão sendo criadas e adicionadas como classes-filho nessa hierarquia (taxonomia) de nível superior.

Por fim em relação à BFO é importante descrever a semântica de cada uma das *entidades* desta ontologia (também chamadas de *classes*) como forma de utilizá-las adequadamente na construção de ontologias de domínio. Para tanto, construiu-se duas tabelas informativas (Tabelas 8 e 9) contendo os seguintes campos: (i) o *nome da classe* na ontologia; (ii) a *classe superior* (nó-pai) a qual ela pertence; (iii) sua *definição* em forma textual; e (iv) alguns *exemplos* reais de uso. As informações apresentadas referem-se à versão mais recente (2.0) da BFO 2.0 (SMITH et al., 2012) e foram separadas considerando a divisão principal desta ontologia em entidades *continuentes* (Tabela 8) e *entidades ocorrentes* (Tabela 9), a seguir.

Tabela 8 - Tipos de entidades da ontologia de continentes da BFO 2.0

Tipo de entidade (ou classe)	is_a (classe pai)	Definição textual	Exemplos
SNAP: independent_continuant	SNAP: continuant	Continuantes portadores de qualidade e realizáveis, que não herdam de qualquer outra entidade (coisa).	Um organismo, uma porção de sangue, um braço, uma pessoa, a parte direita da perna.
SNAP: dependent_continuant	SNAP: continuant	Continuantes que sejam dependentes de um ou outro continuante independente ou sejam suportados por outras entidades.	A cor do tomate, a liquidez do sangue, o papel de ser médico, a função do coração de bombear sangue.
SNAP: material_entity	SNAP: independent_continuant	Continuantes independentes que possuem alguma porção	Um fóton, um ser humano, uma coleção

		de matéria como sua parte própria ou imprópria.	de bactérias, superfície dorsal do corpo.
SNAP: immaterial_entity	SNAP: independent_continuant	Continuantes independentes que não possuem entidades materiais como partes.	Cavidade peritoneal, o porão de um navio, sua cavidade nasal.
SNAP: object	SNAP: material_entity	Entidades materiais que estão espacialmente estendidas e unificadas causalmente ao máximo, isto é, as partes de uma substância não são separadas de outras por intervalos espaciais. A identidade de um objeto substancial independe de outras entidades e pode ser mantida através do tempo.	Um organismo, uma célula, um coração, um relógio, uma maça.
SNAP: fiat_object_part	SNAP: material_entity	Entidades materiais que são parte de um objeto, mas não são demarcadas por qualquer limite físico da realidade, já que são demarcações criadas pelos seres humanos.	Os lobos superiores e inferiores do pulmão esquerdo, as superfícies ventrais e dorsais do corpo, o Hemisfério Oeste da Terra.
SNAP: object_aggregate	SNAP: material_entity	Entidades materiais que consistem na soma mereológica de objetos separados ou na pluralidade de objetos como partes continuantes.	Uma coleção de células num banco biológico de sangue, uma orquestra sinfônica, os átomos de nitrogênio da atmosfera.
SNAP: continuant_fiat_boundary	SNAP: immaterial_entity	Entidades imateriais com zero, uma ou duas dimensões que não incluem uma região espacial física como parte.	Plano do orifício anatômico, a superfície exterior de uma célula ou parede da célula, o Pólo Norte.
SNAP: zero-dimensional_continuant_fiat_boundary	SNAP: continuant_fiat_boundary	Entidade que corresponde a um ponto <i>fiat</i> cuja localização é definida em relação a uma entidade material.	O ponto de origem de algum sistema de coordenada espacial, o Pólo Norte geográfico.
SNAP: one-dimensional_continuant_fiat_boundary	SNAP: continuant_fiat_boundary	Entidade que corresponde a uma linha <i>fiat</i> contínua, cuja localização é definida em relação a uma entidade material.	O Equador, todas linhas de latitude e longitude, o sulco mediano de sua língua.
SNAP: two-dimensional_continuant_fiat_boundary	SNAP: continuant_fiat_boundary	Entidade que corresponde a uma superfície auto-conectada, cuja localização é definida em relação a uma entidade material.	A superfície de um lado de um CD, a superfície exterior de uma célula, o lúmen do seu intestino.
SNAP: site	SNAP: immaterial_entity	Entidades imateriais de três dimensões que é, parcialmente ou totalmente, limitada por uma entidade material ou uma parte imaterial tridimensional.	Um buraco no interior de um pedaço de queijo, o interior de seu escritório, sua narina esquerda, o lúmen do seu intestino.
SNAP: spatial_region	SNAP: immaterial_entity	Entidades imateriais que são partes continuantes de um determinado espaço R. Todos tipos de entidades continuantes são entidades espaciais, já que se relacionam com alguma região espacial. A localização espacial exata da região espacial é essa própria região.	A soma total de todo espaço no universo, um ponto no espaço, uma aresta de um cubo no espaço, a superfície de uma esfera no espaço, a região do espaço ocupado pelo cubo ou pela esfera.
SNAP: zero-dimensional_spatial_region	SNAP: spatial_region	Entidade que corresponde à uma região espacial zero-dimensional.	Um ponto no espaço.
SNAP: one-dimensional_spatial_region	SNAP: spatial_region	Entidade que corresponde à uma região espacial unidimensional.	Uma aresta de um cubo no espaço

SNAP: two-dimensional spatial_region	SNAP: spatial_region	Entidade que corresponde à uma região espacial bidimensional.	A superfície de uma esfera no espaço
SNAP: three-dimensional spatial_region	SNAP: spatial_region	Entidade que corresponde à uma região espacial tridimensional.	A região do espaço ocupado por um cubo ou por uma esfera.
SNAP: generically_depend_continuant	SNAP: depend_continuant	Continuantes que dependem genericamente de um ou mais independentes continuantes. Possuem a propriedade de migração (cópia exata) de uma entidade portadora para outra.	O arquivo pdf do seu computador, a sequência de uma molécula de proteína ou a sequência que é uma cópia dela.
SNAP: specifically_depend_continuant	SNAP: depend_continuant	Continuantes que são herdados ou surgem/nascem de alguma entidade independente continuante.	A cor do tomate, a liquidez do sangue, o papel de ser médico, a função do coração de bombear sangue.
SNAP: quality	SNAP: specifically_depend_continuant	Continuantes dependentes que, ao contrário de papéis e disposições, não requerem qualquer outro processo para serem realizados.	O tamanho de um chipanzé, o cheiro de um pedaço de queijo, a temperatura ambiente de uma porção de ar.
SNAP: relational_quality	SNAP: quality	Qualidades que envolvem uma relação de dependência recíproca entre independentes continuantes.	Um vínculo matrimonial um caso de amor, um acordo entre uma pessoa e outra.
SNAP: realizable_entity	SNAP: specifically_depend_continuant	Continuantes dependentes que herdam de uma entidade material e que são realizados apenas em processos da entidade correlacionada.	A disposição do sangue em coagular, a função de seus órgãos reprodutivos, o papel de ser médico.
SNAP: role	SNAP: realizable_entity	Entidades realizáveis que participaram ou serviram a um continuante em alguns tipos de contextos naturais, sociais, trazendo algum resultado final, mas não sendo essenciais para a existência do continuante.	O papel de um estudante numa universidade, o papel de um componente químico no experimento, o papel da matéria ingerida na digestão.
SNAP: disposition	SNAP: realizable_entity	Entidades realizáveis que provocam um processo específico ou transformações no objeto do qual é inerente, sob um conjunto de circunstâncias específicas e em conjunto com as leis da natureza.	A disposição do metal de conduzir eletricidade, a disposição do sangue em coagular, a disposição genética de um paciente em ser imune a uma doença.
SNAP: function	SNAP: disposition	Disposições existentes em um continuante, devido às suas características físicas e adquiridas em sua evolução, que existem a fim de realizar de processos de uma determinada espécie.	A função de um coração no corpo para bombear sangue ou para receber sangue oxigenado, a função de reprodução na transmissão genética.

Fonte: elaborado pelo autor.

Tabela 9 - Tipos de entidades da ontologia de ocorrentes da BFO 2.0

Tipo de entidade (ou classe)	is_a (classe pai)	Definição textual	Exemplos
SPAN: process	SPAN: occurrent	Ocorrentes que tem sua existência vinculada a um acontecimento ou ocorrência, que possui partes temporais próprias e que possui dependência de uma ou mais entidades materiais.	A vida de um organismo, o voo de um pássaro, o curso de uma doença, o processo de meiose.

SPAN: process_boundary	SPAN: occurrent	Processos que são limites fiat ou bonafide temporariamente, ou seja, por um instante de tempo. Não possuem partes temporais próprias.	Aniversário, morte, a formação de uma sinapse, a incisão no início de uma cirurgia, a separação de duas células no final de uma divisão celular.
SPAN: history	SPAN: process	Processo que é a soma da totalidade dos processos que estão ocorrendo numa região espaço-temporal ocupada por uma entidade material ou um local (site), incluindo processos na superfície da entidade ou nas cavidades em que ela serve ao hospedeiro.	Os movimentos dos neutrinos no interior de uma entidade, à medida que passam por ela.
SPAN: process_profiles	SPAN: process	Partes estruturais inseparáveis dos processos que de alguma forma podem ser mensuradas com base na abstração seletiva.	Frequência respiratória, pulsação, queda de temperatura corporal, aumento da aceleração do metabolismo.
SPAN: spatiotemporal_region	SPAN: occurrent	Entidades ocorrentes que são partes da relação espaço-tempo. Todas instâncias de ocorrentes são entidades espaço-tempo, pois entram na relação de localização com a região espaço-temporal.	A região espaço-temporal ocupada por uma vida humana, a região espaço-temporal onde se desenvolveu um tumor cancerígeno.
SPAN: temporal_region	SPAN: occurrent	Entidades ocorrentes que são partes do tempo, tal como definido em um sistema de referência. Todas instâncias de ocorrentes são também entidades temporais.	O momento no qual um dedo é cortado num acidente industrial (zero-dimensional), a região temporal durante o qual o processo ocorre.
SPAN: zero-dimensional temporal_region	SPAN: temporal_region	Entidades que são regiões temporais sem extensão.	O momento que uma criança nasce; agora mesmo; o momento da morte.
SPAN: one-dimensional temporal_region	SPAN: temporal_region	Entidades que são regiões temporais que se estendem.	A região temporal durante a qual o processo ocorre.

Fonte: elaborado pelo autor.

Outra importante ontologia para a presente pesquisa refere-se à **Relation Ontology** (SMITH et al., 2005) que, atualmente, é uma ontologia de referência de relações ontológicas para biomedicina.

Historicamente, a Relation Ontology (RO) é o resultado de um trabalho em conjunto realizado por grupos de pesquisadores em ontologias biomédicas, que reúnem os autores da BFO, da GO, da FMA e da GALEN, com o propósito de definirem um conjunto restrito de relações para uso em ontologias biomédicas, que sejam logicamente bem definidas para promover a interoperabilidade entre tais ontologias (SMITH e CEUSTERS, 2010). Para tanto, o ponto de partida das relações da RO consistiu num reexame detalhado das definições básicas das relações usadas na GO e em ontologias relacionadas, para a construção de um conjunto coerentemente definido de relações bio-ontológicas formais que pudesse ser facilmente aprendido e utilizado por profissionais da área.

A interoperabilidade pretendida com o uso da RO é uma estratégia para tratar os problemas de uso e integração das relações empregadas nas ontologias que fazem parte da *Open Biomedical Ontologies* (OBO). Segundo Smith et al. (2005), as ontologias da OBO e ontologias similares utilizavam (algumas ainda utilizam) diferentes tipos de relações e, além disso, quando o faziam era de forma ambígua e inconsistente. Um exemplo são os três tipos de relacionamentos da GO representados simultaneamente pela expressão *part_of*: (i) *part_of* usado para relações entre vocabulários; (ii) *part_of* usado para indicar uma possível relação parte-todo entre entidades biológicas; e (iii) *part_of* usado para indicar uma relação parte-todo necessária (obrigatória) entre entidades biológicas. Outro exemplo é a ausência de uma distinção clara entre as relações *derives_from* e *develops_from* na *Cell Ontology*, entre outros exemplos de outras ontologias.

Smith et. al (2005) destaca também que as relações da RO podem ser consideradas genuinamente ontológicas, isto é, representam relacionamentos entre entidades na realidade, independentemente da forma como foi obtido o conhecimento da realidade e independentemente de como é feita a representação ou processamento desse conhecimento nos computadores. Nesse sentido, as relações da RO são usadas no desenvolvimento de ontologias para estabelecer a conexão entre classes do domínio, de forma que o termo “classe” é usado para se referir a uma entidade da realidade e, assim, equivalente aos termos “universal” e “tipo”, usados na literatura da área. Além desse tipo de conexão, relações da RO servem para conectar instâncias à classes, de forma que o termo “instância” é usado como equivalente aos termos “particular” e “individual” usados na literatura da área. Em ambos os casos, as relações da RO podem ser aplicadas a diferentes domínios de conhecimento, especialmente, biomédicos e, desta forma, tais relações além de exclusivamente ontológicas são também consideradas de propósito geral.

Considerando as possíveis conexões entre classes e instâncias, Grenon e Smith et al. (2004) descrevem os três tipos de relações binárias contidas na RO:

- <class, class>: relação estabelecida entre duas classes (universais). Exemplo: *Exocitose is_a Secreção*.
- <instance, class>: relação estabelecida entre uma instância (particular) e uma classe. Exemplo: *Uma membrana vesicular particular instance_of Membrana Vesicular*.
- <instance, instance>: relação identificada entre dois particulares (instâncias). Exemplo: *Uma membrana vesicular particular part_of um sistema de endomembrana de uma determinada célula*.

Além dessa categorização das relações contidas na RO, tais autores destacam mais dois tipos de classificação para elas: (i) considerando o critério de *divisibilidade*, temos: (a) relações *primitivas*, as quais são naturais e indivisíveis. Exemplo: *instance_of* e (b) relações *derivadas*, que são aquelas compostas a partir de relações primitivas. Exemplo: *is_a*. Por outro lado, ao considerar os *tipos de entidades conectadas* por uma relação ontológica da RO, temos: (a) *relações fundamentais*: correspondem às relações estritamente formais, usadas para a construção de taxonomias e partonomias; (b) *relações espaciais*: usadas para conectar entidades que são regiões espaciais às entidades que correspondem aos ocupantes dessas regiões; (c) *relações temporais*: usadas para a conexão de entidades existentes em diferentes instantes de tempo; e (d) *relações de participação*: usadas para conectar entidades processuais (processos) aos seus portadores. Como exemplo de tais tipos de relações, temos a primeira versão da RO publicada em 2005 contendo ao todo 10 relações ontológicas, conforme pode ser visto na Tabela 10, a seguir.

Tabela 10 - As dez relações formais da RO em sua primeira versão

Grupo	Relações
Relações fundamentais	<i>is_a</i> <i>part_of</i>
Relações espaciais	<i>located_in</i> <i>contained_in</i> <i>adjacent_in</i>
Relações temporais	<i>transformation_of</i> <i>derives_from</i> <i>preceded_by</i>
Relações de participação	<i>has_participant</i> <i>has_agent</i>

Fonte: adaptado de Smith et al. (2005).

Outro ponto importante a considerar sobre o emprego das relações RO no desenvolvimento de ontologias é o uso de uma sintaxe semi-formal para definir as relações de tal ontologia. A RO foi desenvolvida a partir de uma rigorosa metodologia para a construção de relações providas de definições formais claras e precisas no domínio biomédico (SMITH et al., 2005). Apesar do rigor formal de tais relações da RO, os detalhes técnicos usados em suas formulações são transparentes ao trabalho dos curadores e desenvolvedores de ontologia, cuja tarefa principal deve-se concentrar em saber qual

relação e como utilizá-la em sua ontologia. Entretanto, para os objetivos da presente pesquisa a compreensão dessa sintaxe é necessária afim de apresentar a semântica das relações da RO.

Na prática, a sintaxe semi-formal da RO é equivalente a um padrão de notação lógica usado para especificar as relações. As convenções básicas deste padrão notacional podem ser resumidas conforme abaixo (SMITH et. al, 2005):

- As variáveis **C**, **C1**, **Cn** (com letra maiúscula) são usadas para representação dos *universais continuantes* e **c**, **c1**, **cn** (letra minúscula) usadas para representar *particulares continuantes*;
- As variáveis **P**, **P1**, **Pn** (com letra maiúscula) são usadas para representação dos *universais ocorrentes* e **p**, **p1**, **pn** (letra minúscula) usadas para representar *particulares ocorrentes*;
- As variáveis **r**, **r1**, **rn** (com letra minúscula) são usadas para representar *regiões espaciais tridimensionais*.
- As variáveis **t**, **t1**, **tn** são usadas para representar intervalos de tempo.
- As relações entre dois universais são escritas em itálico, tal como *C is_a C1*.
- As relações entre dois particulares (exemplo: **c1 part_of c**) e entre um particular e um universal (exemplo: **c instance_of C**) são grafadas em negrito.

Usando tal padrão notacional, as relações ontológicas incluídas na RO possuem cada uma definição textual e outra definição semi-formal que permitem ao desenvolvedor distinguir entre elas e escolher aquela mais apropriada para representação desejada. Conforme mencionado anteriormente, a primeira versão da RO continha 10 relações formais e, atualmente, esse número de relações³¹ está em torno de 160. Optou-se na presente pesquisa por descrever aquelas que, possivelmente, tem aplicação direta na construção da ontologia proposta.

Para descrever tais relações, construiu-se uma tabela informativa (Tabela 11), baseada em referências sobre esta ontologia (GRENON e SMITH, 2004; SMITH et. al, 2005; SPEAR, 2006), contendo os seguintes campos: (i) *Relação*: que serve para apresentar a relação, utilizando a sintaxe semi-formal usada na RO; (ii) *Definição textual*: definição da relação RO por meio descritivo; (iii) *Composição da relação*: descreve como a

³¹ Disponível em: <http://code.google.com/p/obo-relations/>. Acesso em: 26 de Abril de 2013.

relação é formada a partir de suas relações primitivas. De forma óbvia, apenas as relações derivadas possuem esse campo preenchido; e (iv) *Exemplos*: exemplos de uso da relação. A Tabela 11 apresenta as relações contendo tais informações. O acrônimo RO é utilizado na frente de cada relação da tabela para identificar que tal relação pertence à ontologia *Relation Ontology* (RO).

Tabela 11 - Atributos e exemplos das relações da RO

Relação	Definição textual	Composição da relação	Exemplos
RO: c instance_of C at t	O particular continuante c é uma instância do universal continuante C num dado tempo t.	_____	your heart instance_of Heart at the present
RO: p instance_of P	O particular ocorrente p é uma instância do universal ocorrente P, independentemente de tempo.	_____	your life instance_of Human Life
RO: C is_a C1	O universal continuante C é um universal continuante C1, somente se cada particular continuante c for uma instância de C num dado tempo t e também for uma instância de C1 neste mesmo tempo t.	C instance_of C at t and C instance_of C1 at t	Blood Cell is_a Cell Heart is_a Organ
RO: P is_a P1	O universal ocorrente P é um universal ocorrente P1, somente se cada particular ocorrente p for uma instância de P num dado tempo t e também for uma instância de P1 neste mesmo tempo t.	P instance_of P at t and P instance_of P1 at t	LungCancer Development is_a CancerDevelopment
RO: c part_of c1 at t	O particular continuante c é parte do particular continuante c1 num dado tempo t.	_____	this cells nucleus part_of this cell at the present
RO: p part_of p1	O particular ocorrente p é parte do particular ocorrente p1, independentemente de tempo.	_____	this tumor's growth part_of John's life
RO: r part_of r1	A região espacial r é uma subregião (ou parte de) de uma outra região r1, independentemente de tempo.	_____	the spatial region occupied by heart part_of the spatial region occupied by body
RO: C part_of C1	O universal continuante C é parte de um outro universal continuante C1, somente se, para cada particular c de C num tempo t qualquer, exista algum particular c1 de C1, talque c seja parte de c1 no tempo t.	c is_a C at t and c1 is_a C1 at t and c part_of c1 at t	Human Heart part_of Human Circulatory System
RO: P part_of P1	O universal ocorrente P é parte de um outro universal ocorrente P1, somente se, para cada particular p de P exista algum particular p1 de P1, talque p seja parte de p1, independentemente de	p is_a P and p1 is_a P1 and p part_of p1	Second Trimester part_of Human Fetal Developmente

	tempo.		
RO: c <i>inheres_in</i> c1 at t	O particular continuante c é inerente (herdeiro) de um outro particular continuante c1 num tempo t qualquer.	_____	the shape of John's body <i>inheres_in</i> John at July, 2006
RO: C <i>inheres_in</i> C1	O universal continuante C é inerente (herdeiro) de um outro universal continuante C1, somente se, para cada particular c de C num tempo t qualquer, exista algum particular c1 de C, tal que c é inerente (herdeiro) de c1 no tempo t.	c <i>is_a</i> C at t and c1 <i>is_a</i> C1 at t and c <i>inheres_in</i> c1 at t	Rational <i>inheres_in</i> Human Being
RO: c <i>located_in</i> r at t	O particular continuante c está localizado numa região especial r em um dado tempo t.	_____	John <i>located_in</i> the dinning room at dinner time
RO: C <i>located_in</i> C1	O universal continuante C está localizado em um universal continuante C1, somente se todos os particulares continuantes c são instâncias de C num tempo t qualquer e exista algum particular continuante c1 instância de C1, tal que c seja localizado em c1 no tempo t.	c <i>is_a</i> C at t and c1 <i>is_a</i> C1 at t and c <i>located_in</i> c1 at t	Kidney <i>located_in</i> Torso
RO: p <i>has_participant</i> c at t	O particular ocorrente p é um processo participante do particular continuante c num dado tempo t.	_____	John's life <i>has_participant</i> John at January, 1984
RO: P <i>has_participant</i> C	O universal ocorrente P é um processo participante do universal continuante C, somente se, para cada particular p de P exista algum particular continuante c instância de C num tempo t, tal que p seja um processo participante de c no tempo t.	p <i>is_a</i> P and c <i>is_a</i> C at t and p <i>has_participant</i> c at t	Second trimester <i>has_participant</i> Fetus

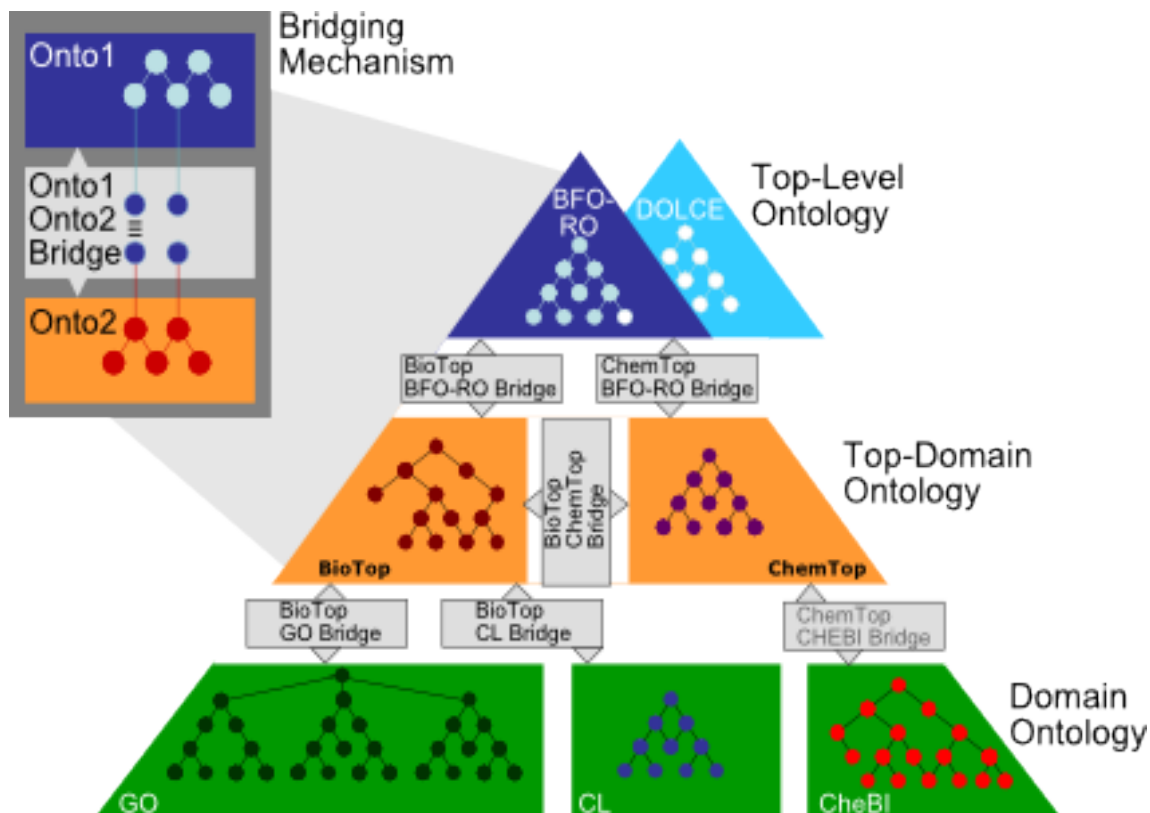
Fonte: elaborado pelo autor.

Outra importante ontologia de fundamentação criada para biomedicina é a **Biological Top-Level (BIOTOP)**, que foi desenvolvida com o objetivo de prover uma camada ontológica para a ligação e integração de diversas ontologias de domínios específicos em ciências da vida (BEISSWANGER et al., 2007). Considerando esse propósito, a BIOTOP pode ser considerada uma ontologia de alto nível para o domínio biomédico, especificamente, a área de biologia molecular.

Diferentemente das outras ontologias de fundamentação, aqui apresentadas, a BIOTOP não se propõe a descrever conceitos genéricos de domínios diversos, mas sim fornecer uma camada ontológica intermediária entre as ontologias de fundamentação e as ontologias biomédicas de domínio. Essa ponte é necessária, segundo Beisswanger et al. (2007), para garantir uma transição mais suave e natural entre as entidades de uma

ontologia de alto nível (fundamental) e as entidades de uma ontologia de domínio muito específico. Esses mesmos autores destacam ainda que apesar dos esforços empregados nesse processo, seus resultados dependem, criticamente, do aprofundamento das ontologias de domínio em uma estrutura formal rígida das ontologias de fundamentação. A Figura 22, a seguir, ilustra essa proposta de arquitetura ontológica do projeto BIOTOP.

Figura 22 - Arquitetura ontológica do projeto BIOTOP.



Fonte: adaptado de BIOTOP (2012).

A arquitetura, apresentada na Figura 22, é constituída de três camadas: (i) na primeira temos as ontologias de alto nível (fundamentais), tais como a BFO-RO e a DOLCE; (ii) a segunda camada contém as ontologias que farão essa ponte de integração, chamadas de *ontologia de alto domínio*: a própria BIOTOP e a CHEMTOP, que é uma versão mais atual da BIOTOP numa nova estrutura ontológica (BIOTOP, 2012); e (iii) a terceira camada engloba as ontologias de domínio, aqui representadas através da *Gene Ontology (GO)*, *Cell Ontology (CL)* e *Chemical Entities of Biological Interest (CheBI)*. A integração entre as três camadas é realizada pelas chamadas pontes BIOTOP com a ontologia correspondente, que nada mais são do que as relações ontológicas entre as entidades destas ontologias.

A integração proposta com a BIOTOP inclui também preocupações com aplicações de processamento de linguagem natural (PLN) nas ciências da vida, tais como a extração de informação e a mineração de textos. Isso se deve ao fato de que a BIOTOP foi desenvolvida como uma extensão e remodelagem da ontologia GENIA (SCHULZ et al., 2006).

A ontologia GENIA é um modelo formal para sinalização de reações celulares em humanos, usada como base para dicionários semânticos e tesouros em aplicações de PLN, que envolvam recuperação e extração de informação, classificação e categorização de termos (SCHULZ et al., 2006). As categorias incluídas na GENIA restringem-se às entidades continuantes biológicas, que correspondem a um total de 48 termos (chamados de classes), dispostos em uma estrutura taxonômica com profundidade máxima de 6 nós. A única relação incluída na ontologia GENIA é a *is_a* e também não existem axiomas.

O trabalho conduzido pelos criadores da BIOTOP foi tomar a estrutura semântica da GENIA como ponto de partida para o desenvolvimento de uma ontologia superior para domínios de biologia molecular e de biomedicina. Nesse trabalho, Beisswanger et al. (2007) afirma que a GENIA foi remodelada, tendo parte de suas classes com problemas removidas, novas classes adicionadas, inclusão de relações mereológicas e melhorias sensíveis na descrição das classes, que no projeto original de GENIA eram descritas, informalmente, por meio de notas de escopo verbais. De maneira geral, pode-se afirmar que o propósito inicial de uso da GENIA para anotação semântica foi estendido, com a BIOTOP, para a modelagem de domínios do conhecimento usando esta ontologia.

Em 2008, a BIOTOP era composta de 175 classes, ligadas a 171 instâncias por meio de relações binárias pertencentes aos 9 tipos de relações semânticas previstas na ontologia, além das subrelações destas e suas relações inversas. Neste mesmo ano, a BIOTOP passou por um processo de reengenharia ontológica para tornar mais claro seu foco biomédica e, assim, um número significativo de seus axiomas foram migrados para uma nova estrutura, dando a origem à ontologia ChemTop (BIOTOP, 2012).

Sobre o conteúdo semântico da BIOTOP e sua interface com outras ontologias (GENIA, BFO, GO, CL e ChEBI), Schulz et al. (2006) e Beisswanger et al. (2007) explicam da seguinte forma:

- A BIOTOP herda a distinção das classes BFO em entidades continuantes e ocorrentes, e também a distinção entre continuantes independentes e continuantes dependentes. Entretanto, as subclasses BFO de continuantes independentes não foram incorporadas na BIOTOP. Como subclasses destes tipos, a BIOTOP utiliza entidades moleculares equivalentes às

entidades GENIA, tais como *Organismo*, *Tecido*, *Célula*, *Componente Celular* e *Átomo*. Da BFO, a BIOTOP ainda herda: (i) a classe *BFO:Process*, para iniciar a hierarquia dos processos biológicos; (ii) a classe *BFO:Function*, para iniciar a hierarquia das funções biológicas; (iii) a classe *BFO:Quality*, para definir algumas qualidades como *Massa Física* e *Estado Canônico*; e (iv) a classe *BFO:Role*, para definir papéis como *Papel de Sinalização*.

- A partição principal da GENIA em *Source (Fonte)* e *Substance (Substância)* também é incorporada pela BIOTOP. *Substance* refere-se às substâncias envolvidas em reações químicas, enquanto *Source* indica as localizações biológicas onde as substâncias são encontradas e suas reações nesses locais. Na BIOTOP os continuantes independentes *Organismo*, *Tecido*, *Célula*, *Componente Celular* e *Átomo* fazem parte do braço *Source*, enquanto as subclasses da *Entidade Mono Molecular* pertencem ao braço *Substance*.
- Pelo fato da *Gene Ontology (GO)* restringir o significado de “molécula” para “produto genético” e de “processos biológicos” para “processos que envolvam produtos genéticos”, temos relações de subsunção entre a BIOTOP e GO: o braço *Molecular Function* da GO é subsumido à classe *Molecular Function* da BIOTOP e o braço *Biological Process* da GO à *Biological Process* da BIOTOP. Com relação ao terceiro braço GO – o *Cellular Component* – apenas algumas de suas partes são subsumidas à classe *Cellular Component* e não toda esta parte da hierarquia, porque na BIOTOP os componentes celulares são definidos como partes próprias das células, ao contrário da GO.

Um último registro sobre a BIOTOP vai para as relações ontológicas incluídas nesta ontologia. Além da relação de subsunção *is_a*, herança da GENIA, a BIOTOP incluiu algumas relações semânticas da *OBO-Relation Ontology (OBO-RO)*: *proper_part_of*, *located_in*, *derives_from*, *has_participant* e suas relações recíprocas; e também novas relações não contidas na OBO-RO, como: *has_inherence* (relação recíproca: *inheres_in*), *realization_of* (relação recíproca: *has_realization*) e duas subrelações de *has_part*: viz *has_grain* (relação recíproca: *grain_of*) e *component_of* (relação recíproca: *has_component*) (BEISSWANGER et al., 2007).

4.4 Ontologias biomédicas de domínio

No desenvolvimento da teoria ontológica do sangue humano foram reutilizadas classes tanto de ontologias de fundamentação quanto de ontologias biomédicas de domínio. Nesta seção, são descritas algumas ontologias que representam o conhecimento de domínios específicos da biomedicina e que tiveram seus termos reutilizados para a construção da teoria ontológica do sangue. São elas: *Generalized Architecture for Language, Encyclopaedias and Nomenclatures (GALEN)*; *Gene Ontology (GO)*; *Foundational Model Anatomy (FMA)*; *Cell Ontology (CL)*; *Protein Ontology (PRO)* e *Chemical Entities of Biological Interest (ChEBI)*.

O programa ***Generalized Architecture for Language, Encyclopaedias and Nomenclatures (GALEN)*** de pesquisa em terminologias médicas se iniciou em 1991 e, em 1999, o projeto **OpenGALEN** foi formado para prover uma fonte de aberta de disseminação dos resultados desse programa e um framework para seu desenvolvimento (RECTOR et al., 2003). Desde então a Fundação OpenGALEN tem disponibilizado recursos para a organização do domínio biomédico, que incluem uma ontologia clínica de fonte aberta e um modelo de referência comum.

Rector et al. (2003) descreve tais recursos ou componentes do projeto OpenGALEN:

- O *Modelo de Referência Comum OpenGALEN* é considerada uma grande ontologia, que inclui: (i) uma ontologia de alto nível, usada para fornecer uma estrutura geral de categorização; e (ii) o CORE ou modelo de referência, que contém as definições reutilizáveis da área de anatomia, além de conteúdos relacionados à fisiologia humana, patologia e sintomatologia – os sintomas.
- *GALEN Representation and Integration Language (GRAIL)* é o nome dado à linguagem de lógica descritiva usada pela OpenGALEN para representação da ontologia definida. Freitas e Schulz (2009) ressaltam que a linguagem GRAIL permite a definição de classes de forma semelhantes à feita pelo SNOMED-CT, porém com uma sintaxe mais rica.
- *OpenKnoME* é um ambiente de desenvolvimento de código aberto, usado para construir e manter o Modelo de Referência OpenGALEN.
- Um repositório de documentação, que incluem especificações e metodologias relacionados à abordagem OpenGALEN.

A OpenGALEN contém aproximadamente 25.000 conceitos, que estão dispostos em múltiplas hierarquias *is-a*, além de um total de 26 tipos de relações. Rector, Rogers e Pole (1996) descrevem sobre tais conceitos e relações do CORE OpenGALEN, apresentado um resumo do Esquema OpenGALEN. Em tal descrição, eles apresentam as categorias, os modificadores e a hierarquia de atributos OpenGALEN. A Tabela 12, a seguir, ilustra a taxonomia das categorias e dos modificadores.

Tabela 12 - Extrato da taxonomia das categorias principais da OpenGALEN

Entity	Example	Entity	Example
GeneralisedProcess		GeneralisedSubstance	
SpecificProcess		Energy	
BiologicalProcess		Substance	
BodyProcess	Peristalsis, Clotting	BodySubstance	
Behaviour	VolitionalAct, ClinicalAct	Tissue	MuscleTissue
NonBiologicalProcess		NamedBodySubstance	Urine, Bile, Sputum
PhysicalProcess	Irradiation	ChemicalSubstance	Drugs, Sodium
ChemicalProcess	Histological Staining	NamedSubstance	Air, Wood
GenericProcess	Transport, Opening		
GeneralisedStructure		ModifierConcept	
AbstractStructure		Aspect	
PsychosocialConstruct	Clinic, Hospital	Feature	Sex, Shape, Chronicity
LogicalStructure	Protocol, Plan	State	Male/female, acute/chronic
PhysicalStructure		Selector	leftSelector/ rightSelector
LinearStructure	Displacement	Status	pathological/ physiological
PlanarStructure	Triangle, Square	Collection	Rash
SolidStructure		Modality	FamilyHistory
MicroscopicStructure	Cell, Microorganism	Role	DrugRole, PatientRole
InerdSolidStructure	Building, Device	Unit	second, kilogram
OrganicSolidStructure		LevelofSpecification	uniquelySpecified
BodyStructure			
BodyPart	HeartLeg, Head, Femur		
GenBodyStructure	Artery, Lump, Orifice		

Fonte: adaptado de Rector, Rogers e Pole (1996).

Através da Tabela 12 é possível identificar que as categorias OpenGALEN, incluem estrutura, substância, processo e modificadores. Esses últimos são divididos em

três grupos principais: (i) os aspectos, tais como localização e forma; (ii) as modalidades, como História Familiar ou Risco; e (iii) as coleções, como, por exemplo, História Familiar de Diabetes.

Um comentário final sobre a OpenGALEN, que merece registro, refere-se a sua abrangência. Segundo Freitas e Schulz (2009), o propósito da OpenGALEN é semelhante ao do SNOMED-CT, no entanto, esse projeto jamais alcançou o escopo e granularidade do SNOMED-CT. Ainda segundo esses autores, a principal contribuição do OpenGALEN é que ele pode ser considerado pioneiro na utilização da lógica formal nas terminologias biomédicas, introduzindo um recurso extremamente útil e necessário das ontologias.

Por sua vez, a **Gene Ontology (GO)** é considerada uma das ferramentas (recursos) mais importantes para a representação e processamento de informação de produtos genéticos e suas funções (SMITH, WILLIAMS e SCHULZE-KREMER, 2003). Mantida pelo *Gene Ontology Consortium*, a GO foi criada, originalmente, para dar suporte a apontamentos compartilhados em três bancos de dados sobre modelos de organismos: *FlyBase*, *Mouse Genome Informatics* e *Saccharomyces Genome Database (SGD)*, como citado em GO (2012).

Apesar do seu nome, a *Gene Ontology* dificilmente pode ser considerada uma ontologia em si. A definição mais precisa para a GO é a de um vocabulário controlado usado para descrever os produtos genéticos e suas funções em qualquer organismo (SMITH, WILLIAMS e SCHULZE-KREMER, 2003). A razão para essa confusão conceitual deve-se ao uso que é feito do termo ontologia em áreas como a Ciência da Computação e a própria Ciência da Informação, compreendendo, erroneamente, ontologias como um tipo de vocabulário controlado.

Desde a sua criação, o escopo da GO foi ampliado para cobrir todo o domínio de biologia, independentemente das características de organismos específicos. Atualmente, a GO engloba um total de 38137 termos - todos com definições completas. Desse total, 23928 são processos biológicos, 3050 são componentes celulares e 9467 são funções moleculares (GO, 2012).

A GO é dividida em três hierarquias disjuntas (arquitetura tripartite), às vezes chamadas de sub-ontologias, que são (GO, 2012; SMITH, WILLIAMS e SCHULZE-KREMER, 2003):

- *Componente celular*: refere-se ao local da célula onde o produto genético é ativo. Os componentes celulares incluem termos como ribossomo e proteossoma, especificando onde os múltiplos produtos genéticos serão

encontrados. Essa hierarquia também inclui termos como membrana celular ou aparatos Golgi.

- *Função (ou atividade) molecular*: é a atividade bioquímica de um produto genético. Ela descreve o que está concluído, sem especificar onde ou quando o evento realmente ocorre. As funções moleculares normalmente correspondem às atividades que podem ser realizadas por produtos genéticos individuais, mas também atividades realizadas por conjuntos complexos de produtos genéticos. Exemplos de termos funcionais mais amplos são: enzima, transportador ou ligante. Um exemplo de termo funcional específico é adenilato ciclase.
- *Processos biológicos*: uma série de eventos realizados de um ou mais conjunto ordenados de funções moleculares. Os processos também envolvem transformação física ou química. Exemplos de termos que descrevem processos biológicos de alto nível são: processo celular fisiológico ou crescimento e manutenção celular. Já alguns exemplos de termos para descrever processos mais específicos são: processo metabólico pirimidina ou transporte alfa-glucosídeo.

Cada uma dessas hierarquias apresentadas constitui uma hierarquia múltipla, que contém 24.500 nós, chamados de termos na GO. Esses termos podem ser compreendidos como identificadores semânticos que padronizam a descrição de dados sobre os genes (ou produtos genéticos) como afirmam Freitas e Schulz (2009). Assim, a GO é bastante útil para as pesquisas nas bases de dados mantidas pelo *GO Consortium*, tornando mais simples o acesso ao conhecimento descoberto pelos pesquisadores.

Uma pesquisa na GO por uma determinada proteína, por exemplo, pode obter a seguinte descrição deste produto genético (em três dimensões): (i) a mitocôndria – o compartimento celular onde seu gene é expresso; (ii) a sinalização – função em que a proteína está associada; e (iii) a mitose – processo biológico do qual a proteína participa, conforme exemplifica Freitas e Schulz (2009).

Sobre a estrutura semântica da GO pode-se afirmar ainda que: (i) os termos GO representam classes de entidades reais; (ii) esses termos são caracterizados por identificadores, chamados de números de inclusão, e possuem como atributos sinônimos e suas definições e (iii) as relações GO são de dois tipos: *is-a* e *whole-part* (GO, 2012).

A Figura 23, a seguir, ilustra o registro da classe Célula na GO, incluindo sua representação na hierarquia *is-a*, indicada pela letra “I”, e na hierarquia *whole-part*, indicada através da letra “P”.

Figura 23 - Registro da classe Célula na Gene Ontology (GO).

(I) GO:0005623 : cell
 (P)GO:0044464 : cell part
 (I) GO:0009334 : 3-phenylpropionate dioxygenase complex
 (I) GO:0020007 : apical complex
 (P) GO:0020032 : basal ring of apical complex
 (P) GO:0020010 : conoid
 (P) GO:0033289 : intraconoid microtubule
 (P) GO:0020009 : microneme
 (P) GO:0070074 : mononeme
 (P) GO:0020031 : polar ring of apical complex
 (P) GO:0020008 : rhoptry
 (P) GO:0020025 : subpellicular microtubule

Cell

Term Information

Accession: GO:0005623

Ontology: cellular component

Synonyms: None

Definition: The basic structural and functional unit of all organisms. Includes the plasma membrane and any external encapsulating structures such as the cell wall and cell envelope.

[source: GOC:go_curators]

Fonte: adaptado de Freitas e Schulz (2009).

Embora seja um importante recurso na organização da informação biomédica e de uso intensivo por parte dos biocuradores, especialmente àqueles ligados ao GO Consortium, a GO apresenta algumas falhas (problemas ontológicos), que foram identificadas por alguns pesquisadores (SMITH, WILLIAMS e SCHULZE-KREMER, 2003; KUMAR e SMITH, 2003), como pode ser visto a seguir:

- São identificadas evidências incorretas na fixação do relacionamento *whole-part* entre os elementos das três hierarquias existentes, especialmente, na realização de algumas funções e de processos biológicos mais amplos, que, às vezes, não são encontrados.
- Alguns termos incluídos na hierarquia Função Molecular, na realidade, são substâncias e não funções em si. Assim, é possível encontrar termos como anti-coagulante (definido como: uma substância que retarda ou previne a coagulação) e enzima (definida como: uma substância que cataliza [...]) em Função Molecular.
- A relação *is-a* é utilizada de forma imprecisa na GO, já que em alguns casos é possível encontrar tal relação sendo usada no lugar da relação

whole-part e também casos onde a relação *is-a* é destacada como uma subsunção não necessária – ao contrário de sua definição.

Outra ontologia biomédica de domínio descrita nesta revisão refere-se ao ***Foundational Model of Anatomy (FMA)***. Desenvolvido pela Escola de Medicina da Universidade de Washington, o FMA foi concebido como uma representação de classes ou tipos necessários para a representação simbólica da estrutura fenotípica do corpo humano (FMA, 2012). Especificamente, o FMA é uma ontologia de domínio que representa o conhecimento declarativo e explícito sobre a anatomia humana.

A estrutura FMA compreende objetos materiais moleculares dos níveis macroscópicos que constituem o corpo humano e associações com entidades não materiais (espaços, superfícies, linhas, pontos) requeridas para descrição dos relacionamentos estruturais (ROSSE e MEJINO, 2003).

O FMA foi desenvolvido, inicialmente, como um aprimoramento do conteúdo anatômico dos vocabulários integrados ao UMLS, com o objetivo de facilitar a correlação entre os conceitos anatômicos representados por esses diferentes vocabulários, como afirma Rosse e Mejino (2003). Desde então, o vocabulário *Digital Anatomist*, conhecido como o componente UWDA do UMLS, está presente na estrutura semântica UMLS.

Atualmente, o FMA contém aproximadamente 75.000 classes anatômicas, que representam desde complexas estruturas macromoleculares aos componentes de células do corpo humano, 120.000 termos associados a essas classes e 168 tipos de relacionamentos diferentes (FMA, 2012). As classes anatômicas se relacionam com uma outra classe por um tipo de relacionamento específico, formando um total de 2,1 milhões de instâncias de relacionamentos.

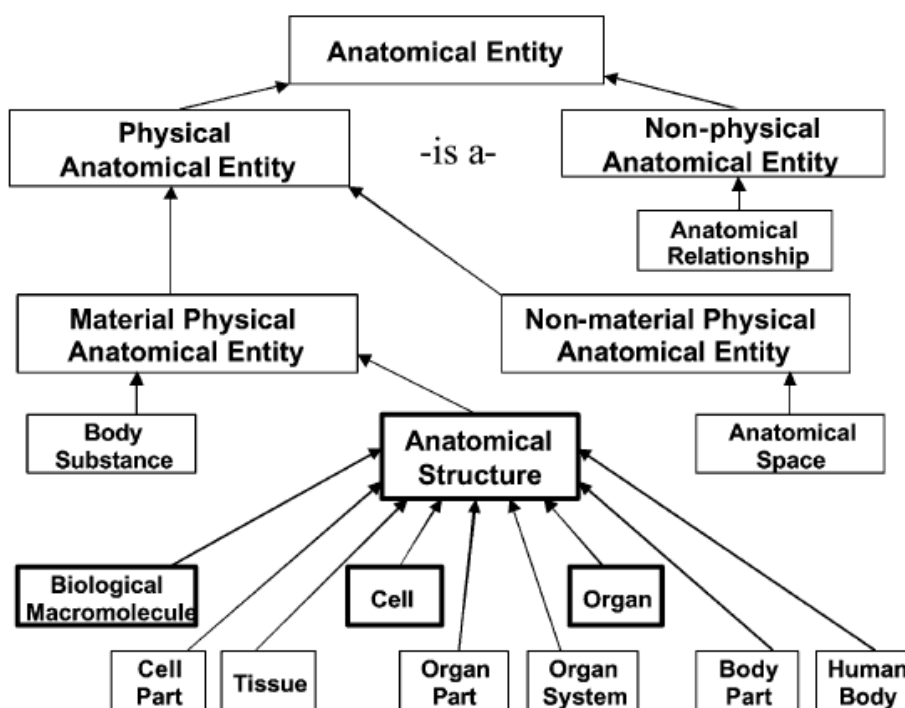
É também importante registrar que a ontologia FMA possui quatro componentes interrelacionados (FMA, 2012), conforme mostrado a seguir:

- *Anatomy taxonomy (At)*: classifica as entidades anatômicas de acordo com suas características similares e com as características que as diferenciam em relação às outras;
- *Anatomical Structural Abstraction (ASA)*: especifica os relacionamentos todo-parte e espaciais existentes entre as entidades representadas em At;
- *Anatomical Transformation Abstraction (ATA)*: especifica a transformação morfológica das entidades representadas em At durante o ciclo de vida de desenvolvimento pré-natal e pós-natal;

- *Meta-knowledge (Mk)*: especifica os princípios, regras e definições, segundo os quais as classes e os relacionamentos são representados nos outros três componentes FMA.

Desta forma, costuma-se referenciar a FMA, na literatura, a partir da abstração FMA = (At, ASA, ATA, Mk). Afim de ilustrar essa ontologia, apresenta-se, a seguir, a Figura 24 para descrever um dos componentes FMA – a *Anatomy taxonomy* – representando suas principais classes.

Figura 24 - Hierarquia *is-a* das principais classes da Anatomy taxonomy do FMA.



Fonte: adaptado de Rosse e Mejino (2003).

Utilizando-se de uma abordagem de modelagem do conhecimento disciplinada, assim denominada por Rosse e Mejino (2003), o FMA foi desenvolvido com base em alguns princípios fundamentais de modelagem (contexto unificado, nível de abstração, princípio de definição, conceito dominante, entre outros) e também em definições aristotélicas sobre os objetos do mundo. Por causa dessa abordagem, os nós das hierarquias FMA são denominados de *classes* ou *tipos*, amparando seu comprometimento com entidades do mundo real, ao invés de comprometimento com os significados de termos.

O desenvolvimento do FMA, sob essa perspectiva, justifica-se pelo fato do FMA caracterizar-se como uma ontologia de referência no domínio anatômico do corpo

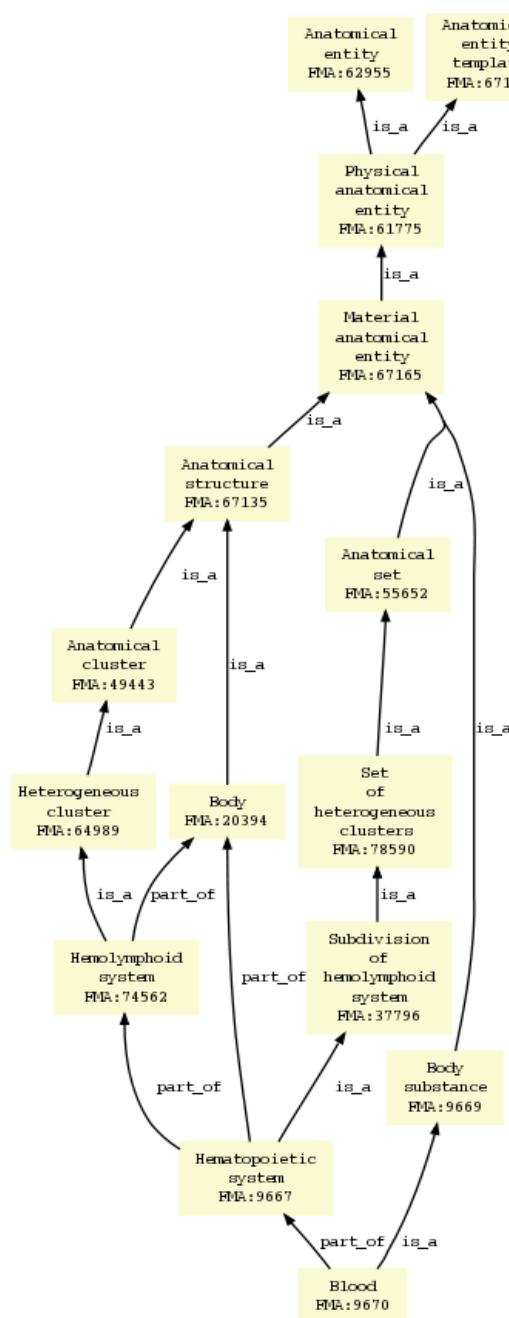
humano, que possa ser utilizável e generalizável no desenvolvimento de outras ontologias do domínio biomédico. Por isso, seus criadores utilizam o termo “*Foundational*” para descrevê-la, acrescentando ainda que: (i) a anatomia é fundamental para todos os domínios biomédicos; e (ii) os conceitos anatômicos e seus relacionamentos no FMA generalizam todos esses domínios.

Um dos elementos da anatomia humana descritos no FMA é o sangue e algumas de suas especializações. Por esse motivo, a ontologia FMA é de extrema importância para o desenvolvimento aqui proposto. Nesta pesquisa, o FMA é tomada como ponto de partida para compreensão da fisiologia do sangue humano, porém não é a única referência utilizada para a aquisição e representação do conhecimento da área, porque o conteúdo contido em tal ontologia não é suficiente para descrever o domínio sob estudo e também por faltar definições para todos os termos do sangue incluídos no FMA. Assim, consulta-se manuais técnicos, normas técnicas e publicação científicas recentes da área para representar o conhecimento sobre o sangue humano.

Tomando o FMA como ponto de partida para a compreensão da fisiologia do sangue humano, é possível identificar que a entidade sangue (*FMA: Blood*) é definida, nesta ontologia, como “substância do corpo que é formada de plasma (*FMA: Plasma*) e células sanguíneas” e classificada como uma entidade anatômica material (*FMA: Material anatomical entity*).

Em sua estrutura semântica, o FMA apresenta uma taxonomia para a chamada *entidade anatômica do sangue*. Em tal taxonomia, abaixo da categoria básica *FMA Anatomical Entity*, encontram-se duas categorias disjuntas, chamadas de *Physical Anatomical Entity* e *Non-Physical Anatomical Entity*. Abaixo de *Material Anatomical Entity* encontra-se a categoria *Body Substance* e, por último, temos a categoria *Blood* (veja Figura 25).

Figura 25 - Taxonomia FMA da entidade anatômica do sangue.



Fonte: site da FMA.

No desenvolvimento ontológico da HEMONTO as entidades *FMA: Blood* e *FMA: Body* são utilizadas para iniciar algumas taxonomias e partonomias construídas. Tais estruturas de representação do conhecimento vão sendo construídas à medida que outras entidades, não contempladas no FMA e específicas do sangue, são incluídas para formar a representação pretendida.

Um último registro sobre a ontologia FMA é que embora ela possa ser considerada de referência para o desenvolvimento de outras ontologias biomédicas, identifica-se alguns casos de inconsistência em seu conteúdo semântico, principalmente porque ela trata suas classes como entidades anatômicas *padrão*, tal como um atlas anatômico, que descreve um corpo humano ideal, sem nenhuma deficiência, alteração anatômica ou malformação (FREITAS e SCHULZ, 2009). Um exemplo de inconsistência é o axioma “*O trato gastrointestinal inferior tem-parte Apêndice*”. Essa característica do FMA não impede ou atrapalha o uso de suas classes na teoria ontológica do sangue humano, mas atenta-se para este fenômeno de representação de elementos do corpo humano.

Por sua vez, a **Cell Ontology (CL)** é uma ontologia representa o conhecimento biomédico relativo aos tipos de células, incluindo células procariotas, células fúngicas, células animais e células vegetais. Ao todo, incluindo sinônimos, a CL engloba, aproximadamente, 680 tipos de células, descritas em função de suas propriedades funcionais, histológicas, de descendência ou linhagem, entre outras. (BARD, RHEE e ASHBURNER, 2005).

Segundo esses mesmos autores, a motivação para o desenvolvimento da CL veio da necessidade de se definir um vocabulário estruturado de tipos de células para auxiliar na anotação de bases de dados de fenótipos e outros objetos biológicos. Isso ocorre porque os tipos de células são informações necessárias para a descrição de objetos biológicos, tais como os dados de expressões genéticas.

Nesse sentido, a CL foi proposta para descrever os tipos de células dos principais modelos de organismos, que incluem, por exemplo, humanos, ratos, fungos e plantas, como a *Drosófila* e a *Arabidopsis*. Sua construção baseou-se nas regras de desenvolvimento de ontologias definidas no GO Consortium.

A referência a diferentes tipos de organismos é talvez a característica principal que diferencia a *Cell Ontology* de outras ontologias biomédicas (FMA, SNOMED e outras ontologias mais específicas sobre anatomia – exemplo: *American Type Cell Collection (ATCC)*), que também incluem tipos de células em seu conteúdo semântico. Segundo Bard, Rhee e Ashburner (2005), tais ontologias assumem, explicitamente, que os tipos de células representados nelas estão associados a um organismo particular, o que impede a anotação semântica e a descrição de outros organismos.

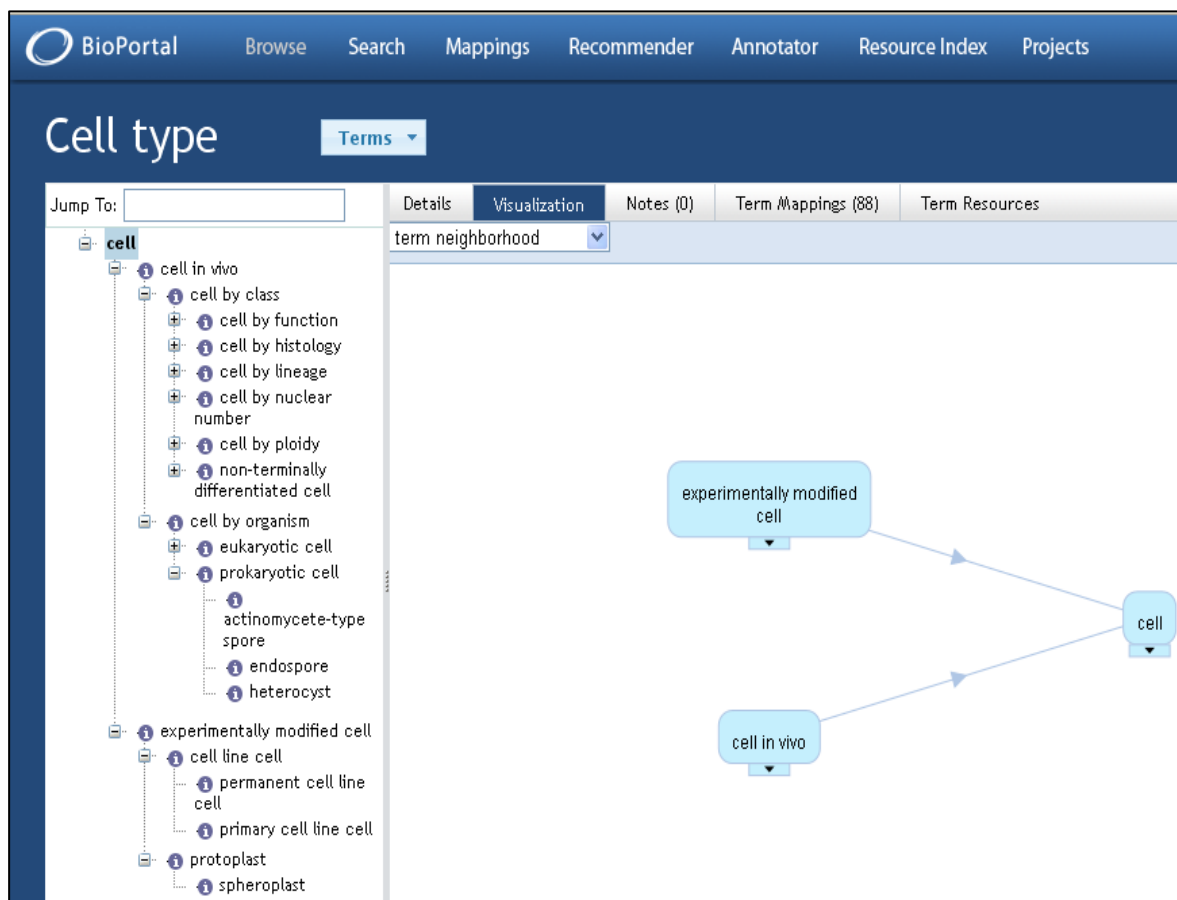
Outros dois aspectos que diferem a *Cell Ontology* dessas outras ontologias referem-se à (BARD, RHEE e ASHBURNER, 2005): (i) primeiramente, o fato de que as ontologias específicas sobre anatomia definem tipos de células como constituintes de tecidos, o que encapsula e limita o conhecimento representado sobre as células; (ii) em

segundo lugar, algumas ontologias dessa área não tem identificadores publicamente disponíveis para cada termo.

Quanto à estrutura semântica da *Cell Ontology*, a mesma inclui uma taxonomia representada através de uma grafo acíclico dirigido que possui, aproximadamente, 10 nós de profundidade, nos quais estão distribuídos as entidades que representam os tipos de células. Tais entidades ou nós são conectados somente através de dois tipos de relações formais: *is_a* e *develops_from*. Como exemplo do uso destas relações, temos que: *chondrocyte is_a mesenchyme_cell*; *a hepatocyte develops_from a mesenchymal_cell*. Além disso, cada termo é representado na *Cell Ontology* através de um identificador único, através da sintaxe CL: número inteiro. Por fim, cada termo da CL contém definições em texto livre e pode ter um ou mais sinônimos, que estão inclusos nesta ontologia.

Parte da estrutura taxonômica da CL, conforme descrito, pode ser visualizada através da Figura 26, a seguir. Tal figura, obtida através do site BioPortal da OBO Foundry apresenta algumas das classes (por exemplo: *eukaryotic cell*, *prokaryotic cell*, *spheroplast*) incluídas na CL e seu posicionamento na taxonomia desta ontologia.

Figura 26 - Estrutura taxonômica da Cell Ontology apresentada por meio do BioPortal.



Fonte: adaptado de <http://bioportal.bioontology.org/ontologies/46604>.

Já a **Protein Ontology (PRO)** é uma ontologia que representa o conhecimento relativo às proteínas e aos complexos protéicos, como resultados de diversos processos biológicos. A representação das classes de proteínas, nesta ontologia, incluem representações estruturadas de proteínas isomórficas, variantes e formas modificadas de diferentes organismos, tais como: humanos, ratos, *Escherichia coli*, além de complexos protéicos (NATALE et al., 2011).

Assim como outras ontologias biomédicas, citadas nesta pesquisa, a *Protein Ontology* é membro da *OBO Foundry*. A PRO está incluída no nível molecular de granularidade da OBO, que corresponde a uma das dimensões em que a OBO é organizada – a outra dimensão refere-se à relação com o tempo (objetos, qualidades e processos). Especificamente, a PRO representa moléculas que contém cadeias de aminoácidos, tais como: proteínas, peptídeos produzidos ao longo da cadeia de aminoácidos e proteínas que contém complexos específicos. A PRO representa cerca de 12 genomas de referência da *Gene Ontology*.

Para representação desses diversos tipos de proteínas, Natale et al. (2011) destaca que a PRO além de definir seus próprios termos, também faz uso de termos contidos em outras ontologias ou bases de dados:

- A *Sequency Ontology*³² (SO) é usada, por exemplo, para definição de termos de partes de proteínas removidas durante a maturação.
- A *Protein Modification Ontology*³³ (PSI-MOD) é usada, por exemplo, na definição de termos referentes aos resíduos de aminoácidos.
- A *Chemical Entities of Biological Interest Ontology*³⁴ (ChEBI) fornece termos usados para definir os objetos PRO que contém anexos não-protéicos e componentes.
- A *Pfam*³⁵ fornece termos usados para descrever regiões de proteínas.
- A *Gene Ontology* é usada para definição de relações como, por exemplo, a disposição que uma determinada proteína tem para desempenhar um processo numa certa localização celular.

³² Disponível em: <http://www.sequenceontology.org/>. Acesso em 28 de setembro de 2012.

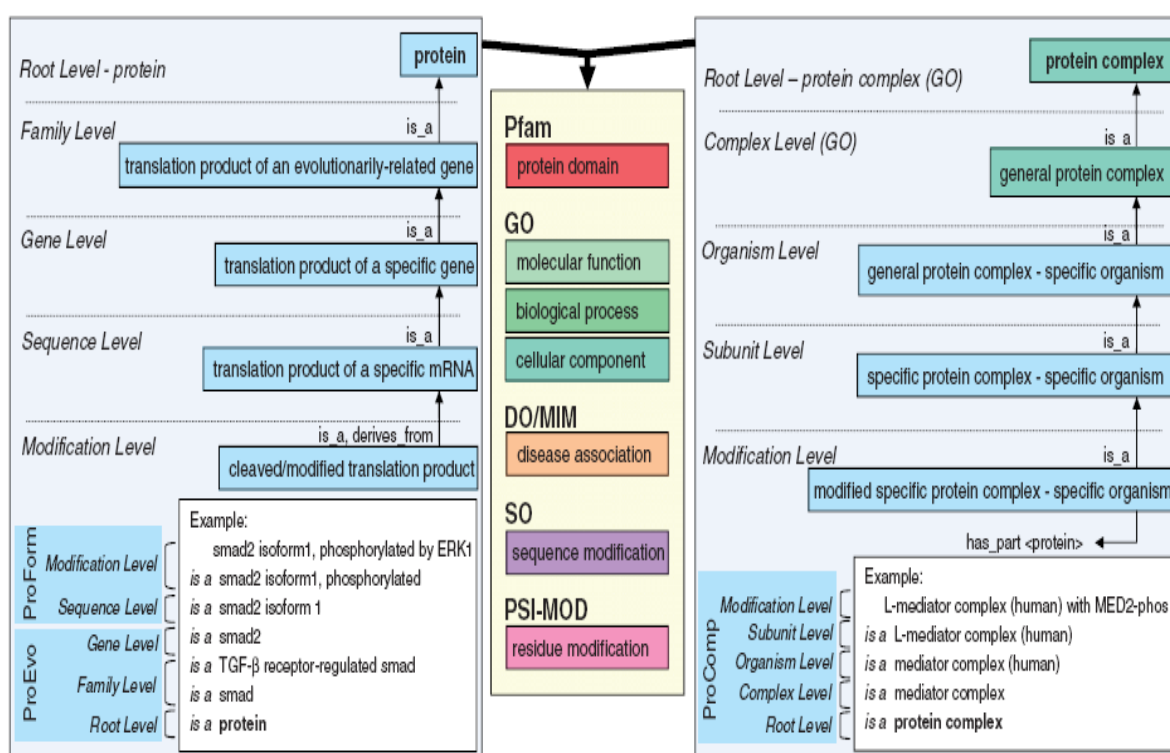
³³ Disponível em: <http://www.ebi.ac.uk/miriam/main/datatypes/MIR:00000056>. Acesso em 28 de setembro de 2012.

³⁴ Disponível em: <http://www.ebi.ac.uk/chebi/>. Acesso em 28 de setembro de 2012.

³⁵ Disponível em: <http://pfam.sanger.ac.uk/>. Acesso em 28 de setembro de 2012.

Além de mencionar o uso de tais ontologias na definição de termos da *Protein Ontology* é importante destacar também que sua estrutura engloba três subontologias: (i) a ProEvo, que inclui proteínas com base em seu parentesco evolutivo; (ii) a ProForm, que engloba as formas protéicas produzidas a partir de um locus de um determinado gene; e (iii) a ProComp, que representa os complexos protéicos. Essas três subontologias da PRO, bem como as outras ontologias utilizadas por ela para definir os termos estão representadas na Figura 27, a seguir, que apresenta a arquitetura geral da *Protein Ontology*.

Figura 27 - Arquitetura geral da Protein Ontology (PRO).



Fonte: adaptado de Natale et al. (2011)

No lado esquerdo da Figura 27 estão representadas as subontologias ProEvo e ProForm, enquanto no lado direito está representada a subontologia ProComp. A ProEvo representa proteínas traduzidas de diferentes genes, porém relacionados. A ProForm representa todas as formas de proteínas codificadas por um único gene. Já a ProComp representa cadeias de aminoácidos específicos que contêm complexos protéicos. Sobre o uso de relações formais para conectar os termos da *Protein Ontology*, a ProEvo e a ProCom utilizam apenas a relação *is_a*, enquanto a ProForm faz uso das relações *is_a* e *derives_from*.

Outra informação representada na arquitetura da figura 20 refere-se às categorias (*Family-level*, *Gene-level*, *Sequence-level*, *Modification-level*) usadas para refletir a hierarquia *is_a* desde a ProEvo até a ProForm. Os termos PRO são rotulados com as categorias que refletem a organização geral da ontologia. Assim, um termo pode ser rotulado como na forma de um complexo ou como organismo específico, de acordo com a necessidade.

Sobre tais categorias ou níveis temos as seguintes definições: (i) *Family-level*: contém os produtos protéicos de uma família genética distinta de seu ancestral comum; (ii) *Gene-level*: termos que se referem a produtos protéicos de um gene diferente; (iii) *Sequence-level*: contém produtos protéicos que possuem uma sequência diferente de sua tradução inicial e (iv) *Modification-level*: termos que se referem a produtos protéicos derivados de uma mesma espécie de RNA e que se diferem devido a alguma mudança após o início da tradução, como, por exemplo, alterações químicas ou clivagem.

Um último registro sobre a *Protein Ontology* é sobre a definição de seus termos. Primeiramente, para cada termo PRO são armazenadas as seguintes informações: identificador (ID) do termo, nome, definição, sinônimo, categoria (classe superior), status (se o termo está obsoleto ou não) e relacionamentos com outros termos. Sobre as definições dos termos elas podem ser encontradas no formato de arquivo da OBO e, além disso, o projeto PRO permite que usuários do sistema de busca³⁶ desta ontologia possam sugerir a inclusão de novos termos nela, os quais passarão por avaliação e poderão ser incluídos em sua base de conhecimento.

Por fim, descreve-se o ***Chemical Entities of Biological Interest (ChEBI)***³⁷, um dicionário, disponibilizado gratuitamente, de entidades químicas de interesse biológico - entidades moleculares - criado no Instituto de Bioinformática Europeu (European Bioinformatics Institute - EBI), em 2002, com o propósito de promover o uso correto e consistente da terminologia biomédica nas bases de dados moleculares do EBI (DEGTYARENKO et al., 2008).

É importante destacar que o ChEBI é um dicionário de “entidades moleculares” focado nos chamados componentes químicos “pequenos” ou “moléculas pequenas”, que são assim chamados para se diferenciar das “macromoléculas”. Desta forma, o ChEBI não inclui em seu conteúdo macromoléculas como ácidos nucleicos, proteínas e peptídeos derivados de proteínas pelo processo de clivagem. Além disso, as “entidades moleculares”

³⁶ Disponível em: <http://pir.georgetown.edu/pirwww/search/>. Acesso em: 28 de setembro de 2012.

³⁷ Disponível em: <https://www.ebi.ac.uk/chebi/>. Acesso em: 22 de fevereiro de 2014.

em questão são produtos naturais ou sintéticos utilizados para intervir nos processos de organismos vivos.

Segundo Degtyarenko et al. (2008), o ChEBI inclui mais de 12000 entidades moleculares, classes de entidades moleculares (por exemplo: os alcanos) e classes de grupos (por exemplo: grupos alquilo). O grupo, em questão, é um conjunto definido de átomos ligados ou um único átomo dentro de uma entidade molecular. Além dos produtos “bioquímicos”, o ChEBI contém também produtos farmacêuticos, agrotóxicos, reagentes de laboratórios, isótopos e partículas subatômicas.

Todo esse conteúdo incluído no ChEBI advém de três fontes de dados, que foram fundidas em uma estrutura única que forma tal dicionário. Essas três fontes de dados são (DEGTYARENKO et al., 2008):

- *IntEnz*: o banco de dados relacional de enzimas integrado do EBI, que contém a Nomenclatura das Enzimas - recomendações do NC-IUBMB para nomear as enzimas.
- *KEGG COMPOUND*: uma parte do banco de dados composto da Kyoto Encyclopedia of Genes and Genomes (KEGG)
- *ChEBI Ontology*: a ontologia incluída no projeto ChEBI, desenvolvida por Michael Ashburner and Pankaj Jaiswal, originalmente com o nome de Chemical Ontology.

Na presente pesquisa, o componente do ChEBI usado para o reaproveitamento de termos na ontologia de produtos do sangue humano (HEMONTO) foi a ChEBI Ontology. Tal ontologia é formada por 4 (quatro) subontologias, que incluem classes e relacionamentos de entidades moleculares específicas: (i) subontologia da Estrutura Molecular; (ii) subontologia de Partículas Subatômicas; (iii) subontologia do Papel Biológico; e (iv) subontologia de Aplicação. A Tabela 13, a seguir, apresenta um fragmento de tais subontologias da ChEBI.

Tabela 13 - Estrutura da ChEBI Ontology

Subontologia	Definição	Exemplos
Estrutura molecular (Molecular structure)	Uma descrição das entidades moleculares ou parte delas, baseada em suas composições e na conectividade entre seus átomos constituintes.	CHEBI:23091 ChEBI ontology ↑◇CHEBI:24431 molecular structure ↑△ CHEBI:23367 molecular entities ↑△ CHEBI:33259 homoatomic molecular entities ↑△ CHEBI:33262 elemental oxygen ↑△ CHEBI:33263 diatomic oxygen

		↑Δ CHEBI:15379 dioxygen ↑Δ CHEBI:26689 singlet dioxygen ↑Δ CHEBI:27140 triplet dioxygen
Partícula subatômica (Subatomic particle)	Partículas menores que um átomo.	CHEBI:23091 ChEBI ontology ↑◇CHEBI:36342 subatomic particle ↑Δ CHEBI:33233 fundamental particle ↑Δ CHEBI:36338 lepton ↑Δ CHEBI:10545 electron
Papel biológico (Biological role)	Papeis realizados por uma entidade molecular ou parte dela em um contexto biológico.	CHEBI:23091 ChEBI ontology ↑◇CHEBI:24432 biological role ↑Δ CHEBI:33280 molecular messenger ↑Δ CHEBI:24621 hormone ↑Δ CHEBI:28918 (R)-adrenaline
Aplicação (Application)	Usos pretendidos de uma entidade molecular ou parte dela pelos humanos.	CHEBI:23091 ChEBI ontology ↑◇CHEBI:33232 application ↑Δ CHEBI:25944 pesticide ↑Δ CHEBI:22153 acaricide ↑Δ CHEBI:38593 fenazaquin

Fonte: adaptado de Degtyarenko et al. (2008).

Cada subontologia da ChEBI Ontology contém uma lista de classes de componentes químicos, as quais são representadas a partir dos seguintes atributos (DEGTYARENKO et al., 2008): (i) *ChEBI ID*: um identificador único e estável da classe; (ii) *ChEBI Name*: o nome da entidade (classe) recomendado para uso pela comunidade biológica; (iii) *ChEBI ASCII Name*: é um ChEBI Name em formato ASCII, que inclui caracteres especiais e requerem uma apresentação em UNICODE. Esse atributo é utilizado pela ChEBI ontology para indicar os relacionamentos semânticos entre as classes desta ontologia. Assim, por exemplo, na Tabela 9, temos a entidade “molecular structure”, cujo *ChEBI ID* é 24431, *ChEBI Name* é a própria “molecular structure” e o *ChEBI ASCII Name* é indicado pelos caracteres ↑ (que indica que esta classe é filha da classe superior) e ◇ (que denota o relacionamento semântico “part_of” entre essa classe e sua superior).

Quanto aos relacionamentos entre as classes da ChEBI Ontology é importante destacar que tal ontologia reutiliza duas relações próprias da Relation Ontology (RO): *is_a* (símbolo na ChEBI: Δ) e *part_of* (símbolo na ChEBI:◇), e inclui também outras relações específicas da ChEBI, tais como: *is conjugate acid of* (exemplo: lactic acid (CHEBI:28358) *is conjugate acid of* lactate (CHEBI:24996)) e *has parent hydride* (exemplo: perfluorodecane (CHEBI:38851) *has parent hydride* decane (CHEBI:32894)).

5 Ontologias: construção, avaliação e implementação

Nos dois capítulos anteriores desta tese, discorreu-se sobre o tema “ontologia” destacando seus conceitos básicos, no capítulo 3, e alguns fundamentos ontológicos e aplicações das ontologias de fundamentação e de domínio para representação do conhecimento, no capítulo 4. O capítulo atual continua explorar o tema “ontologia”, que é central na presente pesquisa, agora apresentando e discutindo as tarefas primordiais no desenvolvimento de ontologias: construção, avaliação e implementação. Sob esses três assuntos, o presente capítulo fornece o background teórico necessário para o entendimento e a justificativa do resultado principal desta tese: uma metodologia própria para construção de ontologias por parte dos cientistas da informação.

Para atingir o propósito deste capítulo, o mesmo foi estruturado da seguinte maneira: a seção 5.1 descreve, brevemente, as atuais metodologias para construção de ontologias nos diversos domínios do conhecimento; a seção 5.2 apresenta uma análise do cenário atual de tais metodologias; a seção 5.3 destaca alguns aspectos essenciais para a construção de ontologias; a seção 5.4 apresenta alguns critérios, métodos e técnicas de avaliação de conteúdos ontológicos; e, por fim, a seção 5.5 aborda alguns aspectos sobre o uso de linguagens lógicas na implementação de ontologias.

5.1 Metodologias para construção de ontologias

Devido ao uso e aplicações de ontologias a partir dos anos de 1990 para representação do conhecimento, foram sendo desenvolvidas a partir de então uma série de metodologias, métodos e propostas de como conduzir a construção de uma ontologia. Através da literatura da área, é possível perceber que houve uma proliferação de metodologias e métodos de construção entre meados da década de 90 até o início dos anos 2000. Nesse período surgiram as metodologias até hoje mais referenciadas por pesquisas na área, tais como: a metodologia de Gruninger e Fox – TOVE (GRUNINGER e FOX, 1995), a *Methontology* (GÓMEZ-PEREZ, FERNANDEZ-LOPES, VICENTE, 1996), o método 101 (NOY e GUINNESS, 2001), entre outras. De meados dos anos 2000 até hoje, surgiram novas metodologias para construção de ontologias, tais como: a metodologia *UP for ONtology* (UPON) (DE NICOLA, MISSIKOFF, NAVIGLI, 2009) e, entretanto, em um número menor se comparado ao período anterior. Na presente seção, são descritas as metodologias de construção mais conhecidas na literatura da área, dando um enfoque maior às atividades das etapas de cada metodologia.

Antes de descrever tais metodologias, é importante destacar a confusão terminológica, identificada na literatura da área, em relação ao uso dos termos

“metodologia” e “método” ao nomear as propostas de construção de ontologias. De acordo com o IEEE (1995), *metodologia* é “uma série integrada de técnicas ou métodos criando uma teoria geral de sistemas de como uma classe de pensamento pode ser executada”. Enquanto *método* “é um conjunto de processos ou procedimentos ordenados usados na engenharia de um produto ou na realização de um serviço”, tendo-se em vista que tais processos são compostos de atividades, que, por sua vez, são compostas por tarefas atribuídas a um ou mais membro do projeto. A partir de tais definições é possível compreender que metodologia e método são conceitos distintos, de forma que uma metodologia determina “como” e “quando” uma dada atividade pode ser realizada, sendo composta de métodos que possuem suas próprias técnicas. Na presente pesquisa, considera-se essa distinção entre esses dois conceitos ao citar as metodologias e métodos para construção de ontologias.

Sobre as metodologias e métodos em si, destacou-se aquelas mais citadas na literatura da área. Para descrição de tais metodologias e métodos, inicialmente, fez-se uma introdução de cada uma, apresentando sua origem, seu propósito principal e domínios de aplicação. Essa breve descrição das metodologias e métodos pesquisados é apresentada na Tabela 14, a seguir.

Tabela 14 - Metodologias para construção de ontologias

Metodologia	Origem e Propósito Principal	Domínio de aplicação
Metodologia de Gruninger e Fox – TOVE (GRUNINGER e FOX, 1995)	Metodologia criada com base no desenvolvimento do projeto <i>Toronto Virtual Enterprise (TOVE)</i> , cujo objetivo era o de criar um modelo de senso comum ou conhecimento compartilhado sobre empresas. Esta metodologia serviu de base para o projeto e avaliação de ontologias integradas em domínios corporativos, incluindo propostas de construção de ontologias e extensões de ontologias já existentes.	Negócios (empresarial)
Metodologia de Uschold e King – ENTERPRISE (USCHOLD e KING, 1995)	Método desenvolvido com base na prática da construção da ontologia de alto nível <i>Enterprise Ontology</i> . Tem como propósito principal descrever o conhecimento sobre domínios corporativos ou de negócios.	Negócios (empresarial)
<i>Methontology</i> (GÓMEZ-PEREZ, FERNANDEZ-LOPES e VICENTE, 1996)	Metodologia desenvolvida no Laboratório de Inteligência Artificial da Universidade Politécnica de Madri entre 1996 e 1997, que possibilita a construção de uma ontologia por reengenharia sobre outra ontologia, utilizando-se do conhecimento do domínio tratado. Tal metodologia pode ser usada, segundo seus autores, em quaisquer domínios do conhecimento, embora seu uso mais conhecido e citado ocorreu na criação de uma ontologia no domínio da química.	Diversos

Método <i>Kactus</i> (BERNARAS, LARESGOTTI, CORERA, 1996)	Método recursivo derivado do projeto Kactus que permitiu a reutilização de conhecimento em sistemas de complexidade técnica, tal como o domínio de redes elétricas, e a construção de ontologias nesse domínio como suporte a tais sistemas.	Sistemas de complexidade técnica
Método <i>Sensus</i> (SWARTOUT et al., 1996).	Método derivado da ontologia <i>Sensus</i> , a qual foi desenvolvida pelo grupo <i>Information Sciences Institute</i> (ISI) com o propósito de ser usada para fins de processamento de linguagem natural. O método <i>Sensus</i> propõe alguns processos para estabelecer as ligações entre os termos específicos e os termos da ontologia de alto nível, que corresponde à ontologia <i>Sensus</i> . Na prática, o método <i>Sensus</i> foi aplicado no desenvolvimento de uma ontologia no domínio de planejamento de uma operação militar aérea.	Diversos
Método 101 (NOY e GUINNESS, 2001)	Método concebido a partir da experiência no desenvolvimento de uma ontologia de vinhos e alimentos, utilizando o editor de ontologias <i>Protégé</i> ³⁸ .	Diversos
Método CYC (REED, LENAT, 2002)	Método usado na construção da ontologia CYC, que considera o conhecimento consensual do mundo e é indicada pelos autores na criação de ontologias para fundamentar sistemas inteligentes.	Diversos
<i>On-to-Knowledge Methodology (OTKM)</i> (SURE, STAAB e STUBER, 2003)	Metodologia desenvolvida para a construção de ontologias para aplicações de gestão do conhecimento, com o foco em Processo de Conhecimento (<i>Knowledge Process</i>) e em Conhecimento do Processo Meta (<i>Knowledge Meta Process</i>). Na prática, tal metodologia foi aplicada em um estudo de caso em gestão de competências de uma empresa internacional localizada na Suíça – a <i>Swiss Life</i> .	Gestão do conhecimento empresarial
Metodologia <i>UP for ONtology</i> (UPON) (DE NICOLA, MISSIKOFF e NAVIGLI, 2009)	Metodologia de construção de ontologias derivada e baseada no padrão de engenharia de software conhecido como Processo Unificado – do inglês <i>Unified Software Development Process</i> ou <i>Unified Process</i> (UP) – do qual foram derivados metodologias de software como o <i>Rational Unified Process</i> (RUP) e o PRocesso para Aplicativos eXtensíveis e Interativos (PRAXIS). Apesar da <i>UP for ONtology</i> poder ser usada em diferentes domínios do conhecimento, segundo seus autores, seu uso mais comum é no domínio do <i>e-bussines</i> .	Negócios (<i>e-bussines</i>)
Metodologia <i>NeON</i>	Metodologia para construção de redes ontológicas baseado	Diversos

³⁸ O *Protégé* é um software para criação, edição e visualização de ontologias. Sua interface permite a importação e exportação de ontologias em alguns formatos, entre eles usando a linguagem recomendada pela W3C – a OWL. Disponível em <http://protege.stanford.edu>. Acesso em: 01 de Maio de 2012.

(SUARÉZ-FIGUEROA, 2010)	em um desenvolvimento colaborativo e argumentativo de ontologias. Tal metodologia foi desenvolvida em uma abordagem híbrida que combina o trabalho metodológico da área de Engenharia de Software e algumas metodologias para construção de ontologias, especificamente, a <i>Methontology</i> , a <i>On-To-Knowledge</i> , a <i>DILIGENT</i> e outros métodos ontológicos, como o de Grüninger e Fox. A <i>NeOn Methontology</i> inclui: (i) o Glossário NeOn de Processos e Atividades (ao todos tem-se 59 processos e atividades definidos), o qual identifica e define os processos e atividades potencialmente envolvidos quando redes ontológicas são construídas colaborativamente; e (ii) duas redes ontológicas sobre os modelos de ciclo de vida	
Metodologia <i>MFPFO</i> (LIM, LIU e LEE, 2011)	Metodologia de construção de ontologia multi-facetada, anotada semanticamente, para a modelagem de uma família de produtos. Tal metodologia é capaz de sugerir, automaticamente, anotações semanticamente relacionadas, baseadas no design e no repositório de construção.	Domínios que possuem uma família de produtos
Ciclo de Vida de Schiessl e Bräscher (2011)	Embora não seja propriamente uma metodologia de construção de ontologias, o ciclo de vida ontológico, descrito por Schiessl e Bräscher (2011), inclui todas as etapas necessárias no processo de construção de ontologias, destacando o papel de cada etapa e as tarefas contidas em cada uma delas.	Diversos

Fonte: elaborado pelo autor.

Apesar de existirem outras metodologias para construção de ontologias na literatura da área, de maneira geral, as metodologias apresentadas na Tabela 14 são as mais referenciadas na área, como já demonstrado em Uschold e Gruninger (1996); Jones, Bench-Capon e Visser (1998); Fernández-López (1999); Fernández-López e Corcho (2004); Silva, Souza e Almeida (2008); Santos (2014) e comprovado por uma simples busca por publicações científicas na área usando a expressão “*methodologies for building ontologies*”.

Dando sequência à descrição de tais metodologias, uma característica fundamental que precisa de ser destacada corresponde às etapas incluídas em cada metodologia e de que forma elas são apresentadas aos ontologistas para sua realização. Essa característica é de extrema importância para a presente pesquisa, uma vez que a partir da análise das etapas incluídas em cada metodologia é possível identificar quais etapas e passos podem ser reaproveitados e quais novas etapas ou novos passos precisam de ser criados na metodologia proposta para direcioná-la para a área de Ciência

da Informação. Nesse sentido, a Tabela 15, a seguir, apresenta, resumidamente, as etapas incluídas em cada uma das metodologias para construção de ontologias pesquisadas.

Tabela 15 - Etapas das metodologias para construção de ontologias

Metodologia	Etapas (processos e atividades)
Metodologia de Gruninger e Fox – TOVE (GRUNINGER e FOX, 1995)	<ol style="list-style-type: none"> 1. Elaboração de cenários de motivação, que objetivam identificar problemas no ambiente atual; 2. Especificação de questões de competência informal, que objetivam especificar em linguagem natural os requisitos que a ontologia deverá ser capaz de atender; 3. Concepção da terminologia formal, em que, mediante declarações em lógica de primeira ordem, os conceitos e suas propriedades são organizados em uma taxonomia; 4. Especificação de questões de competência formal, em que problemas são definidos de modo consistente perante os axiomas na ontologia; 5. Especificação de axiomas formais, que restringem a interpretação dos termos envolvidos nas questões de competência formal; 6. Verificação de teoremas completos, que determinam as condições sobre as quais as soluções das questões são completas.
Metodologia de Uschold e King – ENTERPRISE (USCHOLD e KING, 1995)	<ol style="list-style-type: none"> 1. Identificação do propósito da ontologia, que objetiva identificar a necessidade de construção, o grau de formalismo (desde o informal com uso de linguagem natural até o rigorosamente formal com uso de declarações lógicas) e as classes de usuários da ontologia, incluindo desenvolvedores, mantenedores e usuários das aplicações; 2. Construção da ontologia em três etapas: <ol style="list-style-type: none"> a) captura ou concepção da conceitualização da ontologia; b) codificação ou implementação através de uma linguagem de representação de ontologias; c) integração com ontologias já existentes; 3. Avaliação da ontologia através dos requisitos especificados; 4. Documentação acerca das pretensões da ontologia e das primitivas usadas para expressar as definições na ontologia.
<i>Methontology</i> (GÓMEZ-PÉREZ, FERNÁNDEZ-LOPES e VICENTE, 1996)	<ol style="list-style-type: none"> 1. Especificação: identificar o propósito da ontologia, incluindo os usuários pretendidos, cenários de uso, o grau de formalidade requerido, etc., e o escopo da ontologia, incluindo o conjunto de termos a ser representados. 2. Aquisição de Conhecimento: etapa realizada em paralelo com a etapa (1). Esta etapa é não-prescritiva e, assim, qualquer método pode ser usado, embora seja mais comum as entrevistas com especialistas e análises de textos do domínio tratado. 3. Conceitualização: identificação de termos do domínio, tais como conceitos, instâncias, relações verbais e propriedades, com cada termo tendo uma representação intermediária. 4. Integração: deve-se analisar termos de outras ontologias, tal como o padrão Ontolingua, que possam ser reutilizados na ontologia em construção. 5. Implementação: a ontologia deve ser representada formalmente através de uma linguagem formal, tais como a Ontolingua e a lógica descritiva. 6. Avaliação: são usadas técnicas baseadas nos métodos de validação e verificação dos sistemas de base de conhecimento. Há também um conjunto de diretrizes para avaliar incompletudes, inconsistências e redundâncias. 7. Documentação: recomenda-se a especificação de uma documentação em linguagem da natural ao final de cada fase do ciclo de vida do desenvolvimento da ontologia.
Método <i>Kactus</i> (BERNARAS, LARESGOTTI, CORERA, 1996)	<ol style="list-style-type: none"> 1. Desenvolvimento de uma lista de necessidades ou requisitos que precisam ser atendidos pela aplicação; 2. Identificação de termos relevantes para o domínio da aplicação a partir de tais requisitos, construindo, assim, um modelo preliminar; 3. Refinar e estruturar a ontologia a fim de obter um modelo definitivo; 4. Buscar por ontologias já desenvolvidas por outras aplicações no sentido de sua reutilização.
Método <i>Sensus</i> (SWARTOUT et al., 1996).	<ol style="list-style-type: none"> 1. Identificar termos-chave do domínio; 2. Ligar manualmente os termos-chave à ontologia SENSUS; 3. Adicionar caminhos até o conceito de hierarquia superior da Sensus; 4. Adicionar novos termos para o domínio;

	5. Adicionar subárvores completas.
Método 101 (NOY e GUINESS, 2001)	<ol style="list-style-type: none"> 1. Determinar escopo da ontologia; 2. Considerar o reuso de termos de outras ontologias; 3. Enumerar termos; 4. Definir classes na ontologia; 5. Organizar as classes em uma taxonomia; 6. Definir propriedades (slots) e descrever seus valores permitidos (facetadas), através de restrições; 7. Adicionar valores de slots para as instâncias, isto é, criar instâncias.
Método CYC (REED, LENAT, 2002)	<ol style="list-style-type: none"> 1. Extração manual do conhecimento de senso comum; 2. Extração auxiliada por computador de senso comum; 3. Extração gerenciada por computador de senso comum; 4. Desenvolvimento de representação do conhecimento e ontologia de alto nível (CYC) contendo os conceitos mais abstratos; 5. Representação do conhecimento de diferentes domínios usando tais primitivas.
<i>On-to-Knowledge Methodology (OTKM)</i> (SURE, STAAB e STUBER, 2003)	<p>Conhecimento do Processo Meta (Knowledge Meta Process)</p> <ol style="list-style-type: none"> 1. Estudo de viabilidade: identificação de problemas e oportunidades das áreas e potenciais soluções. 2. Kickoff: compreende dois passos principais: (i) captura da especificação de requisitos a partir do documento de especificação de requisitos da ontologia - Ontology Requirements Specification Document (ORSO); e (ii) criação da descrição de uma ontologia semi-formal. 3. Refinamento: engloba três passos: (i) refinamento da descrição da ontologia semi-formal; (ii) formalização da ontologia semi-formal para ontologia-alvo; (iii) criação do protótipo. 4. Avaliação: consiste na avaliação do projeto em três tipos de avaliação: (i) focada em tecnologia; (ii) focada no usuário; (iii) focada na ontologia. 5. Aplicação e Evolução: nesta etapa tem-se a aplicação da ontologia nos sistemas produtivos ou, mais especificamente, o uso dos sistemas baseados em ontologias. Além dessa tarefa, a evolução da ontologia como um processo organizacional. <p>Processo de Conhecimento (Knowledge Process)</p> <ol style="list-style-type: none"> 1. Geração de conhecimento e/ou importação de documentos e meta-dados. 2. Captura dos itens de conhecimento a fim de elucidar a importância ou a interligação entre eles. 3. Recuperação e acesso aos conhecimentos requeridos. 4. Recuperação dos conhecimentos para posterior utilização em seu contexto.
Metodologia <i>UP for Ontology (UPON)</i> (DE NICOLA, MISSIKOFF e NAVIGLI, 2009)	<ol style="list-style-type: none"> 1. Workflow de Requisitos: envolve as seguintes atividades: (i) determinação do domínio de interesse e escopo da ontologia; (ii) definição do propósito do negócio ou cenário de motivação, com usuários e seus objetivos; (iii) especificação de um ou mais <i>storyboards</i> (contextos de modelagem e situação de maneira narrativa); (iv) criação de um Léxico de Análise, usando ferramentas automáticas para extração de conhecimento de documentos textuais (OntoLearn, Text-to-Onto, etc.); (v) identificação de questões de competência; (vi) identificação e priorização de casos de uso. 2. Workflow de Análise: consiste no refinamento e estruturação dos requisitos ontológicos identificados no workflow anterior. Engloba as seguintes tarefas: (i) aquisição dos recursos do domínio e construção de um Léxico do Domínio; (ii) construção de um Léxico de Referência; (iii) modelagem do cenário de aplicação da ontologia usando diagramas UML; (iv) construção do Glossário de Referência. 3. Workflow de Desenvolvimento: o objetivo deste workflow é construir a estrutura ontológica para o conjunto de entradas do Glossário de Referência. Para tanto, tem-se as seguintes atividades: (i) modelagem de conceitos, categorizando-os em três tipos primários: ator do negócio, objeto do negócio e processo do negócio; e dois complementares: mensagem e atributo; (ii) modelagem de hierarquias de conceitos e relacionamentos específicos do domínio. 4. Workflow de Implementação: consiste em codificar a ontologia em uma rigorosa linguagem formal. Tem-se duas etapas neste workflow: (i) seleção de uma linguagem formal; e (ii) formalização da ontologia nesta linguagem, por exemplo em OWL. 5. Workflow de Teste: o objetivo deste workflow é verificar a qualidade semântica e pragmática da ontologia desenvolvida. A qualidade pragmática é assegurada pela codificação em OWL, enquanto a qualidade semântica é medida pela verificação da consistência da ontologia, realizada da seguinte maneira: verificação da cobertura da ontologia e das respostas da ontologia às questões de competência elaboradas

	anteriormente.
Metodologia <i>NeON</i> (SUARÉZ-FIGUEROA, 2010)	Os 59 processos e atividades contidos no Glossário NeOn de Atividades e Processos são agrupados em três grandes grupos de atividades, conforme a seguir: <ol style="list-style-type: none"> 1. Gerenciamento de processos e atividades: inclui o escalonamento, controle e medida da qualidade da ontologia. 2. Desenvolvimento orientado de processos e atividades: é dividido em três partes: <ol style="list-style-type: none"> a. Pré-desenvolvimento de processos e atividades: (i) estudo do ambiente; (ii) estudo de viabilidade; (iii) reuso de ontologias; (iv) reengenharia ontológica; (v) reuso de recursos não-ontológicos; e (vi) reengenharia de recursos não-ontológicos. b. Desenvolvimento de processos e atividades: (i) especificação de requisitos da ontologia; (ii) conceitualização da ontologia; (iii) formalização; (iv) implementação da ontologia; (v) integração com outras ontologias; (vi) modularização da ontologia; (vii) reestruturação da ontologia; (viii) merge da ontologia; (ix) alinhamento da ontologia; (x) atualização da ontologia; (xi) modificação da ontologia; (xii) localização da ontologia; (xiii) tradução da ontologia; (xiv) anotação da ontologia; (xv) customização. c. Pós-desenvolvimento de processos e atividades: (i) atualização da ontologia; (ii) versionamento da ontologia; (iii) evolução da ontologia. 3. Suporte aos processos e atividades: engloba as seguintes etapas: (i) aquisição de conhecimento; (ii) avaliação da ontologia; (iii) documentação da ontologia; (iv) resumo da ontologia; (v) avaliação da ontologia; e (vi) gerenciamento da configuração da ontologia.
Metodologia <i>MFPFO</i> (LIM, LIU e LEE, 2011)	<ol style="list-style-type: none"> 1. Construção de uma taxonomia da família de produtos; 2. Extração de entidades; 3. Identificação do conceito e geração da unidade facetada; 4. Anotação semântica e modelagem faceta; 5. Construção de uma ontologia de família de produtos multi-facetada e anotada semanticamente; 6. Avaliação e validação da ontologia.
Ciclo de Vida de Schiessl e Bräscher (2011)	<ol style="list-style-type: none"> 1. Especificação: identifica o propósito e o âmbito da ontologia; 2. Conceitualização: descreve, em modelo conceitual, a ontologia a ser construída, de forma que atenda às especificações do passo anterior. O modelo conceitual de ontologia consiste no domínio de conceitos e as relações entre eles. As relações reforçam as conexões mais fortes entre grupos de conceitos. Os grupos de conceitos fortemente relacionados geralmente correspondem a diferentes módulos (subontologias) em que o domínio pode ser decomposto; 3. Formalização: transforma a descrição conceitual em modelo formal, isto é, a descrição do domínio no passo anterior é representada em linguagem formal, ainda que não seja a forma final. Conceitos são normalmente definidos através de axiomas que delimitam as interpretações possíveis para o significado desses conceitos. Conceitos são geralmente organizados hierarquicamente através de uma relação estruturante, tal como “é-um” (classe-superclasse, instância-classe) ou “parte-de”; 4. Aplicação: implementa a ontologia formalizada em linguagem de representação de conhecimento. Para isso, escolhe-se uma linguagem para representação e escreve-se o modelo formal na linguagem escolhida; 5. Manutenção: atualiza e corrige a ontologia aplicada. Reforça-se que o encadeamento das atividades não pretende ser o melhor recurso, pois não é este o foco. Eles são uma opção viável e consolidada por estudos anteriores e, como tudo em ciência, são passíveis de melhoramentos. Logo, esses passos não esgotam o tema, pois há atividades paralelas ao ciclo de vida que podem e devem ser realizadas. São elas: <ul style="list-style-type: none"> • Aquisição de conhecimento: adquire o conhecimento sobre o assunto utilizando técnicas de dedução junto aos especialistas de domínio ou por referência à bibliografia relevante. Várias técnicas podem ser utilizadas para a aquisição de conhecimentos, tais como o brainstorming, entrevistas, questionários, análise de textos, análise e técnicas de indução; • Avaliação: Julgamento técnico, baseado em técnicas disponíveis, da qualidade da ontologia; • Documentação: relata o que foi realizado, como foi feito e o porquê. A documentação associada aos termos representados na ontologia é particularmente importante não apenas para melhorar sua clareza, mas também para facilitar a manutenção, a utilização e a reutilização.

Fonte: elaborado pelo autor.

A revisão realizada sobre as metodologias para construção de ontologias possibilitou à presente pesquisa selecionar àquelas mais apropriadas para a elaboração da metodologia proposta, as quais correspondem a: **Methontology** (FERNÁNDEZ et al., 1997); **Metodologia NeON** (SUARÉZ-FIGUEROA, 2010) e **Método 101** (NOY e GUINESS, 2001).

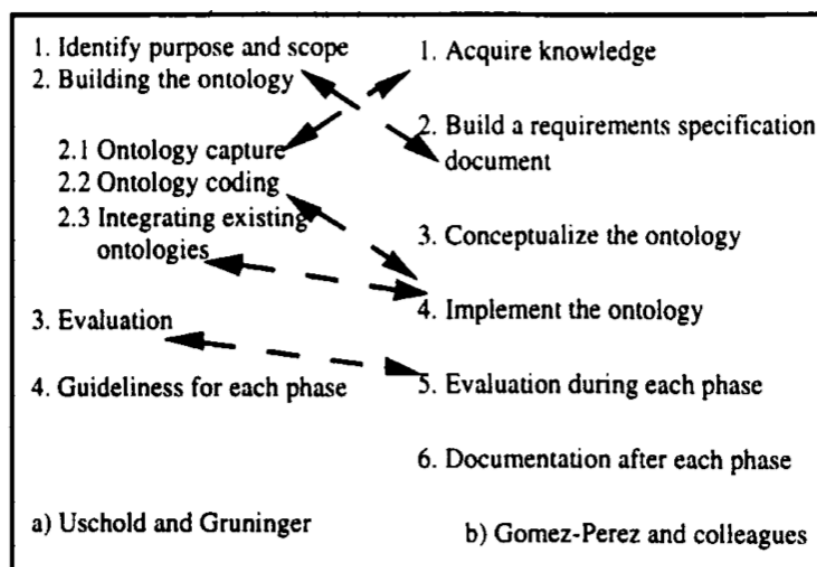
Por serem metodologias mais complexas e selecionadas para uso na proposta desenvolvida, a NeOn e a Methontology são descritas com mais detalhes nas subseções seguintes deste capítulo (5.1.1 e 5.1.2). Por ser um método mais simples, não foi necessário descrever detalhadamente o método 101, tal que as referências aos seus passos podem ser percebidas no guia metodológico da proposta desenvolvida, na metodologia da *OntoForInfoScience*.

5.1.1 Methontology

A Methontology é uma metodologia bem estruturada para a construção de ontologias, que foi desenvolvida, entre os anos de 1996 e 1997, com base na experiência de seus autores na construção de uma ontologia no domínio bioquímico (GÓMEZ-PÉREZ et al., 1996) (FERNÁNDEZ et al., 1997). Como aconteceu com outras metodologias, a Methontology passou a ser utilizada no desenvolvimento de outras ontologias em diferentes domínios do conhecimento e, até hoje, é bem aceita e utilizada na área de engenharia ontológica. Seu desenvolvimento inspirou, por exemplo, a criação da metodologia NeOn, também muito utilizada atualmente.

A elaboração da Methontology começou a partir dos trabalhos de Gómez-Pérez, Fernández e De Vicente em 1996 na modelagem do domínio químico, através do qual definiram um ciclo de desenvolvimento ontológico que incluía seis atividades principais: i) aquisição de conhecimento; ii) construção do documento de especificação de requisitos; iii) conceitualização da ontologia; iv) implementação da ontologia; v) avaliação durante cada fase do ciclo de desenvolvimento; vi) documentação após cada fase. Em uma comparação com a metodologia mais conhecida da época – a metodologia de Uschold e Gruninger (1996) no projeto TOVE – Gómez-Pérez et al. (1996) relacionaram as fases destas duas metodologias de acordo com a Figura 28, a seguir.

Figura 28 - Relação das etapas das metodologias de Uschold e Gruninger e Methontology



Fonte: adaptado de Gómez-Pérez et al. (1996)

No ano seguinte, Fernández, Gómez-Pérez e Juristo (1997) elaboraram a versão final da Methontology, incluindo etapas e passos bem estruturados, além de detalhes do desenvolvimento de uma ontologia. Nessa versão, foi acrescentada a etapa de Integração, que corresponde a reutilização de termos de outras ontologias, às etapas da versão anterior. Embora as etapas da Methontology já tenham sido mencionadas anteriormente nesta pesquisa, essa parte do referencial teórico descreve-as com maiores detalhes, importantes para a elaboração da metodologia *OntoForInfoScience* proposta.

A Methontology engloba sete etapas no ciclo de desenvolvimento de uma ontologia, as quais podem ser descritas da seguinte maneira (GÓMEZ-PÉREZ et al., 1996) (FERNANDÉZ et al., 1997):

1) **Especificação:** deve-se produzir um documento formal da ontologia escrito em linguagem natural, usando um conjunto de representações intermediárias ou questões de competência. As informações que devem estar incluídas nesse documento são:

- a) O propósito da ontologia, incluindo seus usos pretendidos, cenários de uso e usuários finais;
- b) Nível de formalidade da ontologia implementada, que pode ser (USCHOLD e GRUNINGER, 1996): altamente informal, semi-informal, semi-formal ou rigorosamente formal;

c) Escopo da ontologia, que inclua o conjunto de termos que deve ser representado, suas características e granularidade;

Além disso, um bom documento de especificação deve contemplar as seguintes propriedades:

- *Concisão*: todo termo deve ser relevante ao domínio, não havendo termos duplicados e irrelevantes;
- *Compleitude parcial*: está relacionada com a cobertura dos termos e o nível de granularidade de cada termo;
- *Consistência*: todos os termos e seus significados devem fazer sentido no domínio tratado.

2) **Aquisição de Conhecimento**: é considerada uma etapa independente do processo de desenvolvimento da ontologia. Geralmente, é realizada em paralelo com a etapa de especificação. Embora não existam regras específicas para a aquisição de conhecimento do domínio, sugere-se a consulta a especialistas, livros, manuais, figuras, tabelas e outras fontes de conhecimento para ontologia em conjunção com as técnicas de: brainstorming, entrevistas, análise formal ou informal de textos e ferramentas de aquisição do conhecimento.

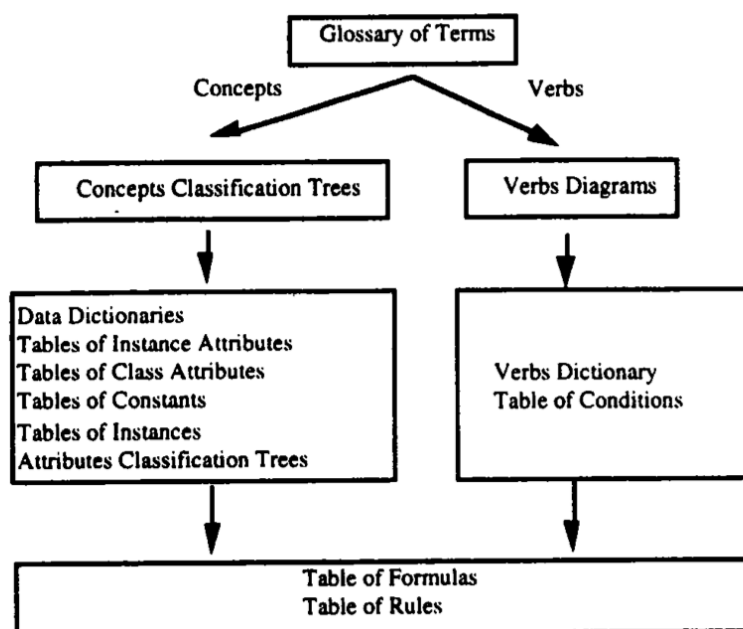
3) **Conceitualização**: etapa correspondente à estruturação do conhecimento do domínio, através de um modelo conceitual que descreva o problema e sua solução em termos do vocabulário do domínio identificado na etapa de especificação da ontologia. Conceitos, instâncias, verbos e propriedades do domínio devem ser definidos através das seguintes representações intermediárias:

- a) Glossário de Termos: incluem conceitos, instâncias, verbos e propriedades;
- b) Dicionário de dados: deve descrever a usabilidade e potencialidade dos conceitos do domínio, seus significados, atributos e instâncias;
- c) Tabelas de atributos-instância: provê informação sobre o atributo ou sobre os valores da instância;
- d) Tabelas de atributos-classe: descreve o próprio conceito e não suas instâncias;
- e) Tabelas de constantes: usada para especificar informações do domínio que tem sempre o mesmo valor;
- f) Tabelas de instâncias: contém as definições de instâncias;
- g) Árvores de classificação de atributos: que mostram graficamente atributos e constantes relacionadas na sequência de inferência desde a raiz;

- h) Dicionário de verbos: expressa o significado dos verbos de maneira declarativa;
- i) Tabelas de condições: especifica um conjunto de condições que devem ser satisfeitas antes da execução de uma ação ou condições garantidas após a execução de uma ação;
- j) Tabelas de fórmulas e regras: reúnem o conhecimento sobre fórmulas e regras.

Esse conjunto de representações intermediárias é visto na metodologia como dois ramos (branches) da árvore de representação, conforme Figura 29 abaixo.

Figura 29 - Conjunto de representações intermediárias da Methontology



Fonte: adaptado de Gómez-Pérez et al. (1996)

4) **Integração:** deve-se considerar o reuso de termos e suas definições de outras ontologias no desenvolvimento da ontologia em construção. Nesta etapa é proposto que:

- a) Se inspecione meta-ontologias, tais como a CYC e a OntoLingua, para selecionar aquela que melhor se adequa a conceitualização do domínio já desenvolvida e, então, fazer a reutilização de termos dessa ontologia selecionada.
- b) Procurar por bibliotecas de ontologias que provêm definições dos termos, os quais sejam coerentes semanticamente com os termos identificados na conceitualização desenvolvida.

O resultado desta etapa é um documento de integração, que descreve para cada termo aproveitado, o nome do termo e a definição do termo na meta-ontologia, o nome da meta-ontologia e o nome do termo na ontologia em construção.

5) **Implementação:** a implementação de ontologias requer o uso de um ambiente que suporta a meta-ontologia e as ontologias selecionadas na fase de integração. O resultado desta etapa é a ontologia codificada em uma linguagem formal, tais como: CLASSIC, BACK, LOOM, Ontolingua, Prolog, C++.

6) **Avaliação:** a metodologia inclui um framework formal para avaliação do conhecimento compartilhado (software, ontologias e documentação), o qual usa técnicas baseadas nos métodos de validação e verificação dos sistemas de base de conhecimento.

Baseado na experiência de verificação de ontologias na Ontolingua, há ainda um conjunto de diretrizes para avaliar incompletudes, inconsistências e redundâncias apresentadas na ontologia.

7) **Documentação:** a principal estratégia na documentação da Methontology é a especificação de uma documentação em linguagem da natural ao final de cada fase do ciclo de vida do desenvolvimento da ontologia. Assim, fazem parte da documentação desta metodologia: i) um documento de especificação de requisitos; ii) um documento de aquisição do conhecimento; iii) um documento do modelo conceitual; iv) um documento de formalização; v) um documento de integração; vi) um documento de implementação; e vii) um documento de avaliação.

5.1.2 Metodologia NeOn

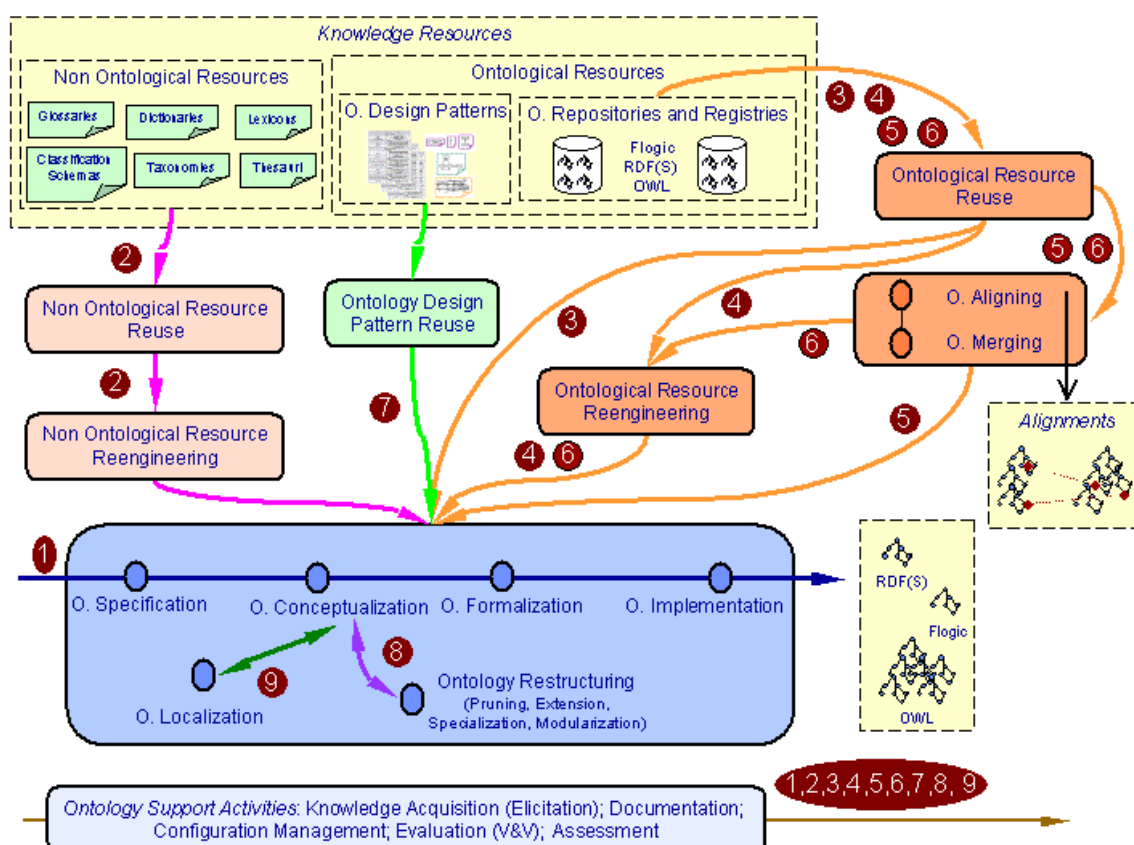
A *NeOn Methodology* se caracteriza por ser uma metodologia para construção de redes ontológicas³⁹ baseado no desenvolvimento colaborativo e argumentativo de ontologias, que inclui um glossário de processos e atividades e duas redes ontológicas sobre os modelos de ciclo de vida (SUÁREZ-FIGUEROA et al., 2008). Recentemente na área de engenharia ontológica, a metodologia *NeOn* tem sido amplamente utilizada no desenvolvimento de ontologias direcionadas à área de engenharia de software, por ter um

³⁹ Uma rede ontológica é uma estrutura complexa formada por algumas ontologias mais ou menos conectadas. Estas conexões podem estar em forma de meta-relações entre as ontologias ou também a partir de mapeamentos e alinhamentos entre pares de elementos individuais das ontologias (por exemplo: conceitos, propriedades e instâncias) (SUÁREZ-FIGUEROA et al., 2009).

conteúdo bastante completo quanto às atividades necessárias no ciclo de vida de um software desenvolvido com base em ontologias.

Com base em estudos de caso na construção de ontologias em diferentes projetos nacionais e internacionais, a metodologia *NeOn* traçou um panorama de um conjunto de cenários possíveis de construção de ontologias e redes ontológicas, conforme pode ser visto na Figura 30, a seguir, extraída de Suárez-Figueroa (2008).

Figura 30 - Cenários para construção de redes ontológicas pela NeOn



Fonte: adaptado de Suárez-Figueroa (2008)

Na arquitetura da metodologia *NeOn* (Figura 30) estão apresentados os 9 (nove) cenários para construção de redes ontológicas, os quais correspondem a:

- *Cenário 1*: Construção de redes ontológicas heterogêneas sem o reuso de recursos ontológicos existentes.
- *Cenário 2*: Construção de redes ontológicas pelo reuso e reengenharia de recursos não ontológicos.

- *Cenário 3*: Construção de redes ontológicas pelo reuso de recursos ontológicos.
- *Cenário 4*: Construção de redes ontológicas pelo reuso e reengenharia de recursos não ontológicos.
- *Cenário 5*: Construção de redes ontológicas pelo reuso e fusão (*merging*) de recursos ontológicos.
- *Cenário 6*: Construção de redes ontológicas pelo reuso, fusão (*merging*) e reengenharia de recursos ontológicos.
- *Cenário 7*: Construção de redes ontológicas pelo reuso de padrões de desenvolvimento ontológico (*ontology design patterns – ODPs*).
- *Cenário 8*: Construção de redes ontológicas pela reestruturação de recursos ontológicos.
- *Cenário 9*: Construção de redes ontológicas pela localização de recursos ontológicos.

Na arquitetura NeOn da Figura 30, processos e atividades a serem realizados estão representados pelos círculos coloridos ou caixas arredondadas. As setas direcionadas com círculos numerados associados representam os diferentes cenários descritos nesta metodologia, os quais incluem processos e atividades como: especificação ontológica, reuso de recursos ontológicos e também não ontológicos, reuso de padrões de desenvolvimento ontológico (ODPs), entre outros. Já as caixas pontilhadas apresentam fontes de conhecimento a serem reutilizadas e possíveis outputs (alinhamentos e redes ontológicas implementadas) que resultam da execução de alguns dos cenários apresentados.

A Figura 30 representa também que os cenários especificados pela NeOn podem ser combinados de diferentes maneiras como, por exemplo: o cenário 2 (reuso e reengenharia de recursos não ontológicos) pode ser combinado com os cenários 3 a 8; e o cenário 9 (localização de ontologias) pode ser realizado ou não junto com os cenários 1 a 8. É importante destacar também que a NeOn distingue cenários que envolvam reuso de recursos ontológicos de outros cenários que envolvam reuso e reengenharia de recursos não ontológicos.

Por fim, outra importante informação contida na Figura 30 diz respeito às atividades de suporte ontológico. É possível observar que as atividades de aquisição de conhecimento (elicitação), documentação, gerenciamento da configuração, avaliação e análise ontológicas devem ser realizadas durante todo o desenvolvimento da rede

ontológica. E é na etapa de avaliação e análise ontológica que se concentra o foco da presente pesquisa com relação à metodologia *NeOn*.

5.2 Análise do cenário atual das metodologias

Apresentadas as metodologias para construção de ontologias mais citadas e referenciadas na literatura da área, o passo seguinte deste referencial teórico foi realizar uma análise do cenário atual das metodologias, destacando problemas identificados tanto na literatura da área quanto na análise realizada a partir de prática, além de considerações sobre o uso de tais metodologias em pesquisas científicas. Em tal análise, foram destacadas falhas e omissões das atuais metodologias e também seus conteúdos que precisam ser melhor explicados aos desenvolvedores de ontologias, incluindo os cientistas da informação.

A partir da revisão de literatura sobre deficiências e limitações no desenvolvimento de ontologias, pode-se afirmar que os maiores e mais recorrentes problemas encontrados nas metodologias atuais para construção de ontologias são:

- I. **Não há um padrão de construção ou uma metodologia unificada** que seja amplamente aceita para o desenvolvimento de ontologias (USCHOLD e GRUNINGER (1996); (JONES, BENCH-CAPON e VISSER (1998); FERNANDÉZ-LOPEZ (1999); FERNANDÉZ-LOPEZ e CORCHO (2004); BREITMAN (2005); CARDOSO (2007); SILVA, SOUZA e ALMEIDA (2008)). Tais metodologias, ainda hoje, não são consideradas suficientemente “maduras” como as metodologias empregadas na área de Engenharia de Software.
- II. As metodologias atuais apresentam **abordagens e características diversas, sendo direcionadas a propósitos e aplicações específicas** (FERNANDÉZ-LOPEZ (1999)), tais como aquelas dedicadas ao domínio dos negócios.
- III. **Muitas das metodologias atuais assumem uma tarefa como ponto de partida na construção da ontologia**, por exemplo, a forma de aquisição de conhecimento ou avaliação a partir de questões de competência. Se por um lado essa estratégia é interessante, por outro ela não pode vir a ser quando ela impõe limitações na reutilização da ontologia (JONES, BENCH-CAPON e VISSER, 1998). Esses autores acrescentam ainda que amarrar uma ontologia a uma tarefa parece admitir a derrota diante do problema da

interação, ou seja, ontologias de natureza distinta devem ser construídas de maneiras diferentes.

- IV. **As atividades necessárias para reutilização e integração de ontologias**, por exemplo, em um repositório ontológico, **são ainda tarefas bastante complexas e para as quais ainda não existe um padrão amplamente aceito para realizá-las**. Jones, Bench-Capon e Visser (1998) destacam a importância de se selecionar uma ontologia adequada e de como estender a ontologia nesse processo. Eles afirmam que para a seleção de uma ontologia precisa-se de uma resposta clara a partir de uma metodologia satisfatória para não levar a uma estrutura inadequada e que a extensão de ontologias, embora bastante abordada nas metodologias, ainda é uma questão complexa a se tratar.
- V. Em relação a detalhes das atividades e dos procedimentos para sua condução, algumas **metodologias e métodos mostram-se superficiais na elucidação dos passos para construção de ontologias**, parecendo até que considerar que o ontologista já domina o assunto sobre construção de ontologias e não necessita de detalhes acerca de atividades e procedimentos envolvidos (SILVA, SOUZA e ALMEIDA, 2008). Um exemplo disso é a ausência de orientações para ajudar o ontologista a fazer escolhas ontológicas sobre a inclusão ou não de um dado conhecimento na ontologia.
- VI. **Algumas abordagens dão mais ênfase em atividades de desenvolvimento**, especialmente a implementação da ontologia (método Cyc e método 101), desconsiderando aspectos importantes relacionados a gerenciamento do projeto, a estudo de viabilidade, à manutenção e à avaliação de ontologias (SILVA, SOUZA e ALMEIDA, 2008).
- VII. Outra questão em aberto sobre as metodologias de construção de ontologias diz respeito ao seu **grau de maturidade**. Para alguns autores, como, por exemplo, Fernández-López (1999), o grau de maturidade de uma metodologia para ontologias deveria ser medido a partir de sua similaridade com os padrões de desenvolvimento de software. Segundo esse autor, o processo de desenvolvimento de ontologias deveria ser fundamentado no padrão internacional IEEE-1074, com algumas adaptações particulares às ontologias. O grande problema dessa abordagem é que esse autor e outros que defendem tal posição consideram uma ontologia como componente de software. Nesse sentido, essa abordagem só funcionaria para ontologias computacionais e não toda a gama de ontologias existentes.

Além desses problemas mencionados, um outro ponto importante a ser analisado sobre as metodologias para construção de ontologias refere-se às etapas incluídas em cada uma delas.

A consulta às etapas das metodologias revisadas nesta pesquisa (seção 5.1) permite verificar que existem diferenças claras entre elas, sendo algumas mais completas e detalhadas, tais como a *Methontology* e a metodologia de Gruninger e Fox - TOVE, e outras mais simples, que dão ênfase a aspectos de desenvolvimento e implementação da ontologia, tais como os métodos 101, CYC e Kactus.

Sobre o uso de tais metodologias na construção de ontologias de domínio, Santos (2014) traça um panorama bastante interessante de como as metodologias atuais vem sendo usadas em pesquisas sobre esse tema e com qual frequência. Segundo essa autora, três metodologias têm se destacado mais na construção de ontologias de domínio: o método 101, a *On-to-knowledge* e a *Methontology*. Além disso, as metodologias de construção atuais não estão sendo usadas isoladamente, na maioria das vezes, são usadas em parceria com outras metodologias como forma de suprir as limitações existentes em cada uma delas, como por exemplo, o uso em conjunto das metodologias *Methontology*, *On-to-knowledge* e o método 101. Como, em geral, algumas metodologias são mais detalhadas em algumas etapas que outras, o uso em conjunto de várias metodologias para construção ontológica parece ser a solução mais viável.

Outra conclusão importante destacada por Santos (2014) é que as metodologias próprias estão ganhando espaço na literatura da área. Os pesquisadores estão fazendo adaptações e elaborando novas metodologias que se adequem as necessidades da área, que inclui esse uso em conjunto (unificação) de diferentes metodologias.

Inserida nesse cenário atual das metodologias de construção, a presente pesquisa propõe uma metodologia própria para o desenvolvimento de ontologias, principalmente, para cientistas da informação e realiza uma aplicação prática desta metodologia – a *OntoForInfoScience* – no desenvolvimento de uma teoria ontológica do sangue.

A ideia por trás da metodologia *OntoForInfoScience* é reutilizar algumas das etapas das metodologias NeOn, *Methontology* e método 101, buscando suprir as limitações de cada uma com sua reutilização em conjunto, além de detalhar etapas destas metodologias que não incluem detalhes e fazer as adaptações necessárias de uma metodologia para a área de Ciência da Informação. A partir dessa estratégia espera-se conseguir uma metodologia útil e bem aceita entre os cientistas da informação.

Cabe ainda citar os motivos que levaram à escolha dessas três metodologias como referências para a elaboração da *OntoForInfoScience*. A partir das metodologias de construção pesquisadas foi realizada uma análise qualitativa das etapas de cada uma, considerando os seguintes aspectos: (i) relevância e utilidade prática da metodologia em trabalhos da área de engenharia ontológica; (ii) quantidade suficiente de etapas para o ciclo de desenvolvimento ontológico; (iii) clareza e detalhamento das etapas; e (iv) coerência na sequência de disposição das etapas. A análise sob esses aspectos levou à escolha das três metodologias, conforme descrito abaixo:

- **Methontology (FERNÁNDEZ et al., 1997)**: foi escolhida por reunir todas as etapas básicas para o desenvolvimento de ontologias, explicadas de uma forma clara, sequencial e resumida, que possibilita o entendimento necessário do conteúdo mínimo e essencial que uma metodologia de construção precisa ter.
- **Metodologia NeON (SUARÉZ-FIGUEROA, 2010)**: foi selecionada por ser considerada a metodologia para construção de ontologias mais completa dentre todas pesquisadas e, talvez, por esse motivo ser a mais usada atualmente na área de engenharia de software. Seus processos e atividades detalhados corroboram para a elaboração da metodologia proposta, com um grande detalhe de que é necessária sua adaptação para o contexto da Ciência da Informação, já que ela usa recorrentemente conceitos oriundos da Ciência da Computação.
- **Método 101 (NOY e GUINNESS, 2001)**: apesar de ser um método mais simples, baseado unicamente na experiência de desenvolvimento com o software Protégé, o método 101 é bastante útil na explicação de passos mais específicos na construção de ontologias, como, por exemplo, a definição de propriedades das classes e relações, criação de restrições de dados, de cardinalidade e de instâncias, entre outros. Portanto, o detalhamento de passos específicos de etapas do desenvolvimento ontológico, não contemplado nas metodologias NeOn e Methontology, é feito com base nos passos do Método 101.

5.3 Aspectos essenciais na construção de ontologias

Como o objetivo principal desta pesquisa é a proposição de uma metodologia própria para a construção de ontologias, torna-se imprescindível levantar e discutir alguns aspectos essenciais para a construção de ontologias fundamentadas em princípios formais

e filosóficos. Nesse sentido, esta parte do trabalho tem como propósito sugerir alguns aspectos importantes no desenvolvimento de ontologias.

O levantamento de aspectos considerados essenciais no processo de construção de ontologias foi possível após a revisão realizada sobre as metodologias atualmente disponíveis. Como já apresentado, procedeu-se com uma análise qualitativa de tais metodologias que possibilitou a identificação das etapas e atividades que são indispensáveis a qualquer ciclo de desenvolvimento ontológico e também as atividades no processo de construção de ontologias que precisam de um maior detalhamento e adaptação ao contexto da Ciência da Informação. Os aspectos levantados referem-se a essas atividades do desenvolvimento ontológico, em geral.

Para apresentar tais aspectos esta seção da tese foi dividida em dois tópicos: a subseção 5.3.1 descreve aspectos gerais considerados essenciais no ciclo de desenvolvimento de uma ontologia e a subseção 5.3.2 trata da caracterização das relações de uma ontologia em seu desenvolvimento, de forma a abordar os tipos de relações existentes e as diferenças entre elas.

5.3.1 Aspectos gerais

Na literatura da área não é possível identificar um número significativo e relevante de pesquisas que abordem aspectos ontológicos essenciais a serem incluídos em metodologias para construção de ontologias. Considerando esse cenário, a presente pesquisa fez um levantamento, com base na literatura e na aplicação prática de ontologias, de alguns aspectos essenciais no desenvolvimento de ontologias. Não se pretende aqui que tais aspectos sugeridos sejam um padrão ou referência para a construção de ontologias em quaisquer domínios, mas que eles sirvam como um instrumento de suporte para a elaboração de metodologias próprias para o desenvolvimento de ontologias, tal como a *OntoForInfoScience*.

Os aspectos gerais considerados essenciais no processo de construção de ontologias, segundo a pesquisa realizada, são:

- I. Avaliação da real necessidade do uso da ontologia para o problema analisado;
- II. Delimitação clara e explícita do escopo de cobertura da ontologia;
- III. Consideração do uso de ontologias de fundamentação no desenvolvimento
- IV. Utilização de algum tipo de colaboração na fase de conceitualização;
- V. Reuso de termos de outras ontologias existentes e de que forma se dá esse reuso;

- VI. Existência de uma descrição informal e outra formal da ontologia;
- VII. Criteriosidade na transição da descrição informal da ontologia para a descrição formal;
- VIII. Completude mínima necessária do conteúdo ontológico;
- IX. Utilização de critérios, métodos ou técnicas para avaliação do conteúdo da ontologia;
- X. Documentação da ontologia ao final de cada etapa.

Os parágrafos seguintes desta seção concentram-se em explicar esses aspectos essenciais.

O **primeiro aspecto** é a inclusão na metodologia de construção de uma **avaliação prévia da real necessidade do uso da ontologia** para o problema e contexto analisados. Essa avaliação prévia, especialmente na área de Ciência da Informação, se faz necessária porque nem todos os problemas a serem tratados demandam o uso de uma ontologia, por exemplo, a indexação e recuperação de informações em um domínio específico. Nesse exemplo, um tesouro do domínio seria suficiente. Esse é apenas um exemplo de outros casos que existem na área, onde nem sempre precisa-se de uma representação formal e o uso de relações extensíveis entre os conceitos do domínio - características próprias de uma ontologia. Nesse sentido, o objetivo com esse primeiro aspecto é ressaltar a importância da existência deste tipo de avaliação em uma metodologia de construção, de forma a orientar o desenvolvedor nessa atividade.

O **segundo aspecto** refere-se a **delimitação clara e explícita do escopo** de cobertura da ontologia. Esse aspecto é levado em consideração na maioria das metodologias de construção atuais. A Methontology (FERNÁNDEZ et al., 1997), a metodologia de Gruninger e Fox (1995), a metodologia de Uschold e King (1995) e o método 101 (NOY e MCGUINNES, 2001), por exemplo, citam a definição do escopo da ontologia como uma etapa de especificação, na qual são identificados os propósitos, usuários, graus de formalismo e aplicações da ontologia. Além disso, essas metodologias recomendam o uso de questões de competência na definição do escopo, tal que elas são utilizadas para especificar tarefas e os problemas que uma ontologia pode solucionar, antes mesmo de sua construção. Trata-se, então, de uma forma de auxiliar na definição do escopo e das características da ontologia.

De fato, todas essas atividades previstas nessas metodologias, tais como a elaboração de questões de competência, identificação de propósitos e aplicações da ontologia, etc.; são imprescindíveis na definição do escopo de cobertura. Entretanto, é necessário complementar essas atividades com uma delimitação clara e explícita do

escopo, de forma a assegurar o ponto de partida do domínio que a ontologia cobre até o limite de cobertura. Nesse sentido, uma delimitação clara e explícita significa especificar formalmente o ponto de partida e o limite do domínio cobertos pela ontologia, incluindo os níveis de conhecimento que serão abordados na ontologia.

O **terceiro aspecto** sugerido é a avaliação da possibilidade de **uso de ontologias de fundamentação** na construção da ontologia pretendida. Como as ontologias de fundamentação ou de alto nível servem para descrever conceitos gerais (espaço, objeto, ação, etc.) independentes do problema ou domínio, elas geralmente são usadas como base ou ponto de partida para a construção de ontologias sobre domínios específicos. Como citado anteriormente, esse tipo de abordagem é uma prática de pesquisa incentivada na literatura sobre ontologias, especialmente as biomédicas (SMITH, 2005; BITTNER e DONNELLY, 2007; DEGEN et al., 2001; MASOLO et al., 2003; GRENON e SMITH, 2004; HERRE, 2010). Dentre as contribuições do uso de ontologias de fundamentação na construção de outras ontologias está o fato de que as primeiras possibilitam distinguir diferenças que são muito sutis nas decisões ontológicas a serem tomadas no processo de construção, tornando tais decisões mais transparentes possível.

As metodologias para construção de ontologias pesquisadas não incluem explicitamente como uma de suas etapas a avaliação de ontologias de fundamentação como ponto de partida para a construção da ontologia requerida. Já na metodologia OntoForInfoScience, tal tarefa é considerada um aspecto ontológico essencial, pelos motivos anteriormente citados, e é uma das primeiras etapas no processo de construção de ontologias. Nesse sentido, após a definição do escopo de cobertura da ontologia, recomenda-se uma pesquisa sobre as atuais ontologias de fundamentação, de forma a escolher qual delas é mais adequada no projeto em desenvolvimento, uma vez que tais ontologias discordam em alguns princípios ontológicos de construção.

O **quarto aspecto** essencial é a **utilização de algum tipo de colaboração na fase de conceitualização**. A palavra “colaboração” nesse contexto significa o processo de compartilhamento de informação e conhecimento entre especialistas de um determinado domínio que, em geral, possuem diferentes visões de mundo, as quais às vezes representam visões concorrentes. Na presente pesquisa, a atividade de conceitualização na construção da ontologia é compreendida do ponto de vista da semântica cognitiva, para a qual a conceitualização é definida como um processo dinâmico de construção do significado (EVANS e GREEN, 2006). Sob esse ponto de vista, torna-se indispensável utilizar a colaboração entre especialistas na conceitualização do domínio tratado, de forma a possibilitar a identificação, análise e negociação dos conceitos que devem ser incluídos

na ontologia, bem como suas características, definição, relacionamentos e estruturação hierárquica.

Portanto, o que se pretende com esse quarto aspecto é ressaltar a importância da inclusão de atividades colaborativas na fase de conceitualização do desenvolvimento ontológico. Tudorache et al. (2008) destaca que as *conceitualizações colaborativas* têm sido adotadas e incentivadas no ambiente web, por exemplo, através de ontologias desenvolvidas colaborativamente, como forma de fomentar a discussão e proporcionar a produção de sistemas realmente representativos das necessidades de uma comunidade. Entretanto, Pereira (2010) afirma que, atualmente, as pesquisas na área de engenharia ontológica ainda geram conhecimento insuficiente sobre métodos e técnicas de apoio à construção colaborativa para conceitualização, sendo, portanto, necessária uma maior exploração e incentivo desta prática.

Na presente pesquisa, a estratégia adotada para incluir e tratar esse aspecto na metodologia proposta foi a utilização do framework *Conceptualization Modelling Environment* (ConceptME), o qual oferece um conjunto de recursos e ferramentas para o desenvolvimento colaborativo de representações conceituais semi-formais e compartilhadas (SOUSA et al., 2013). O ConceptME foi desenvolvido tendo como base um método chamado de *ColBlend* (PEREIRA et al., 2012), que oferece suporte ao processo de conceitualização colaborativa no contexto inter-organizacional e é baseado em uma teoria da semântica cognitiva conhecida como *Conceptual Blending Theory* (Fauconier e Turner, 1998). Sob tal método e tal teoria, o ConceptME utiliza o conceito de espaços semânticos como espaços de ação social dentro dos quais se acumulam significados diferentes e que devem servir como base de colaboração que permita o entendimento necessário para ação coletiva. São nesses espaços semânticos que as pessoas envolvidas no processo de conceitualização participam da atividade de negociação de significados do domínio e, ao final dela, produzem um modelo conceitual comum e aceito por todas as partes envolvidas.

O **quinto aspecto** essencial refere-se ao **reuso de termos de outras ontologias** existentes e de que forma se dá esse reuso na ontologia em desenvolvimento. De forma análoga à definição do escopo da ontologia, esse aspecto é também contemplado pela maioria das metodologias de construção atuais, uma vez que o reaproveitamento de termos, incluindo classes e relações, é um dos princípios universalmente aceitos na área de engenharia ontológica. O ponto crítico nesse reuso de termos na ontologia é a forma pelo qual ele é conduzido.

Uma das tarefas mais difíceis para o desenvolvedor da ontologia é tomar a decisão mais apropriada na escolha do termo a ser reaproveitado, em função de um número relativamente grande de ontologias disponíveis e de muitos sinônimos e

ambiguidades presentes em tais ontologias. Para auxiliar nessa tarefa, os desenvolvedores de ontologia contam, atualmente, com ferramentas de busca em ontologias, tais como: o portal do *National Center for Biomedical Ontology*⁴⁰ (NCBO), o *Ontology Lookup Service*⁴¹ (OLS), o *Ontobee*⁴², entre outras; que lhes possibilitam buscar o termo pesquisado em repositórios de ontologias e conhecer suas propriedades (definição, sinônimos, ancestral na hierarquia, etc.). Após a seleção do termo a ser reaproveitado, o desenvolvedor pode também utilizar editores de ontologia, tal como o Protégé, para importação do termo para a ontologia ou até mesmo importar um ramo inteiro (parte da taxonomia) de uma outra ontologia para aquela em desenvolvimento.

De maneira geral, esses detalhes no reaproveitamento do uso de termos para ontologia não são abordados nas metodologias de construção pesquisadas. Desta forma, esse quinto aspecto é considerado na metodologia proposta, a qual visa auxiliar o desenvolvedor nesse processo de reaproveitamento de termos.

O **sexto aspecto** corresponde à existência de uma **descrição informal e outra formal da ontologia** em desenvolvimento. Entende-se como descrição informal da ontologia a representação textual e gráfica da conceitualização do domínio, que inclui, por exemplo, o dicionário de classes da ontologia, as definições textuais de cada classe, as taxonomias das classes, entre outros. Já a descrição formal da ontologia corresponde à representação do conteúdo ontológico em uma linguagem formal, tais como a lógica descritiva, a lógica de primeira-ordem, a lógica da linguagem OWL, que fornecem recursos para a especificação de axiomas, propriedades e questões de competência da ontologia de maneira formal.

Esse aspecto é incluído quase em todas atuais metodologias de construção de ontologias e, muitas vezes, referenciado como a etapa de implementação da ontologia. Ele é de fundamental importância para assegurar que a ontologia possa englobar todo o conhecimento a ser coberto do domínio e para tornar possível sua interpretação pelas máquinas, através de inferências ao conteúdo formal da ontologia. Para tanto, um aspecto primordial nessas descrições formal e informal da ontologia é distinguir qual parte do conhecimento acerca do domínio pode ser considerado ontológico e assim representável formalmente e qual parte desse conhecimento é descritivo e deve ser representado em forma de anotações e comentários na ontologia. Embora essa distinção seja fundamental no processo de construção de ontologias, a maioria das metodologias existentes apenas

⁴⁰ Disponível em: <http://bioportal.bioontology.org/>. Acesso em: 08 de Março de 2014.

⁴¹ Disponível em: <http://www.ebi.ac.uk/ontology-lookup/>. Acesso em: 08 de Março de 2014.

⁴² Disponível em: <http://www.ontobee.org/>. Acesso em: 08 de Março de 2014.

prescreve a existência dos dois tipos de descrição, não explicando como o conhecimento do domínio pode ser representado ou não em cada um dos tipos.

O **sétimo aspecto** considerado essencial corresponde à **criteriosidade na transição da descrição informal da ontologia para a descrição formal**. O que se quer dizer com “criteriosidade” nesse caso é tornar o mais equivalente possível a descrição formal da descrição informal, uma vez que a primeira é derivada da segunda em ontologias. Sobre esse tema, Jones, Bench-Capon e Visser (1998) explicam que as ontologias têm como propósito ajudarem a estreitar a lacuna entre um sistema formal (executável) e os modelos conceituais do mundo real, fazendo isso através de umnexo identificável entre as descrições formais e informais. Pela análise feita das metodologias de construção pesquisadas, nenhuma delas deixa claro como deve ser feita essa transição entre a descrição informal e a formal.

Um exemplo claro de como ser criterioso nessa transição entre os dois tipos de descrição de uma ontologia está na elaboração das definições textuais (em linguagem natural) das classes de uma ontologia. Num primeiro momento pode parecer trivial criar uma definição em linguagem natural para um termo (classe) da ontologia, uma vez que o pensamento ingênuo sugere que uma definição em linguagem natural é apenas a descrição do significado do termo. Se consultarmos alguns dos mais conhecidos vocabulários médicos e algumas conhecidas ontologias, teremos essa falsa impressão de que as definições de classes de uma ontologia em linguagem natural pode ser feita como uma descrição textual livre e sem muitas restrições. Entretanto, se o objetivo é ser criterioso nessa transição entre a descrição informal e a formal da ontologia, as definições textuais das classes devem ser elaboradas de forma a facilitar suas definições em uma linguagem lógica. O uso de algum método, tal como o método Aristotélico (MUNN e SMITH, 2008), por exemplo, para a criação das definições textuais é ótimo caminho nessa transição para as definições formais (em linguagem lógica).

O **oitavo aspecto** essencial é a **completude mínima necessária do conteúdo ontológico** de cada classe e cada relação da ontologia. Nesse caso, o que se quer dizer com a expressão “completude mínima necessária” de cada classe e relação da ontologia é que cada termo tenha as informações ontológicas mínimas e suficientes para representar na ontologia o objeto real a que ele se refere. Esse último aspecto sugerido, normalmente, não é tratado de forma explícita nas metodologias de construção pesquisadas, que, de uma maneira e outra, descrevem algumas informações ontológicas que devem estar presentes nas classes.

Na presente pesquisa, faz-se uma sugestão das informações ontológicas indispensáveis a toda classe e relação da ontologia. Tais informações foram baseadas nos

atributos de classe e de relação originados de ontologias, normalmente, bem aceitas na área, tais como a *Gene Ontology* (GO), a *Information Artifact Ontology* (IAO), a *Relation Ontology* (RO), e também da estrutura de linguagens de representação, tais como o *Resource Description Framework* (RDF) e o *RDF-Schema* (RDF-S). Tais atributos estão atualmente representados como campos informativos no editor de ontologias Protégé e o preenchimento de cada um desses atributos na ontologia em desenvolvimento assegura a completude mínima do conteúdo ontológico, conforme sugerido pela presente pesquisa. As informações ontológicas ou atributos sugeridos para classes de uma ontologia são apresentados na Tabela 16, a seguir.

Tabela 16 - Informações ontológicas indispensáveis às classes de uma ontologia

Informação ontológica	Origem	Atributo no Protégé	Significado (Descrição)
Identificador da classe	GO	<i>ID</i>	Identificador único da classe na ontologia.
Nome da classe	RDF-Schema	<i>label</i>	Nome preferencial do termo (classe).
Ontologia de origem	IAO	<i>imported from</i>	Ontologia da qual a classe incluída foi importada.
	GO	<i>hasOBONamespace</i>	Namespace de qual ontologia da OBO a classe foi importada.
URI	GO	<i>hasURI</i>	Identificador único da classe, importado da ontologia de origem desta classe.
Sinônimos	GO	<i>hasSynonym</i> <i>hasExactSynonym</i> <i>hasRelatedSynonym</i>	Outros nomes (termos) pelos quais o termo preferencial é referenciado no domínio tratado. São três tipos principais de sinônimos: (i) um sinônimo exato; (ii) um sinônimo relacionado; (iii) um sinônimo sem especificar seu tipo.
Termo alternativo	GO	<i>alternative term</i>	Nome do termo que não é necessariamente um sinônimo, mas em algumas ocasiões é usado para se referir ao termo, como suas flexões verbais ou de gênero (singular/plural).
Definição textual (em linguagem natural)	IAO	<i>definition</i>	Definição da classe descrita em linguagem natural.
Elucidação	IAO	<i>elucidaton</i>	Elucidação ou definição de uma classe que não possui instâncias ou particulares (classe abstrata), descrita em linguagem natural.
Exemplo de uso	IAO	<i>example of usage</i>	Instâncias ou particulares da classe que funcionam como exemplo de uso desta classe.
Definição formal (em linguagem lógica)	-	<i>SubClass Of</i>	Definição formal da classe especificada em lógica OWL e derivada de sua definição em linguagem natural.
Anotações (Comentários)	RDF-Schema	<i>comments</i>	Informações relevantes sobre a classe classificadas como não ontológicas e inseridas como comentários na ontologia.

Fonte: elaborado pelo autor.

O **nono aspecto** essencial é a **adoção de critérios, métodos ou técnicas para avaliação** do conteúdo da ontologia. Assim como ocorre com outros aspectos essenciais, geralmente, todas as metodologias de construção incluem uma etapa para a avaliação do conteúdo da ontologia. O que diferencia uma metodologia da outra quanto à avaliação é o nível de rigorosidade como é feita essa avaliação, que varia desde um

conjunto de critérios básicos até um framework formal de avaliação, e também os tipos de critérios, métodos ou técnicas empregados. O objetivo com esse aspecto é destacar a obrigatoriedade de se incluir uma etapa de avaliação da ontologia no processo de desenvolvimento afim de validar a ontologia de acordo com seus requisitos de especificação e verificar seu conteúdo do ponto de vista de consistência, integridade, precisão, entre outros parâmetros. Como esse aspecto é um assunto mais extenso e de extrema importância, criou-se um tópico nesta parte do referencial teórico da pesquisa (veja subseção 5.4) para tratar desse assunto e apresentar a abordagem de avaliação adotada.

Por fim, o décimo e último aspecto essencial corresponde à **documentação da ontologia ao final de cada etapa do desenvolvimento**. A Methontology (FERNÁNDEZ et al., 1997) foi a primeira metodologia de construção que deu ênfase à especificação de uma documentação em linguagem da natural ao final de cada fase do ciclo de vida do desenvolvimento da ontologia. Nesta metodologia, são previstos desde um documento de especificação de requisitos para a ontologia até um documento de sua implementação, ou seja, um para cada etapa da metodologia. Na presente pesquisa, apoiá-se essa abordagem da Methontology para documentação, entretanto, considera-se também que é bastante complicado na prática a elaboração de um documento formal por etapa. Nesse sentido, o objetivo com este último aspecto essencial é sugerir algum tipo de documentação ao final de cada etapa, que não necessariamente seja um documento formal. O importante nesse ponto é de alguma forma registrar o que está sendo feito a cada etapa para que na elaboração do documento formal final da ontologia todas as informações sobre seu conteúdo estejam contidas nele.

5.3.2 Caracterização das relações ontológicas

Além dos dez aspectos essenciais para a construção de ontologias, mencionados anteriormente, outro importante elemento de uma ontologia que necessita de uma atenção especial em sua criação e definição na ontologia são as relações ontológicas. Uma relação dentro de uma ontologia representa o relacionamento identificado na realidade do domínio entre dois ou mais conceitos, que são representados como classes na ontologia. Assim, a partir das relações entre as classes da ontologia é que o conhecimento do domínio vai sendo organizado e estruturado.

A maior dificuldade na especificação de uma relação na ontologia é identificar o tipo de relacionamento existente entre as classes envolvidas, uma vez que pode haver diferentes visões e interpretações do tipo mais apropriado para representar o relacionamento analisado. O objetivo desta seção é exatamente fornecer um aporte teórico,

com base na literatura da área, para tratar os diferentes tipos de relações ontológicas que podem existir no domínio e, assim, ajudar o desenvolvedor de ontologias na escolha das relações mais apropriadas para sua ontologia. O termo “caracterização” da relação significa descrever todas as propriedades de uma relação ontológica que as diferencie das demais e justifique o motivo de seu uso para representar um determinado relacionamento da realidade.

As primeiras relações ontológicas tiveram origem no quadrado e sexteto ontológico de Aristóteles, conforme apresentado na seção que descreve fundamentos ontológicos básicos. Essas relações compreendiam: *inherence*, *participation*, *instantiation*, *exemplification*, *characterization*. As ontologias atuais, como as de fundamentação, incorporam tais relações, porém caracterizam-as de maneira mais clara como, por exemplo, apresentando suas definições em uma linguagem da lógica formal.

Antes, porém, de apresentar as relações específicas, torna-se necessário caracterizar as relações ontológicas como um todo, ou seja, de maneira geral. Essa caracterização geral foi feita com base no realismo ontológico (GRENON, SMITH e GOLDBERG, 2004) (SMITH et al., 2005) (MUNN e SMITH, 2008). Os parágrafos seguintes tratam dessa caracterização.

A primeira característica abordada refere-se à definição da relação como genuinamente ontológica. De acordo com Munn e Smith (2008) para uma relação ser genuinamente ontológica é necessário que essa relação seja identificada entre entidades na realidade, independente da nossa experiência ou métodos de aprendizagem. Nesse sentido, apenas devem ser incluídas em uma ontologia relações dessa natureza.

A segunda característica que uma relação ontologia deve possuir diz respeito à universalidade da relação, isto é, relações ontológicas devem ser obtidas universalmente do domínio: uma declaração na forma *A relation B* deve ser obtida para todas as instâncias de *A* e não apenas, por exemplo, obtida de algumas instâncias representativas estatisticamente de *A* (MUNN e SMITH, 2008).

Em seguida, temos a caracterização geral da relação ontológica quanto à sua natureza e quanto às entidades relacionadas por ela, chamadas na literatura da área de *relata*. Quanto à *natureza da relação ontológica* temos *relações primitivas e derivadas*. Relações primitivas são relações indivisíveis a partir das quais é possível definir outras relações ontológicas chamadas de derivadas (GRENON, SMITH e GOLDBERG, 2004). Assim, por exemplo na BFO e na RO, as relações *instance_of* e *part_of* apenas entre relata particulares são consideradas primitivas, enquanto as relações *is_a* e *part_of* entre relata universais são consideradas derivadas. Uma observação importante é que relações primitivas devem ser auto-explicativas e neutras quanto a domínios do conhecimento.

Analisando se as *entidades relacionadas* (*relata*) pela relação são *universais ou particulares*, temos a seguinte caracterização (SMITH et al., 2005):

<universal, universal>: ambos os relata são universais (por exemplo, a relação de subsunção *is_a* – *whole portion of blood is_a portion of body substance*).

<instance, universal>: o primeiro relatum é um particular e o segundo é um universal (por exemplo, instanciação – *John's blood instance_of whole portion of blood*)

<instance, instance>: ambos os relata são particulares (por exemplo, a relação de participação – *John's blood participates_in John's blood transfusion*).

Outra caracterização sobre *entidades relacionadas* é se as entidades são *continuentes ou ocorrentes* na relação. Para cada relação é necessário definir o domínio e a imagem da relação de tal forma que se deve especificar se o domínio e a imagem da relação deve conter uma entidade continuante ou uma ocorrente ou ambas. Por exemplo, a relação *participates_in* sempre deve envolver como relata um continuante como domínio e um ocorrente como imagem. Essa caracterização é útil para o processo de avaliação do conteúdo ontológico.

Por fim, temos as propriedades lógicas básicas de uma relação, as quais podem ser usadas para identificar o tipo da relação a partir das propriedades. Essas propriedades básicas são conhecidas na literatura como meta-propriedades (propriedades de propriedades) e são (VARZI, 1996) (SMITH et al., 2005) (BITTNER e DONNELLY, 2007):

(i) *reflexividade*: uma relação é reflexiva se as entidades relacionadas são idênticas e a relação é válida entre elas.

Exemplo: *part_of* - cada objeto é parte de si mesmo (*x part_of x*), por exemplo, o corpo de John é a própria parte de seu corpo (*John's body part_of John's body*).

(ii) *transitividade*: uma relação é transitiva quando é possível inferir essa mesma relação em uma cadeia de entidades relacionadas.

Exemplo: *part_of* - se temos três objetos x, y e z tal que x *part_of* y e y *part_of* z então x é também *part_of* z. Assim, por exemplo, se seu ventrículo direito é parte do seu coração e se seu coração é parte de seu corpo humano, então seu ventrículo direito é parte de seu corpo humano, pela propriedade transitiva.

(iii) *simetria*: uma relação é simétrica quando ela é verdadeira entre as entidades relacionadas nos dois sentidos (direita-esquerda, esquerda-direita).

Exemplo: *part_of* é uma relação assimétrica: se temos dois objetos *x* e *y*, onde *x* *part_of* *y* então *y* não pode ser parte de *x* simultaneamente (*not y part_of x*). Assim, por exemplo, o tórax de João é uma subdivisão do seu tronco, mas o tronco de João não é uma subdivisão do seu tórax.

As caracterizações de uma relação ontológica apresentadas nos parágrafos anteriores fazem parte da caracterização geral. Além desta, é necessário abordar a caracterização específica das relações ontológicas, de forma a descrever propriedades específicas de cada tipo de relação que auxilie o desenvolvedor na modelagem do domínio do conhecimento. Nessa parte, optou-se por descrever unicamente os tipos ontológicos das relações *is_a* e *part_of* por dois motivos principais. Primeiramente, estas duas relações são as mais comuns e utilizadas na representação de domínios do conhecimento tanto em ontologias como em outros instrumentos de representação. Devido a isso, a maior quantidade de erros de modelagem em ontologias pode ser identificada a partir do uso dos tipos dessas duas relações, muitas vezes, devido ao desconhecimento por parte dos desenvolvedores dos tipos existentes e de seus significados. O segundo motivo pela escolha das relações *is_a* e *part_of* deve-se ao fato de que as demais relações ontológicas são bastante específicas e, em geral, suas definições nas ontologias de fundamentação são suficientes para explicar seus significados e como se dá seu uso nas ontologias. Na presente pesquisa, outras relações ontológicas além de *is_a* e *part_of* foram reaproveitadas das ontologias RO e BFO, e compreendidas a partir dessas duas ontologias de fundamentação, cuja descrição já foi apresentada anteriormente no referencial teórico.

Inicialmente, procurou-se definir as relações *is_a* e *part_of* de maneira geral e, posteriormente, suas especializações ou tipos, que devem ser inseridos como subclasses das duas relações gerais na taxonomia de relações da ontologia desenvolvida. Tendo como base as relações primitivas *instance_of* e *part_of*, é possível definir dois tipos de relação *is_a* e dois tipos de relação *part_of* a nível semi-formal (SMITH et al., 2005), conforme apresentado na Tabela 17, a seguir.

Tabela 17 - Definição semi-formal das relações *part_of* e *is_a* usadas na HEMONTO

Relação	Definição semi-formal	Exemplos de uso - HEMONTO
C <i>is_a</i> C1	$\forall c, \forall t$, se <i>c</i> <i>instance_of</i> C no tempo <i>t</i> , então <i>c</i> <i>instance_of</i> C1 no mesmo <i>t</i> .	Platelet concentrate <i>is_a</i> blood component.
P <i>is_a</i> P1	$\forall p$, se <i>p</i> <i>instance_of</i> P então <i>p</i> <i>instance_of</i> P1.	Centrifugation of blood <i>is_a</i> Process.
C <i>part_of</i> C1	$\forall c, \forall t$, se <i>c</i> <i>instance_of</i> C no tempo <i>t</i> , então há algum <i>c1</i> tal que <i>c1</i> <i>instance_of</i> C1 em <i>t</i> e <i>c</i> <i>part_of</i> <i>c1</i> em <i>t</i> .	Fibrinogen <i>part_of</i> Cryoprecipitate

P part_of P1	$\forall p$, se p instance_of P então há algum p1 tal que p1 instance_of P1 e p part_of p1.	Process of freeze part_of Process for obtaining of blood components.
---------------------	--	---

Fonte: elaborado pelo autor.

Na Tabela 17 é importante mencionar a notação formal adotada com base em Smith et al. (2005): (i) variáveis C, C1, Cn (letra maiúscula) representam *universais continuantes*; c, c1, cn (letra minúscula) representam *particulares continuantes*; (ii) variáveis P, P1, Pn (letra maiúscula) representam *universais ocorrentes*; p, p1, pn (letra minúscula) representam *particulares ocorrentes*; (iii) variáveis t, t1, tn são usadas para representar intervalos de tempo.

Em seguida, procurou-se clarificar os diferentes tipos de cada uma dessas relações apresentadas. Essa etapa é importante, pois na prática da conceitualização, especialistas e outros modeladores usam tais relações de forma ambígua e não consideram as diferenças entre seus tipos. Um exemplo dessa ambiguidade é a declaração lógica Actor \equiv Person \sqcap \exists interprets.Actuation \sqcap \exists is.Man (POVEDA-VILLALÓN et al., 2010).

Para a relação *is_a*, usada para construir taxonomias das ontologias, a identificação de seus tipos foi realizada com base em algumas referências da área ((MUNN e SMITH, 2008) (POVEDA-VILLALÓN et al., 2010)), as quais distinguem a relação *is_a* entre seu uso como instanciação, especialização ou especificação, e sinonímia ou equivalência, conforme descrito na Tabela 18.

Tabela 18 - Tipos ontológicos da relação *is_a*

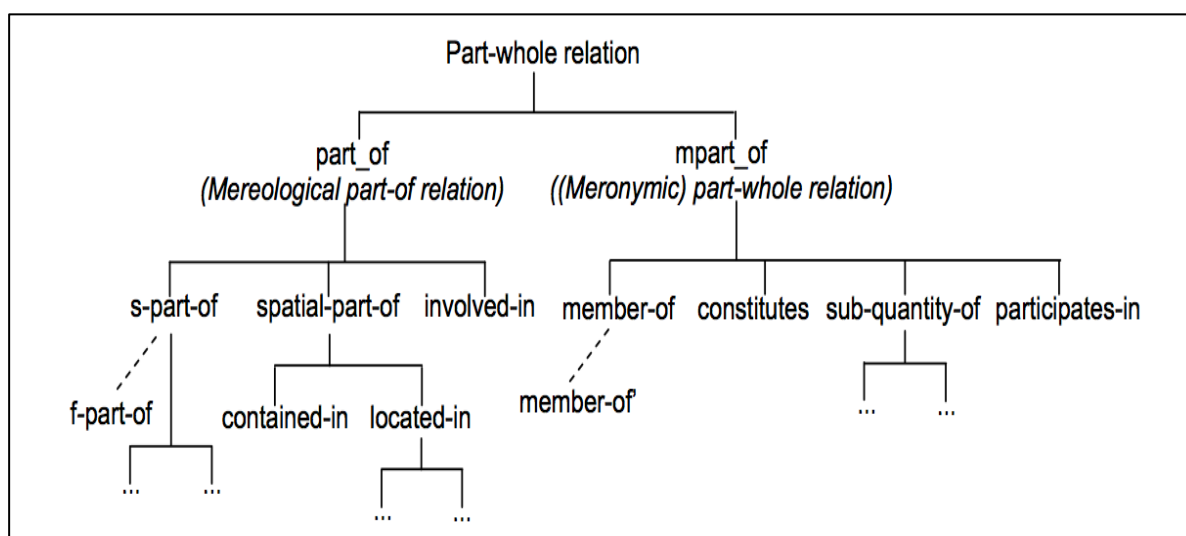
Relação	Definição informal	Exemplos
<i>is_a</i> usada como sinônimo de instance_of (Instanciação)	<i>Subsunção sob um gênero (espécie-gênero)</i> : representa a relação entre um particular (continuante ou ocorrente) e um universal (continuante ou ocorrente) nessa ordem.	John's blood is_a whole portion of blood John's apheresis is_a apheresis
<i>is_a</i> usada como sinônimo de subclass_of (Especificação ou Especialização)	<i>Especialização ou especificação (tipo-subtipo)</i> : representa a relação entre dois universais ou tipos: continuante-continuante ou ocorrente-ocorrente.	portion of plasma is_a portion of body substance apheresis is_a blood collection
<i>is_a</i> usada como sinônimo de same_as (Sinonímia ou Equivalência)	<i>Sinonímia, igualdade ou equivalência</i> : representa uma relação <i>is_a</i> tal que as duas entidades são idênticas. Pode ocorrer entre particulares continuantes ou ocorrentes e também entre universais continuantes ou ocorrentes.	erythrocytes is_a red blood cells leukocytes is_a white blood cells

Fonte: elaborado pelo autor.

Quanto à relação *part_of*, podemos identificar diferentes tipos para ela a partir da taxonomia de relações part-of apresentada em Keet e Artale (2008), a qual engloba

uma série de tipos *part_of* recuperados da literatura da área ((VARZI, 1996) (BITTNER e DONNELLY (2007) WINSTON et al. (1987) GERSTL e PRIBBENOW (1995)), do ponto de vista mereológico e também meronímico, conforme apresentado na Figura 31 e na Tabela 19, a seguir. A Figura 31 traz a taxonomia de relações *part-of* e nela as linhas pontilhadas indicam que o subtipo possui restrições adicionais na participação dos tipos de entidades. A Tabela 19 apresenta as definições formais de cada relação e alguns exemplos. Em tais definições, a sigla POB representa *physical objects* e SOB representa *non-physical social objects* (uma simplificação do trabalho de MASOLO et al. (2004)). O nome da relação na Tabela 19 é apresentado como na taxonomia *part-of* proposta e os sinônimos da relação, encontrados na literatura, estão entre parênteses.

Figura 31 - Taxonomia de relações *part-of* nas visões mereológica e meronímica



Fonte: elaborado pelo autor.

Tabela 19 - Tipos ontológicos da relação *part-of*

Relação	Definição Formal	Exemplo
s-part-of (structural parthood)	$\forall x, y (s_part_of(x, y) \equiv part_of(x, y) \wedge ED(x) \wedge ED(y))$, onde ED é um endurente ou continuante.	car carcass s-part-of car
f-part-of (functional part)	$\forall x, y (s_part_of(x, y) \equiv part_of(x, y) \wedge ED(x) \wedge ED(y))$, function (x, X), function (y, Y), onde ED é um endurente ou continuante e function indica que x é uma função de X.	car engine f-part-of car
contained-in	$\forall x, y (contained_in(x, y) \equiv part_of(x, y) \wedge R(x) \wedge R(y) \wedge \exists z, w (has_3D(z, x) \wedge has_3D(w, y) \wedge ED(z) \wedge ED(w)))$, onde ED é um endurente ou continuante e R é a região especial que ele ocupa. 2D e 3D são usadas para diferenciar o tipo de região especial ocupada.	John's address book contained_in John's bag; Actin filament contained_in cell
located-in	$\forall x, y (located_in(x, y) \equiv part_of(x, y) \wedge R(x))$	Porto located_in North Portugal;

	$\wedge R(y) \wedge \exists z, w(\text{has_2D}(z, x) \wedge \text{has_2D}(w, y) \wedge \text{ED}(z) \wedge \text{ED}(w))$, onde ED é um endurante ou continuante e R é a região espacial que ele ocupa. 2D e 3D são usadas para diferenciar o tipo de região espacial ocupada.	Human brain located_in human head
involved-in (argument restriction Ad2 – DOLCE; or temporal-part-of)	$\forall x, y(\text{involved_in}(x, y) \equiv \text{part_of}(x, y) \wedge \text{PD}(x) \wedge \text{PD}(y))$ onde PD é um perdurante ou ocorrente.	Chewing involved-in eating
member-of (member-bunch, collection, collective or aggregation)	$\forall x, y(\text{member_of}_\eta(x, y) \equiv \text{mpart_of}(x, y) \wedge (\text{POB}(x) \vee \text{SOB}(x)) \wedge \text{SOB}(y))$, onde POB representa physical objects e SOB representa non-physical social objects.	Herd member-of Sheep; Musician member-of Orchestra; Husband and Wife member-of Marriage.
constitutes (material-object relation or constitution)	$\forall x, y(\text{constitutes}_{it}(x, y) \equiv \text{constituted_of}_{it}(y, x) \equiv \text{mpart_of}(x, y) \wedge \text{POB}(y) \wedge \text{M}(x))$, onde M é an amount of matter e POB representa physical objects.	Marble constitutes Statue
sub-quantity-of (quantity-mass or portion-object)	$\forall x, y(\text{sub_quantity_of}_\eta(x, y) \equiv \text{mpart_of}(x, y) \wedge \text{M}(x) \wedge \text{M}(y))$, onde M é an amount of matter.	salt sub-quantity-of seawater; alcohol sub-quantity-of wine; milliliter sub-quantity-of liter
participates-in (participation or noun-feature /activity)	$\forall x, y(\text{participates_in}_{it}(x, y) \equiv \text{mpart_of}(x, y) \wedge \text{ED}(x) \wedge \text{PD}(y))$ onde ED é um endurante ou continuante e PD é um perdurante ou ocorrente.	enzyme participates_in catalytic-reaction; plasma-extractors participates-in process of separation of the plasma

Fonte: elaborado pelo autor.

A taxonomia proposta por Keet e Artale (2008) é bastante completa em relação aos tipos existentes de part-of, entretanto, é necessário complementar tal abordagem para tratar casos da relação part-of que envolvam temporalidade e localização espacial simultaneamente, além de relações entre entidades materiais e imateriais. Esses casos específicos são de extrema importância para a área médica, como pode ser visto pelos exemplos a seguir (SCHULZ et al., 2006):

(i) É possível a existência de relações parte-todo entre entidades materiais e entidades imateriais?

Exemplos: A superfície do meu braço é uma parte do meu braço? A cavidade do meu estômago é parte do meu estômago?

(ii) Partes removidas do todo ainda podem ser consideradas partes dele?

Exemplos: O apêndice cirurgicamente removido de um corpo ainda é parte deste corpo? Uma amostra de tecido em análise no laboratório ainda é parte do organismo do qual ele foi extraído?

(iii) Como tratar a relação parte-todo em relação à localização espacial?

Exemplos: Órgãos transplantados são parte do organismo do doador, do receptor ou de ambos? Oligosacarídeos e lipídios absorvidos pela mucosa intestinal são parte do intestino?

Uma solução para estes casos apontada em Schulz et al. (2006) é introduzir tipos de relações **part-of** que tratem temporalidade, localização espacial e entidades imateriais. Algumas destas relações são apresentadas na Tabela 20, a seguir.

Tabela 20 - Outros tipos de relações **part-of**: espaciais, temporais e imateriais

Relation	Definição informal	Exemplos
Temporary-Part-Of (A, B, t)	Duas classes A e B estão relacionadas pela temporary-part-of se e somente se cada instância de A é parte de B em pelo menos algum instante de tempo.	1) Uma porção de sangue é parte um organismo no instante t_1 (quando ele ainda está no organismo), mas não é uma parte desse organismo no instante t_2 quando essa porção está em algum laboratório. 2) Amputated toe Temporary-Part-Of Body Human
Permanent-Part-Of (A, B, t)	A é uma parte permanente de B, quando A é part_of de B em todos os instantes do ciclo de vida de A. Neste caso A está localizado em B (A located_in B) e não há nenhum momento ou instante de tempo que A não está localizado em B.	1) Human Female Reproductive System Permanent-Part-Of Human Body. 2) My brain Permanent-Part-Of my head 3) A give cell membrane Permanent-Part-Of the cell it surrounds.
Immaterial parts (fiat-part-of)	Partes imateriais do todo como os limites fiat e bona fide, além de buracos (cavidades, orifícios, ...) são consideradas parte da entidade material corpo humano, porque elas não correspondem a discontinuidades da realidade, sendo partes contínuas do todo.	A superfície interna do meu estômago é parte do meu estômago bem como a cavidade do estômago.

Fonte: elaborado pelo autor.

5.4 Critérios de avaliação de conteúdos ontológicos

Nas subseções anteriores foram abordados aspectos importantes no processo de construção de ontologias, já nesta parte da tese o assunto refere-se às formas de avaliação do conteúdo ontológico desenvolvido, focando especialmente no conjunto de critérios de avaliação da **NeOn Methodology** (SUÁREZ-FIGUEROA et al., 2008) (POVEDA-VILLALÓN et al., 2010) (POVEDA-VILLALÓN et al., 2012) e também em algumas propriedades da metodologia **OntoClean** (GUARINO e WELTY, 2004).

Avaliação ontológica é uma técnica de julgamento do conteúdo da ontologia quanto a sua adequação à especificação de requisitos, questões de competência e compromisso com o mundo real, durante cada fase de seu desenvolvimento e entre as fases de seu ciclo de vida (GÓMEZ-PEREZ, 2004). A avaliação ontológica inclui tanto a

verificação quanto a *validação* ontológicas, que podem ser assim definidas (GÓMEZ-PÉREZ, 2004) (SUÁREZ-FIGUEROA et al., 2009):

- *Validação ontológica* é a atividade que compara o significado das definições ontológicas (o modelo ontológico desenvolvido) com o modelo pretendido do domínio que está sendo conceitualizado.
A validação responde à questão: você está desenvolvendo a ontologia correta?
- *Verificação ontológica* é a atividade de verificar a corretude da ontologia, isto é, se ela foi ou está sendo desenvolvida corretamente: se suas definições estão implementadas de acordo com a especificação de requisitos ontológicos e com as questões de competência, além de seu conteúdo estar funcionalmente correto.
A verificação responde à questão: você está construindo a ontologia de maneira correta?

Os primeiros trabalhos sobre avaliação de ontologias surgiram no ano de 1994 ainda no contexto da Ciência da Computação (GÓMEZ-PÉREZ, 1994). Anos depois, novas pesquisas desta autora (GÓMEZ-PÉREZ et al., 2003) (GÓMEZ-PÉREZ, 2004) (GÓMEZ-PÉREZ e LOZANO-TELLO, 2005) e uma conhecida metodologia de avaliação denominada de *OntoClean* (GUARINO e WELTY, 2002) foram os trabalhos mais citados na literatura da área e relevantes sobre avaliação ontológica.

Nos trabalhos de Gómez-Pérez, a avaliação ontológica é realizada com base em uma perspectiva de um dos paradigmas de representação do conhecimento, relativo ao *conteúdo taxonômico*. Na avaliação das relações taxonômicas por estes trabalhos, os critérios utilizados são: *consistência, completude, concisão, expandibilidade e sensibilidade*. Nessa abordagem, não é incluída a avaliação de outros tipos ontológicos: relações, propriedades ou atributos e axiomas; os quais demandam o uso de técnicas de lógica de primeira ordem para avaliação.

A metodologia ***OntoClean*** foi elaborada pelo Ontology Group de LADSEB-CNR de Pádova, Itália, e tem como propósito fazer uma limpeza ou correção nas *taxonomias conceituais* da ontologia avaliada, de forma a remover subclasses incorretas nelas contidas (GUARINO e WELTY, 2004). Para cumprir com este propósito, tal metodologia baseia-se em noções ontológicas altamente gerais originadas de uma abordagem filosófica conhecida como *metafísica analítica* (GUARINO e WELTY, 2002). A partir dessa

abordagem, a metodologia consegue identificar escolhas de modelagem inapropriadas e inconsistentes nas taxonomias construídas.

A identificação dessas inconsistências e relacionamentos inapropriados é feita com base na adoção de meta-propriedades ontológicas, as quais são propriedades de propriedades e permitem impor restrições na construção taxonômica de uma ontologia. A estratégia usada é a de fazer uma avaliação *a posteriori* das relações construídas, analisando as entidades (classes) ligadas por estas relações a partir de suas propriedades. As principais meta-propriedades utilizadas pela *OntoClean* nessa avaliação são: (i) *essência*; (ii) *rigidez*; (iii) *identidade*; e (iv) *unidade*; e podem ser assim descritas conforme Guarino e Welty (2002); Guarino e Welty (2004):

- **Essência:** uma propriedade é considerada essencial para uma entidade quando ela se mantém verdadeira por toda existência desta entidade, em todos os mundos possíveis. Assim, por exemplo, “ter um cérebro” é uma propriedade essencial para todos seres humanos
- **Rigidez:** é um tipo especial da propriedade de essência. Uma propriedade é rígida quando ela é essencial para todas as instâncias possíveis da entidade que a possui. Para esta propriedade temos três tipos de classificações: (i) *rígido*: se a propriedade é essencial para todas as instâncias. Exemplo: “ser pessoa”. Todas as instâncias de “ser pessoa” em todos os mundos possíveis mantém sua existência de pessoa ao longo do tempo. (ii) *semi-rígido*: se a propriedade é essencial para algumas instâncias, mas não para outras. Exemplo: “ser duro” é uma propriedade essencial para martelos, por exemplo, mas não o são para as esponjas, porque em geral esponjas não são duras, porém pode existir alguma que seja. (iii) *anti-rígido*: se a propriedade nunca é essencial para suas instâncias. Exemplos: ser estudante, comida. Ser estudante é um estado temporário, tal que ao longo da vida a pessoa pode deixar de ser estudante. A comida pode deixar de ser uma comida, por exemplo, quando é digerida.
- **Identidade:** refere-se ao problema de ser capaz de reconhecer entidades no mundo que são a mesma entidade ou entidades diferentes. Duas entidades são iguais se e somente se elas tem exatamente as mesmas partes. Para esta propriedade temos duas classificações: (i) *identidade*: duração do tempo = [uma hora; duas horas]. Todas as durações de tempo são idênticas, uma hora é sempre a mesma uma hora; (ii) *não-identidade*: intervalo de tempo = [1:00-2:00

Terça; 1:00-2:00 Sexta]. Dois intervalos deveriam ser idênticos, mais se ocorrem em tempos diferentes são intervalos diferentes.

- **Unidade:** refere-se à capacidade de reconhecer todas as partes que formam ou compõem uma entidade individual. Neste caso, analisa-se se as instâncias de uma dada entidade podem ser uma entidade única e representar a entidade como um todo. Para esta propriedade temos três tipos de classificações: (i) *unidade*: instâncias que podem vir isoladas e representam o todo (classe). Exemplo: *o Oceano Atlântico*, representa o todo da entidade oceano Atlântico; (ii) *anti-unidade*: instâncias que quando aparecem isoladas não representam o todo. Exemplo: *porção de água*. Uma porção de água por maior ou representativa que seja não consegue representar o todo da entidade água. (iii) *não-unidade*: instâncias que compõem um todo, mas são diferentes entre si. Exemplo: *pessoa jurídica* formada pelas entidades *pessoa única*, *empresa*. Neste caso, pessoa jurídica pode representar o todo desta entidade, mas é composta de instâncias de diferente natureza ontológica.

A importância e utilidade de tais propriedades é que a partir delas é possível checar a consistência ontológica das relações taxonômicas. Assim, por exemplo, ao analisar a subsunção *estudante subclassOf pessoa* sob a propriedade de rigidez é possível perceber que há uma clara inconsistência ao interligar uma entidade que é não-rígida (*estudante*) com uma rígida (*pessoa*). De maneira análoga, ao supor *intervalo_de_tempo subclassOf duração_de_tempo* de novo há inconsistência pelo fato de que *duração_de_tempo* é identidade e *intervalo_de_tempo* não identidade sob o critério de identidade. Por fim, sob o critério de unidade verifica-se inconsistência ao analisar *oceano subclassOf porção_de_água*, porque *oceano* é unidade e *porção_de_água* é não-unidade.

Já a avaliação ontológica proposta na **NeOn Methodology** é baseada em uma abordagem conhecida como *descrições ontológicas orientadas a qualidade (Quality-oriented Ontology Descriptions)*, que englobam: (i) uma meta-ontologia O^2 , a qual descreve ontologias como objetos semióticos e descreve também características e elementos da ontologia que podem ser avaliados; (ii) modelos oQual de avaliação ontológica, que são tarefas diagnósticas baseadas na descrição formal das ontologias em O^2 .

Com base na abordagem mencionada, a NeOn propôs um framework formal de avaliação ontológica que engloba três tipos principais de categorias de avaliação (SUÁREZ-FIGUEROA et al., 2009):

- *Avaliações estruturais*: considera a estrutura lógica da ontologia, tal como em um grafo conceitual. Nessa avaliação a ontologia é vista simplesmente como um objeto de informação e, nesse sentido, é avaliada apenas nos níveis sintático e semântico, livre do contexto. Medidas estruturais incluem: profundidade, largura, densidade, adequação lógica, distribuição de grau, etc.
- *Avaliações funcionais*: neste caso avalia-se a ontologia no contexto de seu uso, tal que a conceitualização pretendida é comparada com os requisitos esperados e objetivos alcançados com a ontologia. Medidas funcionais incluem: aceitação dos especialistas, satisfação dos usuários, avaliação de tarefas, avaliação de tópicos.
- *Avaliação de usabilidade*: o foco neste caso é o nível pragmático da ontologia e o contexto no qual ela é comunicada. Neste nível, a ontologia é vista como um objeto semiótico. Os níveis de usabilidade incluem: reconhecimento/identificação da ontologia por anotações, eficiência quanto às necessidades do usuário, nível de interface com o usuário.

Além desses diferentes níveis de avaliação ontológico é também importante considerar que ao lidar com redes ontológicas e não só com ontologias isoladas, a metodologia *NeOn* faz sua avaliação em duas etapas: (i) primeiramente, a avaliação das ontologias constituintes da rede isoladamente; e (ii) em segundo lugar, a avaliação dos mapeamentos e alinhamentos entre pares de ontologias pertencentes à rede. Com essa abordagem, é possível obter uma avaliação da rede ontológica por completo.

Na presente pesquisa, o foco concentra-se na avaliação da ontologia isoladamente, não considerando o contexto de uma rede ontológica. Além disso, estuda-se os tipos de erros cometidos no desenvolvimento ontológico conhecidos na literatura como *pitfalls*. Segundo Poveda-Villalón et al. (2010) (2012), erros comuns no desenvolvimento ontológico podem ser agrupados em dois tipos: (i) erros relacionados aos padrões de desenvolvimento ontológico - *Ontology Design Patterns*⁴³ (ODPs), chamados de *anti-padrão*, que podem ser de dois subtipos: (i.i) casos em que se realizaram *matches* com um ODP, porém a solução não é adequada para o problema de modelagem; (i.ii) casos em que o match com um dado ODP é apropriado, mas o ODP não foi utilizado/aplicado; e (ii) erros não relacionados aos ODPs, chamados de *pitfalls*: são casos onde foi observada

⁴³ Na área de engenharia ontológica são comuns soluções conhecidas como *Ontology Design Patterns* (ODPs), as quais definem soluções para os problemas de desenvolvimento ontológico, auxiliando os desenvolvedores da área na atividade de modelagem ao prover diretrizes de desenvolvimento e melhorar a qualidade das ontologias resultantes (SUÁREZ-FIGUEROA, 2010).

uma solução não adequada ao problema de modelagem e não há um ODP apropriado para tal.

Para o tratamento dos erros conhecidos como *pitfalls*, Poveda-Villalón et al. (2010) (2012) proveêm uma solução alternativa à utilização de ODPs, ao propor um conjunto de guias metodológicos afim de evitar esses erros comuns no desenvolvimento ontológico. A ideia baseia-se no argumento de que mesmo com o auxílio dos atuais ODPs, os desenvolvedores de ontologias enfrentam dificuldades ao reutilizá-los, durante a modelagem ontológica, e continuam a cometer erros nesse processo.

Na primeira versão (POVEDA-VILLALÓN et al., 2010), foram descritos 24 tipos comuns de *pitfalls* e, em uma versão seguinte (POVEDA-VILLALÓN et al., 2012), inclui-se mais 5 tipos de *pitfalls*, totalizando um total de 29 erros agrupados de acordo com a propriedade ontológica que eles violam: (i) *integridade* (exemplo: *pitfall* 11 – ausência de domínio ou intervalo de valores nas propriedades); (ii) *precisão* (exemplo: *pitfall* 17 – existência de muitas especializações hierárquicas); (iii) *consistência* (exemplo: *pitfall* 5 – definições de relações inversas erradas); e (iv) *documentação* (exemplo: *pitfall* 22 – utilização de diferentes critérios de nomeação na ontologia). No presente trabalho, convencionou-se a chamar os tipos comuns de *pitfalls* como critérios de avaliação ontológica da NeOn. Assim, daqui para frente iremos usar tal denominação.

Uma das estratégias adotadas, na presente pesquisa, é reaproveitar alguns desses critérios de avaliação e incluí-los em passos da metodologia de construção ontológica proposta, de forma a auxiliar o desenvolvedor a evitar os erros comuns no processo de desenvolvimento ontológico. Nesse sentido, torna-se fundamental um olhar mais detalhado para tais critérios de forma a incorporá-los na metodologia proposta. Assim, construiu-se uma tabela resumida (veja Tabela 21, a seguir) contendo o tipo de *pitfall* ou critério, sua breve descrição e um exemplo deste tipo de erro. Seguindo a abordagem de Poveda-Villalón (2010), a tabela foi construída separando tais critérios de acordo com o tipo de violação a que eles que correspondem e numerando-os de acordo com a abordagem seguida.

Tabela 21 - Tipos de *pitfalls* encontrados na construção de ontologias

Tipo de violação: Inconsistência			
Cód.	Pitfall	Descrição	Exemplo do erro
1	Criação de elementos polissêmicos.	É criado um termo na ontologia com o mesmo nome e que tem mais de um significado, isto é, representa mais de uma ideia conceitual (conceito).	O termo “Célula” é usado tanto com o significado de “componente biológico do corpo humano” quanto como “local ou lugar de trabalho de um grupo de operários”.
5	Definição de relações inversas erradas.	Duas relações são definidas como inversas na ontologia quando na verdade não são.	“isSoldIn” e “isBoughtIn” são definidas como inversas e quando usada em, por exemplo: “object1

			isSoldIn place3", o classificador irá inferir incorretamente que "place3 isBoughtIn object1".
6	Inclusão de ciclos na hierarquia.	Um ciclo é criado na hierarquia de classes de uma ontologia quando uma classe A tem B como subclasse e ao mesmo tempo B é também superclasse de A.	"Professor" e "Pessoa" são classes de um ontologia, tal que "Professor" é subclasse de "Pessoa" e ao mesmo tempo "Pessoa" é subclasse de "Professor".
7	Mesclar conceitos diferentes na mesma classe.	É criada uma classe na ontologia tal que seu identificador refere-se a dois ou mais conceitos diferentes.	Classe "ProdutoOuServiço" Classe "Estilo&Período"
14	O mau uso da propriedade "allValuesFrom".	Esse mau uso ocorre de duas maneiras de diferentes: (1) o uso da restrição universal ("allValuesFrom") como qualificador padrão ao invés do uso da restrição existencial ("someValuesFrom") e (2) o uso de "allValuesFrom" fechando ou encerrando as possibilidades de adição da propriedade dada.	(1) Auto-explicativo (2) Ex: Book \equiv \exists producedBy.Writer \square \forall uses.Paper', fechando a possibilidade de uso de "Ink" (tinta) na escrita do livro.
15	O mau uso das propriedades "not some" e "some not".	Verifica-se na definição de propriedades de uma classe, quando é utilizado a restrição "not some" onde deveria se usar "some not" e vice-versa.	Na ontologia de pizzas, esse erro ocorre quando se define a pizza vegetariana como qualquer pizza que tem tanto algum (some) topo que não é carne e também algum (some) topo que não é peixe. (vegetarian pizza \equiv any pizza some topping and (topping is not meat and some not fish))
18	Existência de muitas especificações de domínio ou intervalo de valores.	Ocorre quando não se encontra um domínio ou imagem que seja suficientemente geral.	Ao definir o domínio da relação "isOfficialLanguage" apenas à classe "City" estará se restringindo o domínio, que também poderia ser as classes "Country" ou "GeopoliticalObject".
19	Permuta de interseção e união.	Ocorre quando a imagem ou domínio de uma dada relação é definida como a interseção ou união de duas ou mais classes.	Na criação da relação "takesPlaceIn" definiu-se como domínio "OlympicGames" e como imagem a interseção das classes "City" e "Nation".
24	Utilização de definição recursiva.	Quando um termo da ontologia é também usado em sua própria definição, temos uma definição recursiva.	Na criação da relação "hasFork" define-se como imagem a condição "o conjunto de restaurantes que tem pelo menos um relacionamento "hasFork" ".
25	Definição de uma relação inversa a si mesmo.	Verifica-se quando uma mesma relação é a própria relação e ao mesmo a relação inversa.	Própria relação: part_of Relação inversa: proper_part_of
26	Definição de relações inversas para um simétrico.	Ocorre quando são definidas relações inversas para um elemento da ontologia que é simétrico.	Não encontrado.
27	Definição de relações equivalentes erradas	Verifica-se quando duas relações que não são equivalentes no domínio são representadas como equivalentes na ontologia.	Um exemplo comum é tratar as relações "contained_in" e "located_in" como equivalentes, quando na verdade são tipos de relações diferentes.
28	Definição de relações simétricas erradas	Ocorre quando uma relação que é assimétrica no domínio é representada como simétrica na ontologia.	Part_of foi definida em uma ontologia do corpo humano como simétrica. Assim, por exemplo, o tórax de João é uma subdivisão (parte) do seu tronco e o tronco de João é uma subdivisão (parte) do seu tórax, pela propriedade da simetria, onde não verdade não

			deveria ser.
29	Definição de relacionamentos transitivos errados	Ocorre quando uma relação que é intransitiva no domínio é representada como transitiva na ontologia.	A relação <code>is_father_of</code> foi definida como transitiva em uma ontologia. Assim, por exemplo, se <code>John is_father_of Paul and Paul is_father_of Smith</code> , pela propriedade transitiva, <code>John is_father_of Smith</code> – o que é incorreto já que John é na verdade avô de Smith.
Tipo de violação: Imprecisão			
Cód.	Pitfall	Descrição	Exemplo do erro
2	Criação de sinônimos como classes.	Ocorre quando se criam classes diferentes, com identificadores diferentes, para um mesmo conceito do mundo real. E ainda tais classes são definidas como equivalentes na ontologia.	“Car”, “Motorcar”, “Automobile” são criadas como classes da ontologia e como equivalentes.
3	Criação de relacionamento “is_a” em vez de usar “subclassOf”, “instanceOf” ou “sameIndividual”.	Neste caso, a relação “is_a” é usada na ontologia sem distinguir os diferentes tipos de relação “is_a” que existem na realidade, do ponto de vista ontológico, tal como as primitivas OWL “subclassOf”, “instanceOf” ou “sameIndividual”.	subclassOf: Mammal is_a human being instanceOf: Aristoteles is_a human being sameIndividual: Aristoteles is_a Greek philosopher that created ontological square
17	Existência de muitas especializações hierárquicas.	Consiste na criação de instâncias (particulares) como classes, tal que a hierarquia se torna tão especializada que as classes que são nós-folha não tem instâncias - elas são as próprias instâncias.	(superclasse) Place (subclasses) Madrid, Sevilla, Barcelona (superclasse) RatingOfRestaurants (subclasses) 1fork, 2forks
21	Utilização de classe variada ou genérica	Ocorre quando adiciona-se na hierarquia de classes da ontologia uma classe que mistura dois ou mais conceitos diferentes do domínio tratados como uma classe variada ou genérica. Os labels dessa classe são normalmente “Outras” ou “Miscelânea”.	(superclasse) HydrographicalResource (subclasses) Stream Waterfall OtherRiverElement
Tipo de violação: Não integridade			
Cód.	Pitfall	Descrição	Exemplo do erro
4	Criação de elementos da ontologia desconexos	Ocorre quando se criam classes, relações ou atributos que não possuem qualquer relação com o restante (conteúdo) da ontologia.	Criar a relação “memberOfTeam” e não existir na ontologia classes que representam “teams”.
9	Ausência de informações básicas	Verifica-se quando alguma informação necessária não foi incluída na ontologia. Pode ser um erro relacionado aos requisitos dos ODPs ou a ausência de uma informação que tornaria a ontologia mais completa.	Ex1: Criar a relação “startsIn” para representar rotas que começam em um dado lugar e não criar a relação “endsIn” informando o lugar que a rota se encerra. Ex2: Criar a relação “follows” e não criar sua relação inversa “precedes”.
10	Ausência de disjunção	Ocorre quando não são definidos axiomas de disjunção entre classes ou entre propriedades.	Ex1: Classes “Odd” e “Even” criadas na ontologia, mas não definidas como disjuntas. Ex2: Classes “Prime” e “Composite” criadas na ontologia, mas não definidas como disjuntas.
11	Ausência de domínio ou intervalo de valores	Erro verificado quando classes ou propriedades são incluídas na ontologia	Criar a relação “hasWritten” sem definir como domínio, por

	nas propriedades	sem a definição de um conjunto domínio ou imagem ou ambos.	exemplo, "Writer" e como imagem "LiteraryWork".
12	Ausência de propriedades equivalentes	Verifica-se quando se esquece de definir como equivalentes classes ou relações que possuem essa propriedade. Esse é um erro comum após a importação de uma ontologia ou merge entre ontologias.	Ex1: "CITY" e "city" criadas como classes não equivalentes. Ex2: "hasMember" e "hasMember" criadas como relações não equivalentes.
13	Ausência de relações inversas	Ocorre quando cria-se duas relações que são inversas entre si mas não foram definidas como tal na ontologia.	"isLocatedAt" e "isLocationOf" foram criadas na ontologia mas não definidas como relações inversas. Assim se temos, por exemplo, "building1 isLocatedAt site2" não se pode inferir que "site2 isLocationOf building1".
16	O mau uso de classes primitivas e definidas	Neste caso, a ontologia falha ao deixar de fazer a definição "completa" em vez de "parcial". É fundamental entender que, em geral, nada vai ser inferido pelo classificador quando a classe é definida como primitiva.	Não encontrado
Tipo de violação: Não documentação			
Cód.	Pitfall	Descrição	Exemplo do erro
8	Ausência ou falta de anotações.	Verifica-se quando não são inseridas propriedades de anotação para classes ou relações.	Auto-explicativo
20	Permuta de rótulos (labels) e comentários.	Ocorre quando os conteúdos do label e as anotações de comentários são trocados.	Em uma ontologia sobre geografia, a classe "Crossroads" foi anotada com o label "the place of intersection of two or more roads" e na anotação de comentário está a palavra "Crossroads". O correto seria o contrário e a anotação do label viria no lugar da anotação de comentário.
22	Utilização de diferentes critérios de nomeação na ontologia	Ocorre quando não se utiliza qualquer convenção ou padrão de nomeação para os identificadores dos elementos da ontologia.	Algumas relações utilizam o símbolo "-" para separar nomes composto (exs: located-in, composed-by) e outras não (exs: derivesFrom, collectedOf).
23	Utilização de elementos incorretamente	Uma classe, relação ou atributo da ontologia é usado para modelar uma parte da ontologia onde deveria ser usado um elemento diferente da mesma.	Car isEcological Yes or Car isEcological No. Neste caso, o apropriado seria definir a imagem da relação "isEcological" como tipo Booleano (Yes, No).

Fonte: adaptado de Poveda-Villalón (2010) (2012)

5.5 Implementação de ontologias por linguagens lógicas

Nas seções anteriores do presente capítulo, discorreu-se brevemente sobre as metodologias para construção de ontologias, aspectos importantes nesse processo e critérios de avaliação do conteúdo ontológico. Nesta seção, aborda-se outro assunto de extrema importância para os resultados finais de um modelo ontológico: a representação da ontologia em uma linguagem formal. Cabe ressaltar que o propósito desta seção não é esgotar completamente o assunto e nem abordar toda a amplitude do tema, mas sim

discorrer sobre as potencialidades e também as restrições das linguagens lógicas para a representação ontológica de um dado domínio, especificamente da lógica usada na linguagem OWL, a qual é usada na representação da teoria ontológica do sangue humano.

A *Ontology Web Language (OWL)* foi desenvolvida e padronizada pelo *World Wide Consortium (W3C)* como uma linguagem padrão para a representação e integração de ontologias no ambiente web. Segundo McGuinness e van Harmelen (2004), a OWL surgiu da necessidade de se criar uma linguagem formal para representação ontológica na web, a chamada Web Semântica, e estender os recursos de linguagem até então disponíveis para esta tarefa: o XML, o XML-Schema, o RDF e o RDF-Schema. Nesse sentido, a linguagem OWL adiciona mais vocabulário ou recursos para a descrição das classes e propriedades de uma ontologia, tais como: relacionamentos entre as classes, cardinalidade, tipos de propriedades, características de propriedades (por exemplo: simetria) e classes enumeradas.

McGuinness e van Harmelen (2004) destacam também que com o crescimento do uso da linguagem OWL por comunidades específicas de usuários e desenvolvedores, a OWL foi desmembrada em três versões diferentes de forma a atender os requisitos necessários de cada comunidade. Assim as três atuais sublinguagens da OWL são (MCGUINNES e VAN HARMELEN, 2004):

- *OWL Lite*: é a sublinguagem OWL de menor complexidade formal. Ela dá suporte aos usuários que necessitam apenas de uma hierarquia de classificação, como, por exemplo, uma taxonomia ou um tesouro, e restrições simples, como, por exemplo, restrições de cardinalidade cujos únicos valores possíveis são 0 (zero) e 1 (um).
- *OWL Description Logic (OWL-DL)*: sublinguagem OWL que é equivalente à Lógica Descritiva. Ela dá suporte aos usuários que desejam expressividade lógica em suas construções, limitadas pelo poder de decidibilidade inerente à linguagem OWL. A OWL-DL inclui todas as construções da linguagem OWL, mas elas só podem ser usadas sob certas restrições, como, por exemplo, uma classe pode ser uma subclasse de duas ou mais classes, ou então, uma classe não pode ser instância de outra classe.
- *OWL Full*: é a sublinguagem OWL de maior complexidade formal, dedicada aos usuários que esperam o máximo de expressividade da linguagem e liberdade sintática das garantias computacionais de uma descrição RDF, na qual a OWL foi baseada. Assim, por exemplo, em OWL Full uma classe pode ser tratada

simultaneamente como uma coleção de indivíduos (ou instâncias) e também como um indivíduo (ou instância) em si.

Dentre as versões citadas, o foco desta pesquisa é a sublinguagem OWL-DL, uma vez que tal versão da OWL é a mais usada na representação de ontologias biomédicas (SMITH et al., 2007) e, por isso, foi escolhida para representar a ontologia sobre o sangue humano. Além de tal linguagem, é importante ressaltar que, quando se fez necessário, foi também usada a **lógica descritiva pura** para representar parte do conteúdo da ontologia. Na prática, a OWL-DL é um subtipo da lógica descritiva e ambas são amplamente utilizadas na representação de ontologias e terminologias biomédicas.

De acordo com Schulz et al. (2009), desde o esforço pioneiro do projeto GALEN, nos anos 90, no uso em larga escala da lógica descritiva para a representação de domínios biomédicos, a maioria das ontologias e terminologias biomédicas atuais fazem uso de tal linguagem para representação do domínio. Um desses exemplos são as mais de 300.000 classes em lógica descritiva da terminologia clínica SNOMED-CT. Sobre a OWL-DL, Smith et al. (2007) afirma que seu uso como linguagem de representação das ontologias biomédicas do OBO vem aumentando consideravelmente nos últimos anos, tornando-se um padrão atual. Uma variedade de softwares editores de ontologia, tal como o Protégé, suportam o uso da OWL-DL, além de incluírem mecanismos de inferência para tal linguagem. Na presente pesquisa, utilizou-se o editor de ontologias Protégé 4.3 para a representação da teoria ontológica do sangue em OWL-DL.

Ainda sobre tais linguagens lógicas descritivas é importante destacar que uma de suas grandes vantagens em relação às outras linguagens é que as mesmas são bastante favoráveis às aplicações computacionais, porque são linguagens decidíveis, isto é, produzem algoritmos os quais retornam um resultado válido computacionalmente e, assim, podem ser interpretadas pelas máquinas. Por esse motivo, linguagens de lógica descritiva são muito usadas para a representação de quaisquer domínios, inclusive os biomédicos, sendo preferíveis em relação às linguagens lógicas de primeira ordem (FOL – First Order Logical). As linguagens lógicas de primeira ordem são mais expressivas que as da lógica descritiva, porém tem um poder bem menor de decidibilidade.

Entretanto, esse poder de decidibilidade das linguagens de lógica descritiva tem um custo alto para a representação do conhecimento de um dado domínio. Afim de assegurar a decidibilidade, linguagens descritivas como a OWL-DL utilizam expressões lógicas menos complexas, as quais possam ser interpretadas pelos mecanismos de inferência, diminuindo bastante seu poder expressividade. A consequência de tudo isso é que, de

maneira geral, as linguagens de lógica descritiva impõem sérias limitações à representação do conhecimento de um dado domínio e, conseqüentemente, das ontologias.

Alguns autores da literatura da área (SCHULZ et al., 2009; SCHULZ e JANSEN, 2008; NEUHAUS e SMITH, 2007; BODENREIDER et al., 2004; CEUSTERS, ELKIN e SMITH, 2007) descrevem essas restrições impostas pelas linguagens de lógica descritiva à representação ontológica. Algumas dessas restrições são:

- Universais são entidades não representáveis em linguagens de lógica descritiva, embora sejam fundamentais na construção de uma ontologia. Tais linguagens representam as entidades de um dado domínio através de “classes”, “relações” e “instâncias”, normalmente, através da tripla “sujeito – predicado – objeto”. Por exemplo, `Diabetes_Mellitus is_a Frequent_Disease; Blood_Plasma narrower_than Blood`.
- Um problema comum das linguagens de lógica descritiva é limitar o domínio tratado a um domínio normal ou canônico, ou seja, um domínio fechado àquilo que está representado, não considerando exceções. Um exemplo desse problema é a expressão `Hand has_part Thumb`, a qual considera que todos seres humanos que possuem mão tem como parte os dedos, excluindo, porém, os seres humanos com alguma deficiência dessa parte do corpo, humanos em estado embrionário inicial ou humanos com alguma má formação das mãos.
- Especificamente quanto à linguagem OWL-DL, suas propriedades de objetos (object properties) expressam relações binárias sem qualquer referência direta ao tempo, ou seja, índices temporais não são representáveis em OWL-DL, apesar dessa informação ser de extrema importância do ponto de vista ontológico. Uma célula sanguínea, por exemplo, ser formada de duas partes hoje, mas futuramente ser composta de três partes. Numa ontologia, a variação de tempo deve ser considerada.
- Outro problema refere-se à representação dos termos médicos que excede a expressividade lógica das linguagens descritivas. Expressões médicas que incluem a preposição “sem” (por exemplo: `concussão cerebral sem perda de consciência`) são difíceis de serem representadas nessas linguagens, devido à falta de construtores lógicos expressivos e à dificuldade de acordo unívoco sobre seu significado, já que envolve pressupostos tácitos relacionados ao tempo.

Diante dessas restrições impostas pelas linguagens lógicas descritivas à representação ontológica surgem duas questões fundamentais que devem ser consideradas em qualquer pesquisa que envolva a construção de ontologias e sua implementação em linguagens lógicas:

- “Qual o tipo de informação ontológica é apropriado para a representação em linguagens de lógica descritiva?”
- “O que fazer e como representar as informações contidas na ontologia cuja representação não é adequada em linguagens de lógicas descritiva?”

Não existem respostas prontas e exatas para essas questões levantadas, porém algumas referências da literatura da área fornecem dicas e orientações do que de fato deve ser representado em uma ontologia implementada através de uma linguagem formal. Schulz et al. (2009) e Smith e Ceusters (2010), por exemplo, sugerem que apenas deve ser representado em linguagem lógica descritiva os tipos de informação ontológica que tal linguagem é capaz de representar, tais como: classes, relações entre as classes, instâncias, cardinalidade, tipos de propriedades das classes, características das propriedades, definições em linguagem natural das classes que possam originar definições formais, entre outros tipos suportados. Ainda segundo tais autores, os tipos de informação ontológica, cuja representação não é suportada pelas linguagens lógicas descritivas, como a OWL-DL, não devem ser desprezados e incluídos na ontologia em forma de anotações e comentários.

Seguindo a abordagem desses autores, que é geralmente bem aceita na área, a presente pesquisa representa em OWL-DL e na lógica descritiva pura apenas o tipo de informação ontológica do sangue humano apropriado para representação em tais linguagens. Já quanto ao tipo de informação ontológica que não pode ser adequadamente representado em linguagens de lógica descritiva, esse será representado na ontologia em forma de “anotações”, uma vez que esse tipo de conhecimento é também muito importante no contexto biomédico tratado.

6 Metodologia

Nos capítulos anteriores (capítulos 2, 3, 4 e 5) apresentou-se o referencial teórico e empírico desta pesquisa. No presente capítulo, passa-se à descrição da metodologia de pesquisa adotada, incluindo os passos metodológicos executados para a consecução da presente pesquisa.

Para cumprir com os objetivos deste capítulo, optou-se por estruturá-lo nas seguintes seções: *seção 6.1*: apresenta-se a caracterização e o tipo de pesquisa realizada, com base na classificação proposta por referências da literatura; *seção 6.2*: apresenta-se a descrição geral da metodologia de pesquisa adotada, evidenciando os passos gerais que levaram à elaboração de uma metodologia própria para construção de ontologias por parte de cientistas da informação e sua aplicação prática no desenvolvimento da HEMONTO; e *seção 6.3*: descreve-se os passos realizados que permitiram a elaboração de uma metodologia própria para construção de ontologias em uma linguagem apropriada para cientistas da informação denominada de *OntoForInfoScience*.

6.1 Caracterização e tipologia da pesquisa

Na presente seção, é feita a caracterização desta pesquisa quanto à sua tipologia. Para tanto, consultou-se algumas referências (LAKATOS e MARCONI (1991); GIL (1991); CRESWELL (2003)) sobre metodologia de pesquisa científica, na literatura, como forma de classificar a pesquisa em questão.

Para uma melhor compreensão da classificação apresentada para a presente pesquisa, os parágrafos seguintes apresentam um resumo das definições dos tipos de pesquisa prescritos pelos autores das referências consultadas. Junto a essas definições apresentadas, aponta-se o tipo de pesquisa científica ao qual a presente pesquisa se enquadra.

Como forma de traçar a tipologia da pesquisa em questão, optou-se por classificá-la segundo os seguintes critérios (LAKATOS e MARCONI (1991)): (i) objetivos da pesquisa; (ii) natureza da pesquisa; (iii) procedimentos adotados; e (iv) forma de abordagem do problema de pesquisa.

Quanto aos objetivos de pesquisa, Gil (1991) e Lakatos e Marconi (1991) destacam que uma pesquisa pode ser classificada em: (i) *exploratória*, cuja finalidade é proporcionar maiores informações sobre determinado assunto, facilitar a delimitação de um tema de trabalho; (ii) *descritiva*, que pretende descrever as características de determinada população, envolvendo técnicas padronizadas de coleta de dados; e (iii) *explicativa*, que registra, analisa e interpreta os fenômenos estudados, procurando identificar suas razões,

seus fatores determinantes e suas causas. Considerando tais definições, a presente pesquisa pode ser classificada como **exploratória**, já que engloba a criação de artefatos de representação da informação através da ontologia desenvolvida como forma de gerar maiores informações sobre as áreas de hematologia e hemoterapia; e **explicativa**, porque envolve o registro, a análise e a interpretação de dois tipos de fenômenos: (i) aqueles relacionados com o domínio do sangue humano que são registrados através da ontologia desenvolvida; e (ii) aqueles referentes ao processo de construção de uma ontologia.

Quanto à natureza, Lakatos e Marconi (1991) afirmam que uma pesquisa pode constituir-se de um *trabalho original* ou de um *resumo de assunto*. No primeiro caso, trata-se de uma pesquisa realizada pela primeira vez, que venha a contribuir com novas conquistas e descobertas para a evolução do conhecimento científico; no segundo, consiste de uma pesquisa fundamentada em trabalhos mais avançados, publicados por autoridades no assunto, e que dispensa a originalidade, mas não o rigor científico. Com relação a esse critério é trivial deduzir que a presente pesquisa, por ser uma tese de doutorado, constitui-se de um **trabalho original**, que corresponde à construção uma metodologia própria em ontologias para cientistas da informação e uma teoria ontológica do sangue humano, até aqui, inéditas na literatura da área.

Por fim, quanto à forma de abordagem do problema, uma pesquisa pode ser classificada em (CRESWELL (2003); LAKATOS e MARCONI (1991)): (i) *pesquisa quantitativa*, a que traduz em números e taxas as opiniões e informações coletadas, utilizando-se de técnicas estatísticas para serem classificadas e analisadas. Neste tipo de pesquisa, o problema é conduzido por fatores e variáveis e utilizam-se de experimentos, coletas e observação dos dados; (ii) *pesquisa qualitativa*, que é uma pesquisa descritiva, cujas informações não são quantificáveis e os dados obtidos são analisados indutivamente. Assim, a interpretação dos fenômenos e a atribuição de significados são básicas e fundamentais no processo de pesquisa qualitativa; (iii) *pesquisa híbrida*, que envolve as características qualitativa e quantitativa de uma pesquisa em um único estudo, através de coleta e análise de dados. Tal pesquisa engloba métodos de campo como observação (abordagem qualitativa) e levantamento de dados (abordagem quantitativa). Considerando tais definições, a presente pesquisa pode ser classificada como **qualitativa**, porque se conduz um estudo sobre metodologias para construção de ontologias e também sobre um determinado domínio do conhecimento – sangue humano –, nos quais são adquiridas informações sobre ambos e constrói-se uma compreensão de maneira interpretativa destas duas áreas. No caso das metodologias pesquisadas, a interpretação é construída com base na literatura da área e nas análises feitas por este autor, enquanto no caso da ontologia sobre o sangue, a interpretação é desenvolvida com base em fontes de

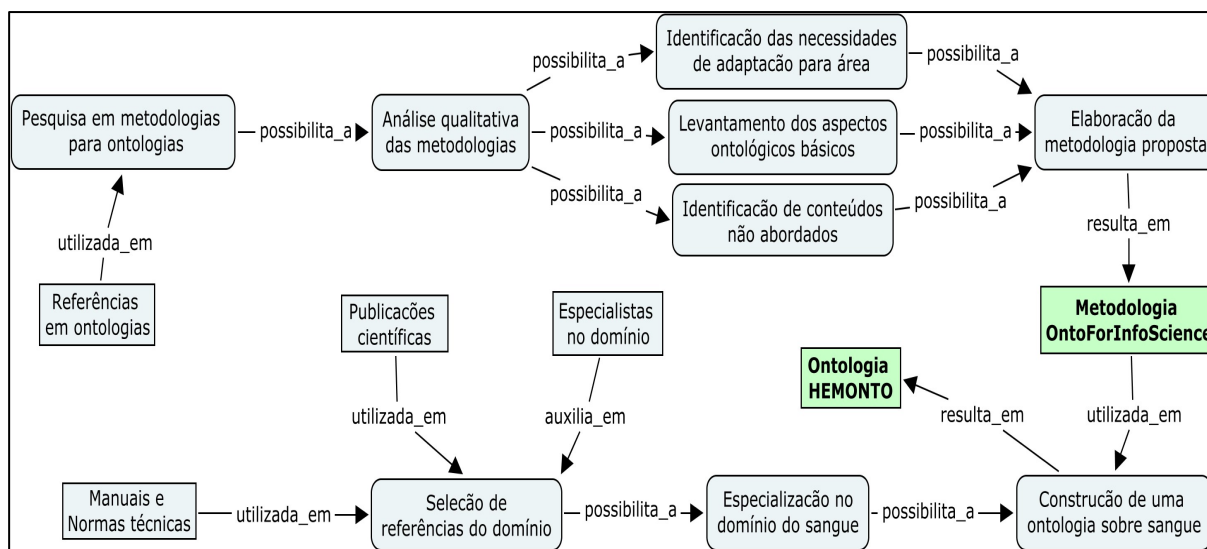
pesquisa da área e nas ideias de especialistas sobre o tema. Sabe-se ainda que o conhecimento apreendido no domínio do sangue humano, do ponto de vista ontológico, não pode ser mensurado quantitativamente.

6.2 Descrição geral da metodologia de pesquisa

A presente seção apresenta a descrição geral da metodologia de pesquisa explicitando os passos metodológicos realizados para a elaboração da metodologia *OntoForInfoScience* e sua aplicação no desenvolvimento da ontologia sobre hemocomponentes e hemoderivados do sangue humano (HEMONTO).

Afim de apresentar tal metodologia de pesquisa, construiu-se um esquema de representação englobando os passos realizados, conforme mostrado na Figura 32, a seguir.

Figura 32 - Esquema de representação da metodologia de pesquisa adotada.



Fonte: elaborado pelo autor.

O esquema de representação da Figura 32 contém três tipos de estruturas informacionais: (i) *objetos* (representados como retângulos), tais como documentos, pessoas e instrumentos de representação da informação; (ii) *processos ou atividades ou tarefas* (representadas como retângulos arredondados), que correspondem às ações (passos) executadas a cada etapa de realização da pesquisa; e (iii) *relacionamentos* (representados por setas), que permitem a interligação dos objetos e processos utilizados na pesquisa, especificando a relação entre eles e explicitando a noção de seqüência na qual são realizados.

Acerca dos relacionamentos utilizados para desenhar o esquema de representação da metodologia de pesquisa, cada um deles é nomeado ou etiquetado com um verbo, que expressa a semântica da relação. As definições de tais relacionamentos são apresentadas na Tabela 22, a seguir.

Tabela 22 - Conjunto de relações usado no esquema de representação da metodologia

Relacionamento	Definição
<i>utilizada_em</i>	Representa a relação de utilização de um objeto em um determinada atividade. Exemplo: Manuais e normas técnicas são utilizadas na seleção de referências sobre o domínio do sangue.
<i>auxilia_em</i>	Indica que um determinado objeto fornece suporte (auxílio) na realização de determina atividade. Exemplo: Especialistas no domínio do sangue auxiliam na seleção de referências deste domínio.
<i>possibilita_a</i>	Representa a relação entre duas atividades (dois processos), na qual a atividade anterior possibilita a execução da atividade seguinte. Exemplo: A análise qualitativa das metodologias para construção de ontologias possibilitou, entre outras atividades, o levantamento de aspectos ontológicos básicos que devem ser incluídos nesse tipo de metodologia.
<i>resulta_em</i>	Representa a relação entre uma atividade e um objeto, através da qual, após a execução da atividade, o objeto é gerado como um produto resultante de sua execução. Exemplo: A construção de uma ontologia sobre o sangue, através da metodologia proposta, deu origem a HEMONTO.

Fonte: elaborado pelo autor.

O esquema de representação da Figura 32 tem dois pontos de partida que se iniciaram paralelamente, especificamente as atividades de: (i) *seleção de referências no domínio do sangue humano* e (ii) *pesquisa exploratória e qualitativa*, na literatura da área, *sobre as atuais metodologias* para construção de ontologias.

Na atividade (i) procedeu-se com a seleção de textos técnico-científicos nas áreas de hematologia e hemoterapia com o apoio de especialistas neste domínio para avaliação do grau de relevância e representatividade de tais textos. Foram selecionadas tanto publicações científicas atuais, contendo os avanços recentes na área, quanto manuais e normas técnicas da literatura norte-americana e também da brasileira, de ampla utilização neste domínio de conhecimento. A seleção e o estudo aprofundado nas referências do domínio possibilitou a atividade de *especialização do conhecimento no domínio do sangue humano*, necessária para a construção da ontologia desenvolvida.

Paralelamente, realizou-se a atividade de *pesquisa exploratória sobre as metodologias de construção para ontologias* mais conhecidas, cujo propósito não foi de

descrevê-las, mas sim levantar informações necessárias de cada metodologia para a atividade seguinte de *análise qualitativa das metodologias*. Tal análise foi realizada do ponto de vista da Ciência da Informação, buscando perceber as metodologias de construção atuais na visão do cientista da informação, suas familiaridades e dificuldades com o conteúdo prescrito em tais metodologias. Nesse sentido, a análise realizada levou à identificação de pontos importantes a serem considerados na elaboração de uma metodologia própria de construção de ontologias para cientistas da informação, os quais, de maneira geral, são: (i) *necessidades de adaptação dos conteúdos das metodologias atuais para a área de Ciência da Informação*; (ii) *aspectos ontológicos básicos a serem incluídos em uma metodologia de construção*; e (iii) *necessidade de inclusão de conteúdos não abordados nas metodologias atuais na metodologia proposta*. Esses três pontos levantados estão representados como atividades do esquema de representação da metodologia de pesquisa e são o resultado da análise qualitativa realizada na presente pesquisa.

Na sequência da metodologia geral de pesquisa, o levantamento de aspectos ontológicos básicos, a identificação de necessidades de adaptação das metodologias atuais e a identificação de conteúdos não abordados, possibilitaram a *elaboração da metodologia proposta – OntoForInfoScience* – baseada em fundamentos filosóficos e princípios ontológicos de construção em uma abordagem direcionada a cientistas da informação. O objetivo foi de criar uma metodologia fundamentada, mas que pudesse ser expressa em uma linguagem textual e gráfica de fácil compreensão por parte dos cientistas da informação.

Ao final da realização dessas duas pequenas cadeias de processos tem-se as condições necessárias para a construção de uma ontologia sobre o sangue, vista como uma teoria ontológica, a partir da metodologia proposta. O resultado dessa construção é a ontologia HEMONTO, cujo conteúdo engloba aspectos fisiológicos do sangue humano e também os produtos (hemocomponentes e hemoderivados) do sangue usados nas áreas de hematologia e hemoterapia, e que serve como um estudo de caso prático para aplicação da metodologia *OntoForInfoScience* proposta.

6.3 Passos adotados na elaboração da metodologia *OntoForInfoScience*

Após a caracterização da presente pesquisa e a apresentação da metodologia geral adotada, passa-se, nesta seção, à descrição dos passos executados para a elaboração de uma metodologia própria para a construção de ontologias por parte dos cientistas da informação.

Afim de atingir seu propósito principal, a metodologia *OntoForInfoScience* foi desenvolvida a partir de uma perspectiva geral da área de Organização e Representação do Conhecimento ligada à Ciência da Informação, diferentemente das metodologias atuais que, em geral, se inserem mais em uma abordagem voltada para a área de Engenharia de Software. Assim, torna-se necessário quebrar um pouco desse paradigma de construção de metodologias para ontologias do ponto de vista computacional e desenvolver uma linguagem de mais fácil entendimento por parte de especialistas em Organização do Conhecimento provenientes da área de Ciência da Informação, tornando as metodologias recursos compreensíveis e utilizáveis por quaisquer desenvolvedores de ontologias.

No caso da *OntoForInfoScience*, a estratégia utilizada para o desenvolvimento de uma metodologia em uma linguagem mais compreensível e utilizável passou, necessariamente, pela apreensão do conhecimento das metodologias de construção ontológica mais relevantes e utilizadas atualmente. Tais metodologias englobam o conhecimento e a experiência adquirida nos últimos anos no desenvolvimento de ontologias, extremamente necessários para atuação na área. Entretanto, é necessário recriar, fazer adaptações ao que já foi ou está sendo desenvolvido, para tornar mais intensa a produtividade no desenvolvimento de ontologias por parte dos cientistas da informação.

Portanto, o primeiro passo em direção à elaboração da metodologia *OntoForInfoScience*, foi uma revisão de literatura sobre as metodologias para construção de ontologias, desde as primeiras, tais como a metodologia de Grüninger e Fox – TOVE (GRÜNINGER e FOX, 2005) e o método de Uschold e King – Enterprise Ontology (USCHOLD e KING, 2005), até as mais recentes, como, por exemplo, a metodologia MFPFO (LIM, LIU e LEE, 2011) e a versão mais recente da metodologia NeOn (SUÁREZ-FIGUEROA, 2010). Essa revisão teve como foco analisar o propósito geral de cada metodologia, suas aplicações práticas e, principalmente, suas etapas para o desenvolvimento de ontologias e foi apresentada no referencial teórico desta pesquisa.

Como segundo passo, realizou-se uma análise qualitativa das metodologias pesquisadas com o objetivo de identificar aquelas que mais seriam apropriadas ao desenvolvimento proposto na metodologia *OntoForInfoScience*. O primeiro resultado dessa análise possibilitou duas atividades primordiais na elaboração da metodologia desenvolvida: (i) o levantamento de aspectos ontológicos básicos que deveria estar incluído em qualquer metodologia para construção de ontologias; e (ii) a seleção das metodologias que serviram de referência para o desenvolvimento da *OntoForInfoScience*. São elas: a Methontology (FERNÁNDEZ et al., 1997), o método 101 (NOY e GUINNESS, 2001) e NeOn Methodology (SUÁREZ-FIGUEROA, 2008).

O segundo resultado da análise qualitativa realizada corresponde a dois processos já mencionados no esquema de representação da metodologia geral de pesquisa: (i) a identificação das necessidades de adaptação das metodologias de construção para a área de Ciência da Informação; e (ii) a identificação de conteúdos não abordados em tais metodologias que são indispensáveis ao desenvolvimento de ontologias por parte dos cientistas da informação. Esses dois processos resultantes da análise qualitativa foram de fundamental importância para a elaboração da metodologia *OntoForInfoScience*, já que a partir deles foi possível desenvolver um plano estratégico de ação, que funcionasse como um guia de desenvolvimento para a metodologia proposta. Tal plano estratégico é apresentado na Tabela 23, a seguir, e inclui o seguinte conteúdo: (i) a funcionalidade esperada para a metodologia de construção; (ii) de que forma as metodologias reutilizadas (*NeOn*, *Methontology* e *método 101*) tratam cada funcionalidade mencionada; e (iii) qual a ação ou estratégia indicadas para a metodologia em desenvolvimento atender a funcionalidade de construção esperada. Esse plano de ação desenvolvido é o guia metodológico na elaboração da *OntoForInfoScience*.

Tabela 23 - Plano de ação para elaboração da metodologia *OntoForInfoScience*

Funcionalidade	Metodologia NeOn	Methontology	Método 101	Ações/Estratégia
Avaliação da necessidade do uso de ontologias para resolução do problema ao invés de um instrumento terminológico	Não inclui essa avaliação.	Não inclui essa avaliação.	Não inclui essa avaliação.	Criar uma etapa preliminar na metodologia para alertar ao desenvolvedor sobre a real necessidade ou não do uso de ontologias no problema.
Linguagem apropriada aos cientistas da informação	Apenas em algumas etapas da NeOn, como na especificação da ontologia, a linguagem é adequada. Nas demais, como nos cenários de desenvolvimento, a linguagem é mais direcionada para engenheiros de software.	Linguagem mais próxima da Ciência da Informação, porém não há muitos detalhes na realização de suas etapas.	Em muitas etapas necessita-se de conhecimentos prévios em Computação, tais como: tipos de dados, cardinalidade, disjunção, classes abstratas.	Continuar desenvolvendo a ontologia com os atributos necessários que tem relação com a Computação (cardinalidade, equivalência, disjunção, herança múltipla, etc.), porém explicando o significado de cada um deles em uma linguagem adequada.
Especificação do ponto de partida e limite do domínio coberto pela ontologia	Realizada na etapa de especificação da ontologia e através de questões de competência.	Realizada na etapa de especificação da ontologia.	Essencialmente através de questões de competência.	Incluir na etapa de especificação da ontologia, entretanto, tornando mais explícito de onde parte a ontologia (classes mais gerais) e onde pára (nível máximo de especialização no domínio).

				Incluir também questões de competência para delimitação do escopo.
Explicação dos métodos de aquisição e extração do conhecimento	Não menciona métodos de aquisição e extração do conhecimento.	Possui uma etapa na metodologia que relaciona alguns métodos de aquisição e extração do conhecimento, porém sem detalhá-los.	Não menciona métodos de aquisição e extração do conhecimento.	Incluir uma etapa de aquisição e extração do conhecimento para a ontologia e detalhar pelo menos o método de aquisição e extração a ser utilizado na ontologia.
Uso de colaboração na conceitualização do domínio	Realizada a partir do uso de redes colaborativas na conceitualização do domínio.	Inclui algum tipo de colaboração no desenvolvimento de artefatos de representação, como os glossários e dicionários, mas sem especificar métodos e formas de colaboração.	Não trata colaboração no desenvolvimento da ontologia.	Tratar a colaboração como um aspecto essencial na etapa de conceitualização. Selecionar algum método ou metodologia de colaboração para conceitualização do domínio e utilizá-lo na etapa de conceitualização, por exemplo, através do uso de ferramentas colaborativas.
Ontologias de fundamentação e reutilização	Incentiva o uso de ontologias de fundamentação na etapa de reuso de recursos ontológicos, na qual tais ontologias são referenciadas como ontologias comuns ou de uso geral, para diferenciá-las das ontologias de domínio, cuja reutilização também é incentivada.	Prevê a utilização das meta-ontologias (equivalentes as ontologias de fundamentação) no desenvolvimento, especificamente, na etapa de integração.	Não faz referência às ontologias de fundamentação. A reutilização de termos é incentivada a partir de quaisquer ontologias existentes – não há uma especificação.	Destinar uma etapa da metodologia proposta para incentivar o uso de ontologias de fundamentação como ponto de partida do desenvolvimento ontológico. Nessa etapa, explicar como deve ser feita essa utilização, incluindo o uso de editores de ontologia.
Detalhamento sobre pesquisa de termos em outras ontologias para reutilização	Na etapa de reuso de recursos ontológicos, é apresentada uma descrição detalhada de como selecionar uma ontologia para reuso a partir de seus princípios e definições. O detalhamento é mais na ontologia como um todo do que na busca de termos específicos.	Na etapa de integração, sugere-se o uso das bibliotecas de ontologias para pesquisa pelos termos e suas definições, sem dar detalhes dessa busca.	Sugere-se como pesquisa aos termos, o acesso às bibliotecas on-line de ontologias (repositórios), porém sem nenhum detalhamento.	Incluir na parte da metodologia que trata da reutilização, um pequeno check-list de como pesquisar por termos relacionados ao domínio para importação na ontologia desenvolvida.
Explicação sobre as propriedades (atributos) das classes e relações	Na etapa de avaliação da ontologia são definidos alguns critérios para identificar erros no	Não inclui essa funcionalidade.	Inclui explicações passo-a-passo de como definir as propriedades descritivas (slots) e também as	Na etapa de desenvolvimento ontológico, descrever o significado de cada propriedade a ser definida e apresentar

	uso incorreto das propriedades. Esses critérios podem ser usados para o entendimento das propriedades das classes e relações		propriedades lógicas (facets dos slots) de classes e relações, porém sem detalhar o significado de cada propriedade a ser definida.	diretrizes para essa definição, inspiradas nas explicações do método 101 e nos critérios para propriedades da NeOn.
Explicação sobre construções formais (lógicas) dos elementos da ontologia	Não inclui essa funcionalidade.	Não inclui essa funcionalidade.	Não inclui essa funcionalidade.	Na etapa de desenvolvimento ontológico, incluir um passo para explicar como se deve construir uma definição ou restrição formais de uma classe ou relação, usando alguma linguagem lógica, cuja explicação seja compreensível para cientistas da informação.
Caracterização das relações ontológicas	Na etapa de avaliação da ontologia são definidos alguns critérios para identificar erros no uso da relação "is_a", a qual pode ser usada com três significados diferentes: "subClassOf", "instanceOf" e "sameIndividual".	Não inclui essa funcionalidade.	A relação "is_a", usada na hierarquia de classes, é tratada como sinônimo das relações "SubClassOf" e "KindOf", representando a relação entre uma superclasse e uma subclasse.	Novamente na etapa de desenvolvimento ontológico, criar um passo que permita caracterizar a relação ontológica no momento de sua inclusão na ontologia. Caracterizar significa distinguir os diferentes tipos da relação ontológica e isso deve ser feito pelo menos para as relações "is_a" e "part_of".
Uso e explicação de critérios de avaliação do conteúdo ontológico	Inclui uma complexa estrutura de avaliação para redes ontológicas, que envolve uma meta-ontologia O^2 para descrever os elementos que devem ser avaliados e modelos de avaliação específicos, denominados oQual. Inclui ainda um conjunto de critérios para identificação de erros comuns no desenvolvimento, chamados de pitfalls.	Inclui um framework formal de avaliação, que engloba técnicas baseadas nos métodos de validação e verificação dos sistemas de base de conhecimento e também um conjunto de diretrizes para avaliar incompletudes, inconsistências e redundâncias.	Não inclui uma etapa própria de avaliação, porém utiliza critérios de avaliação ao longo do ciclo de desenvolvimento, por exemplo, avaliação da hierarquia de classes pela relação "is_a", identificação de ciclos, herança múltipla, etc.	Estabelecer um conjunto de critérios avaliativos do conteúdo ontológico, que envolvam validação e verificação, baseados nos critérios de identificação de erros da NeOn, nas restrições da teoria subjacente a BFO e em critérios do método 101. A avaliação não será implementada através de um framework formal e nem contemplará redes ontológicas, devido ao tempo e escopo da pesquisa.
Documentação da ontologia	Ao final de cada etapa da metodologia, constrói-se um	Recomenda a documentação ao final de cada fase do ciclo de vida do	Não inclui uma etapa própria de documentação, mas cita a importância	Incluir uma etapa para orientar na documentação da ontologia,

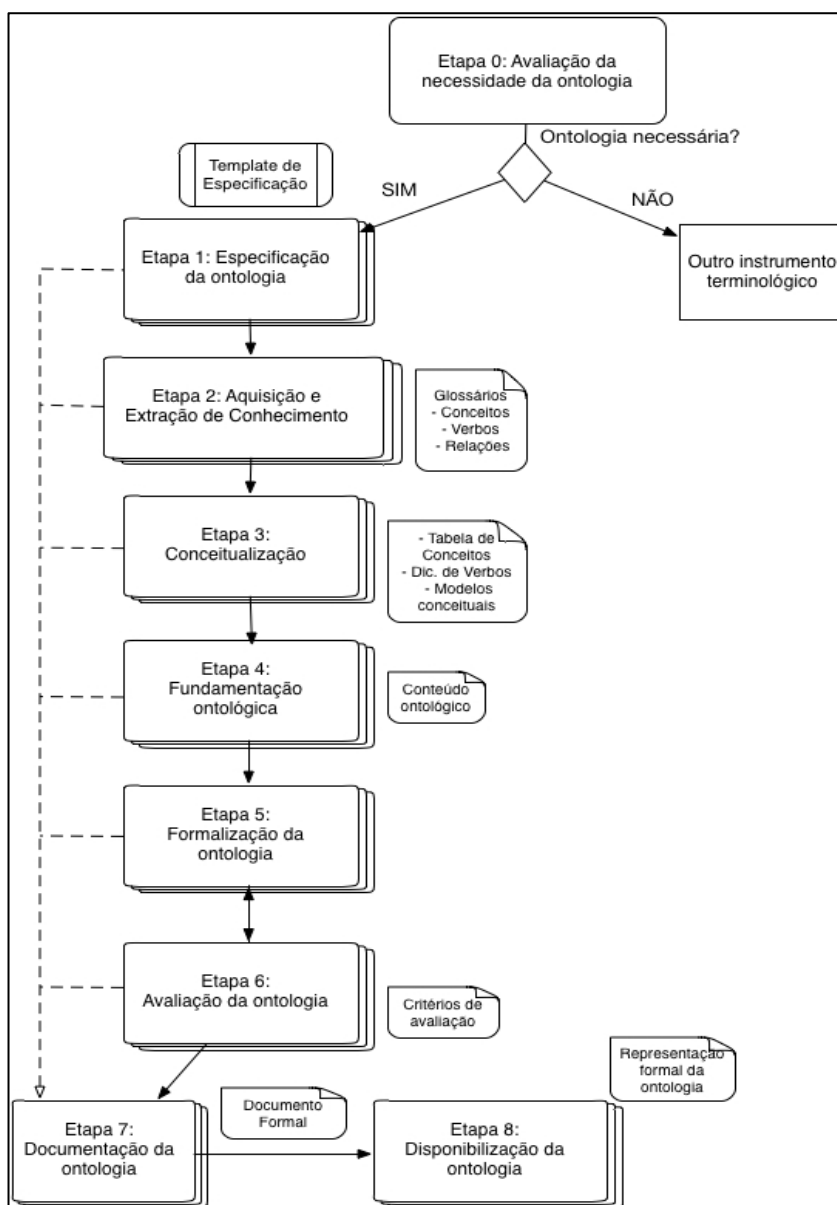
	documento formal dos resultados obtidos, que variam de acordo com o cenário de desenvolvimento.	desenvolvimento da ontologia, prevendo, assim, um documento formal de especificação, outro do modelo conceitual, outro de formalização e assim por diante.	de se documentar a ontologia em seu ciclo de desenvolvimento e também trata a convenção de nomes para os elementos da ontologia.	ressaltando, porém, que o processo de documentação deve ser realizado ao final de cada etapa e atualizado no final.
Formas de apresentação e/ou disponibilização da ontologia	Não se tem recomendações específicas de como apresentar ou disponibilizar a ontologia na web. Porém, isto está implícito com o trabalho em redes ontológicas e as atividades executadas em cada cenário de desenvolvimento.	Prevê a apresentação da ontologia através dos documentos formais da etapa de documentação e sua disponibilização através da codificação da ontologia em alguma linguagem formal, como CLASSIC, LOOM, OntoLingua, Prolog, C++, etc.	Como o método foi desenvolvido com base no Protégé, a forma de apresentação da ontologia são os recursos desse software.	Sugerir ao desenvolvedor formas simples e claras de apresentação do conteúdo ontológico aos seus usuários, incluindo especialistas do domínio, além de uma forma de disponibilização da ontologia para seu acesso via web.

Fonte: elaborado pelo autor.

7 Metodologia *OntoForInfoScience*

Após a apresentação do objetivo principal, justificativas e estratégias adotadas para a metodologia *OntoForInfoScience*, passa-se à descrição detalhada das etapas e passos incluídos em tal metodologia. Nessa direção, o presente capítulo concentra-se em explicar cada uma das etapas da metodologia proposta, bem como os passos (ou subetapas) incluídos em cada uma delas. A cada passo descrito é também apresentado um exemplo de sua execução no desenvolvimento da ontologia sobre os componentes do sangue humano, a HEMONTO. Nesse sentido, o presente capítulo funciona como um guia ou roteiro para a construção de ontologias por parte dos cientistas da informação, especialmente em domínios biomédicos.

Este capítulo 7 foi dividido em seções de acordo com as etapas incluídas na metodologia *OntoForInfoScience*, sendo uma seção para cada etapa prescrita. Ao todo temos 9 (nove) etapas na metodologia, sendo a etapa 0 (zero) preliminar. Elas estão representadas na Figura 33, a seguir, conforme sua ordem de realização na metodologia.

Figura 33 - Etapas da metodologia *OntoForInfoScience*

Fonte: elaborado pelo autor.

A Figura 33 representa as nove etapas da metodologia *OntoForInfoScience*, em sua sequência de execução na metodologia, e os artefatos gerados após cada etapa realizada. A pré-etapa ou etapa zero é representada como um quadrado simples, as demais etapas são representadas por um quadrado composto ilustrando os passos incluídos em cada uma e os artefatos gerados são representados por símbolos indicando um documento, próximos da etapa em que foram gerados.

Na etapa zero da metodologia *OntoForInfoScience* realiza-se a avaliação da necessidade do desenvolvimento de uma ontologia para o problema analisado. Em caso negativo, recomenda-se o uso de outro instrumento terminológico, tais como tesouros,

taxonomias ou vocabulários controlados, na solução proposta. Por outro lado, caso o desenvolvimento da ontologia se faz necessário, tem-se a sequência de etapas do processo de construção. Na etapa 1 realiza-se a especificação da ontologia, através do template de especificação, que quando preenchido é o artefato resultante desta etapa. Na etapa 2 tem-se a fase de aquisição e extração de conhecimento para construção da ontologia e seus artefatos resultantes são três glossários: um de conceitos, um de verbos e outro de relações, os quais são utilizados na etapa 3. Nesta etapa 3 realiza-se a conceitualização do domínio sob estudo e seus artefatos resultantes são: a tabela de conceitos e propriedades, o dicionário de verbos e os modelos conceituais gráficos. A etapa 4 envolve a seleção e utilização de ontologias de fundamentação para a ontologia em desenvolvimento. Por sua vez, a etapa 5 corresponde à representação formal da ontologia cujo conteúdo engloba classes, relações, propriedades, axiomas e estruturas de representação ontológicas. Esse conteúdo é avaliado na etapa 6 (etapa de avaliação) da metodologia sob o conjunto de critérios avaliativos definidos. Observe que nesta interação das etapas 5 e 6 a seta é bidirecional para indicar que a avaliação do conteúdo ontológico gera alterações que precisam ser atualizadas nele. Na etapa 7, de documentação da ontologia, produz-se o documento formal dessa ontologia, tal que sua elaboração é realizada, praticamente, ao longo de todo processo de construção da ontologia, conforme indicado na Figura 33 através das setas pontilhadas. Por fim, a ontologia desenvolvida é disponibilizada, formalmente e graficamente, aos seus usuários na etapa 8.

De forma a complementar esse esquema de representação sequencial da metodologia *OntoForInfoScience*, construiu-se também uma tabela resumida (veja Tabela 24, a seguir) de todas as etapas previstas na metodologia proposta e os artefatos de representação gerados ao final de cada uma delas. Esta Tabela 24 é uma visão resumida ou mesmo um índice da metodologia *OntoForInfoScience*, cujas etapas e passos são descritos, detalhadamente, nas próximas seções deste capítulo.

Tabela 24 - Passos realizados e artefatos gerados a cada etapa da *OntoForInfoScience*

n.	Etapa	Passos	Artefatos gerados
0	Avaliação da necessidade da ontologia	–	–
1	Especificação da ontologia	Preencher template de especificação com informações sobre: <ul style="list-style-type: none"> - Domínio/Escopo Geral - Propósito geral - Classes de usuários - Uso pretendido - Tipo da ontologia - Grau de formalidade - Delimitação do escopo de cobertura 	Documento de especificação
2	Aquisição e extração de	- Adotar métodos de aquisição e	- Pré-glossário de termos

	conhecimento	extração de conhecimento - Selecionar materiais de referência do domínio	(artefato parcial) - Conjunto automático de candidatos (artefato parcial) - Conjunto manual de candidatos (artefato parcial) - Glossário de conceitos (final) - Glossário de verbos (final) - Glossário de relações (final)
3	Conceitualização	- Construir o dicionário de conceitos - Elaborar a tabela de conceitos e valores - Elaborar a tabela de conceitos e propriedades - Construir o dicionário de verbos - Representar modelos conceituais	- Dicionário de conceitos (artefato parcial) - Tabela de conceitos e valores (artefato parcial) - Tabela de conceitos e propriedades (final) - Dicionário de verbos (final) - Modelos conceituais gráficos
4	Fundamentação ontológica	- Selecionar ontologia(s) de fundamentação para uso - Aplicar ontologia selecionada no desenvolvimento	Conteúdo da ontologia de fundamentação sob uma representação formal
5	Formalização da ontologia	- Construir a taxonomia geral - Definir propriedades descritivas das classes - Criar definições formais das classes - Definir propriedades de dados das classes - Criar instâncias das classes - Especificar relações ontológicas - Definir propriedades das relações	- Classes - Relações ontológico-formais - Propriedades descritivas - Propriedades lógicas - Axiomas lógicos - Taxonomia geral - Estruturas de representação gráfica
6	Avaliação da ontologia	- Definir parâmetros de avaliação - Estabelecer critérios de validação - Estabelecer critérios de verificação	Conjunto de critérios de avaliação
7	Documentação da ontologia	- Produzir o documento final com o conteúdo da ontologia	Documento formal da ontologia
8	Disponibilização da ontologia	- Gerar representação da ontologia em linguagem formal - Apresentar a ontologia em formato e meio eletrônico	- Representação da ontologia em linguagem lógica (por exemplo, em OWL). - URL de acesso à ontologia

Fonte: elaborado pelo autor.

7.0 Avaliação da necessidade da ontologia (Etapa 0)

Como etapa preliminar na metodologia de construção *OntoForInfoScience*, criou-se uma etapa 0 (zero) na qual o(s) responsável(is) pelo desenvolvimento da ontologia deve fazer uma avaliação prévia sobre a real necessidade de criação da mesma, ou seja, antes de iniciar o desenvolvimento deve-se responder a seguinte questão básica:

- “O projeto a ser desenvolvido e seu contexto necessitam ou demandam a construção de uma ontologia? Ou a criação de outro instrumento de representação, tal como um tesouro, seria suficiente?”

Embora, à primeira vista, possa parecer trivial a resposta a essa questão nem sempre está claro para o desenvolvedor a real diferença entre utilizar uma ontologia, um tesauro ou outro vocabulário controlado em um projeto específico. A orientação da metodologia proposta neste caso é:

- Se o objetivo do projeto consiste unicamente na indexação e recuperação de informações em um domínio específico, o uso de um tesauro consistente é recomendado, uma vez que a partir dele é possível formalizar a padronização de termos da linguagem específica do domínio, eliminando ambiguidades no uso dos termos e, assim, auxiliar o usuário no refinamento de suas buscas e localização da informação desejada. Nesses casos, não se faz necessário a construção de uma ontologia, já que um tesauro consegue atender bem a demanda.
- Por outro lado, se o objetivo do projeto corresponde a indexação e recuperação de informação em um contexto dinâmico para descrição de recursos de um dado domínio do conhecimento a utilização de ontologias é fundamental, pois nestes casos alguma dedução e inferência automatizada é necessário, além da necessidade do uso de relações extensíveis, o que é conseguido com ontologias.
- Outras situações que demandam o uso de ontologias referem-se à necessidade de representação de aspectos e objetos do mundo real, a necessidade de mais relações entre os objetos de um domínio além daquelas contidas em vocabulários controlados e o uso de formalismos lógicos para representação da informação e inferência automatizada.

Nesse sentido, a etapa zero da metodologia *OntoForInfoScience* recomenda ao desenvolvedor um exercício prévio e avaliativo do real objetivo do seu projeto para que ele escolha o instrumento de organização e representação da informação adequado, conforme as diretrizes apresentadas.

7.1 Especificação da ontologia (Etapa 1)

Se o resultado da etapa 0 da metodologia proposta indica a construção de uma ontologia para o projeto e contexto sob estudos, passa-se à primeira etapa do processo de construção, o qual é conhecido como fase de *especificação da ontologia*. Nesta etapa deve-se definir o propósito geral e os objetivos específicos da ontologia, o tipo da ontologia

desenvolvida, seu escopo de cobertura, cenários de uso, para que tipos de usuários a ontologia se destina, o grau de formalidade requerido e um pré-glossário de termos. Afim de descrever como o desenvolvedor da ontologia deve proceder para especificar tais informações, construiu-se um template desta etapa detalhando cada uma das informações necessárias à especificação da ontologia, conforme apresentado no Quadro 1, a seguir.

Quadro 1 - Template de Especificação de uma ontologia

Domínio/Esopo Geral
Especificar o domínio do conhecimento que a ontologia irá representar, detalhando quais partes deste domínio serão tratadas na ontologia e quais não serão, isto é, delimitar o escopo geral de cobertura da ontologia desenvolvida.
Propósito Geral
Explicar o motivo e o objetivo da construção da ontologia, ou seja, porquê e para que a ontologia em questão será desenvolvida. Aqui é importante ressaltar que o propósito geral da ontologia deve estar bem claro e definido, de maneira a assegurar que sua utilização por seus usuários traga os resultados esperados. Um texto claro, objetivo e conciso é capaz de explicar bem seu propósito.
Classes de usuários
Especificar todos possíveis usuários da ontologia, agrupando-os em classes de acordo com seus objetivos em relação a mesma. Recomenda-se a definição de classes que envolvam: desenvolvedores e mantenedores da ontologia, usuários colaboradores da ontologia, usuários leitores e usuários das aplicações derivadas da ontologia criada.
Uso pretendido
Elaborar alguns cenários de aplicação para o uso da ontologia, que visam mostrar problemas no ambiente atual onde a ontologia será aplicada e o que se espera com sua utilização.
Tipo da ontologia
Classificar a ontologia desenvolvida quanto à: (i) estrutura ou nível de conhecimento: <i>ontologia de alto nível</i> , <i>ontologia de domínio</i> ou <i>ontologia de tarefa</i> ; (ii) grau de formalidade: <i>ontologia leve</i> , <i>ontologia com médio rigor formal</i> ou <i>ontologia pesada</i> ; (iii) propósito de criação (opcional): <i>ontologia de sistemas de informação</i> ou <i>ontologia para sistemas de informação</i> . Tal classificação é útil na definição do grau de formalidade da ontologia, em suas classes de usuários e seu uso pretendido.
Grau de formalidade
Definir o grau de formalidade requerido para a ontologia com base em seu propósito geral, cenários de uso ou aplicação, classes de usuários e tipo da ontologia. Assim, pode-se requerer, por exemplo, uma <i>ontologia pesada</i> , com alto rigorosidade formal, que use axiomas lógicos bem definidos e completos e exige uma linguagem com mais recursos para sua formalização, tal como a lógica de primeira ordem (First Order Logical-FOL); ou, em outro extremo, uma <i>ontologia leve</i> , com menos rigor formal e cujo objetivo é sua implementação em uma linguagem como OWL-Lite para possibilitar a integração entre sistemas computacionais. Entre esses dois extremos existem

ontologias com médio rigor formal, as quais incluem alguns axiomas lógicos e são representadas em alguma linguagem lógica com maior expressividade em relação a versões como OWL-Lite e, ao mesmo tempo, menor formalização que linguagens como a lógica de primeira ordem (FOL).

Delimitação do escopo de cobertura

Afim de assegurar que a delimitação do escopo de cobertura da ontologia seja clara e explícita para todas as classes de usuários, recomenda-se ao desenvolvedor especificar os seguintes itens, considerados complementares e indispensáveis nessa tarefa:

- **Ponto de partida da ontologia:** corresponde ao conjunto de classes iniciais a partir das quais a ontologia começa a ser desenvolvida. Geralmente, esse conjunto deve ser formado por um conjunto de classes mais gerais reaproveitadas de alguma ontologia de fundamentação ou alto nível e mais uma ou duas classes específicas do domínio tratado.

Exemplo: definição do ponto de partida na ontologia HEMONTO

O ponto de partida da HEMONTO são entidades do mundo real, categorizados por meio das classes da ontologia BFO, e mais o termo “sangue” da FMA (FMA: *portion of blood*). A partir do termo FMA: *portion of blood* são obtidas as especializações do sangue (ou seja, seus componentes e tipos) e as generalizações (ou seja, o “sangue” compreendido como parte do corpo humano).

- **Limite do domínio coberto:** corresponde ao nível de conhecimento máximo do domínio que será representado na ontologia, isto é, em que nível de granularidade do conhecimento do domínio a ontologia pára, tal que o conhecimento além desse nível de granularidade é tratado em outra ontologia ou outro sistema de representação ou até mesmo não é tratado.

Exemplo: definição do limite do domínio coberto na ontologia HEMONTO

No limite do domínio, parte-se do termo “sangue” e, a partir de uma abordagem híbrida que combina a estratégia *top-down* com a estratégia *bottom-up*, define-se todos os componentes, tipos e derivados do sangue humano até atingir o limite de componentes demarcado como *proteínas e enzimas* do corpo humano. Daí para frente, onde temos *enzimas, genes e componentes celulares*, faz-se referência a outras ontologias que definem tais termos, como a GO, a ChEBI e a Cell Ontology, não os definindo na ontologia HEMONTO.

- **Questões de competência:** é o método mais comum em metodologias de construção para auxiliar na definição do escopo de cobertura de modo a especificar os requisitos que a ontologia deverá ser capaz de atender e os problemas que a ontologia pode solucionar antes mesmo de sua construção. O desenvolvedor, portanto, deve elaborar um conjunto de questões de competência que sejam anteriores às etapas propriamente de construção de ontologias de forma que elas deixem claro o propósito geral e os objetivos específicos que a ontologia é capaz de atender. Tais questões de competência podem ser definidas tanto em linguagem natural quanto em uma linguagem semi-formal ou ambas, dependendo da expertise do desenvolvedor quanto ao uso de linguagens formais (lógica), e devem revisadas ao longo do desenvolvimento da ontologia, devido a mudanças naturais

nesse processo.

Exemplo: algumas questões de competência da HEMONTO

QC1. Quais os elementos constituintes de uma porção de sangue e em que proporção são encontrados?

QC2. Quais proteínas são encontradas numa porção de plasma sanguíneo?

QC5. Quais diferenças existentes entre os hemocomponentes e hemoderivados do sangue humano?

QC6. Quais os tipos de hemocomponentes e hemoderivados podem ser obtidos a partir de uma unidade de sangue total?

QC10. Quais elementos estão envolvidos num processo de punção venosa?

QC12. Que tipos de células patológicas existem no sangue humano?

Fonte: elaborado pelo autor.

O exemplo completo de preenchimento desses elementos de especificação de uma ontologia é apresentado na documentação formal da ontologia HEMONTO, no capítulo 8 (oito) desta tese.

7.2 Aquisição e Extração de conhecimento (Etapa 2)

Após a especificação da ontologia, passa-se à etapa de aquisição e extração de conhecimento do domínio sob estudo para representação na ontologia a ser construída. Pelo fato desta etapa ser uma atividade independente do processo de desenvolvimento ontológico, a qual é também comum em outras tarefas de representação do conhecimento, algumas metodologias de construção não a incluem como uma etapa específica, como é o caso da *NeOn Methodology*, ou consideram-na “não prescritiva”, caso da *Methontology*. Mesmo com essa característica e diante das dificuldades de se restringir os passos dessa etapa frente à diversidade de aplicações atuais das ontologias, algumas recomendações podem ser sugeridas na presente etapa, conforme explicado nos três próximos tópicos desta seção, os quais incluem os dois passos incluídos nessa etapa e diretrizes da metodologia proposta para a aquisição e extração de conhecimento.

Passo 2.1: Adotar métodos para aquisição e extração de conhecimento

A primeira atividade essencial para a aquisição de conhecimento de um dado domínio é especificar quais métodos ou técnicas a serem usados para a captura do conhecimento. Existe uma literatura vasta sobre esse tema em diversas áreas científicas, que se necessário devem ser consultadas. Entretanto, o foco da presente pesquisa é

sugerir ao desenvolvedor alguns métodos comuns da área de engenharia ontológica, que é também comum a outras áreas afins, e apresentar os métodos específicos da metodologia *OntoForInfoScience*.

De maneira geral, utilizam-se como métodos e técnicas de aquisição e extração de conhecimento para a construção de ontologias, os seguintes:

- (i) *análise informal de textos* em materiais de referência do domínio: permite o estudo e entendimento dos conceitos nos materiais utilizados;
- (ii) *análise formal de textos* nos documentos de referência do domínio: consiste na identificação de estruturas textuais do domínio, tais como definição e afirmação, e o tipo de conhecimento que tais estruturas podem representar na ontologia: conceitos (classes), propriedades, instâncias, relações, etc.
- (iii) *entrevistas estruturadas, semi-estruturadas ou não-estruturadas* com especialistas da área: possibilitam o entendimento de conceitos do domínio por parte do desenvolvedor e a construção de árvores de classificação de conceitos do domínio.
- (iv) *brainstorming ou grupos focais* com especialistas da área: é uma técnica mais colaborativa que a realização de entrevistas para a compreensão de conceitos do domínio, porém pode se tornar ineficiente quando não se obtém consenso sobre os termos da área, necessitando de intervenção contínua do mediador desta atividade.
- (v) *emprego de técnicas estatísticas e/ou métodos linguísticos* que possibilitem a extração de conceitos e relações relevantes do domínio a partir da análise dos textos dos materiais de referência utilizados.
- (vi) *uso de ferramentas automatizadas* que realizam a extração de conceitos e relações relevantes do domínio. De maneira geral, tais ferramentas são usadas a partir de uma abordagem semi-automática, tal que o desenvolvedor ou especialista do domínio avaliem e validem os termos candidatos extraídos de maneira automática.

Como o emprego destes métodos e técnicas é relativamente complexo e seu detalhamento ultrapassa o escopo de uma metodologia de construção de ontologias, referências na literatura devem ser consultadas para maiores detalhes sobre sua utilização.

Passo 2.2: Selecionar os materiais de referência do domínio a serem utilizados

Em paralelo com a escolha dos métodos para aquisição e extração de conhecimento de um dado domínio é fundamental também fazer uma seleção adequada dos materiais de referência a serem utilizados para obtenção do conhecimento deste domínio. O que se quer dizer com uma seleção adequada de referências significa escolher dentre os materiais disponíveis sobre o tema, aqueles que são mais relevantes e representativos do domínio sob estudo. Para tanto, sugere-se consultar especialistas no domínio sobre a relevância dos materiais a serem utilizados. Os parágrafos seguintes concentram-se em explicar a seleção feita dos materiais de referência para a construção da ontologia sobre o sangue humano proposta, que pode servir como um exemplo para desenvolvedores em outros projetos ontológicos.

Os documentos de referência sobre o sangue utilizados na presente pesquisa são textos técnico-científicos da área e podem ser divididos em dois grupos distintos: (i) normas e manuais técnicos de ampla utilização neste domínio de conhecimento e (ii) publicações científicas atuais, contendo os avanços recentes na área. As normas e manuais técnicos sobre hematologia e hemoterapia utilizados foram: (i) **o guia de hemocomponentes do Ministério da Saúde brasileiro** (BRASIL, 2008); (ii) **o padrão terminológico norte-americano ISBT 28 para o sangue** (ICCBBA, 2010); e (iii) **o manual técnico sobre sangue e terapia celular da associação internacional AABB** (AABB, 2005). Quanto às publicações científicas atuais sobre o tema tratado na presente pesquisa foram usados alguns artigos científicos da área e também o livro sobre hematologia clínica Wintrobe's (GREER et al., 2009), em sua 12ª edição. O processo de seleção de tais documentos sobre o domínio do sangue foi realizado com o apoio de profissionais das áreas de hematologia e hemoterapia, avaliando o grau de relevância e representatividade de tais textos neste domínio.

Sobre o “**Guia para o uso de Hemocomponentes**” do Ministério da Saúde (BRASIL, 2008) pode-se afirmar que é o documento de referência mais utilizado por técnicos e profissionais da área em nosso país. Uma série de profissionais, atuantes em organizações especializadas no tema de hemocomponentes, corroboraram para a confecção de tal documento. Entre essas organizações pode-se citar a Agência Nacional de Vigilância Sanitária (ANVISA), a Sociedade Brasileira de Hematologia e Hemoterapia e o Colégio Brasileiro de Hematologia.

Já o **padrão terminológico ISBT 28 para o sangue, terapia celular e descrição dos produtos dos tecidos** documento provê uma terminologia padrão para descrever produtos envolvidos nos processos de transfusão e transplante, sendo

considerado um manual de referência no domínio do sangue e de terapia celular entre profissionais da área e algumas organizações (ICCBBA, 2010).

Ainda no contexto de manuais e normas técnicas sobre o sangue, foi também utilizado o **manual técnico sobre sangue e terapia celular da associação internacional AABB**, descrito na referência *AABB's Technical Manual* 17ª edição (AABB, 2005), que é considerada atualmente uma das fontes de informação sobre banco de sangue, transfusão e terapia celular, mais referenciadas no mundo por profissionais desta área.

Quanto à utilização de publicações científicas recentes na área, destaca-se, inicialmente, o livro-texto sobre hematologia clínica, denominado ***Wintrobe's Clinical Hematology***, em sua 12ª edição (GREER et al., 2009). Tal referência é uma das mais citadas em todo mundo na área de hematologia e possui uma grande tradição histórica, tendo sua primeira publicação no ano de 1942, como registro do trabalho do doutor Maxwell Wintrobe, um ex-estudante de hematologia e um renomado especialista da área.

Ainda com relação às publicações científicas da área utilizadas na construção da ontologia sobre o sangue proposta, foram utilizados livros texto e material de treinamento das seguintes referências: GREER et al. (2009); BARRETT et al. (2009); VAN DE GRAAFF, RHEES e PALMER (2010). Tais referências contém informações fisiológicas sobre os elementos constituintes do sangue humano e sobre os processos realizados para a separação do sangue total e posterior geração dos produtos derivados do sangue, o que também auxiliou na representação desses objetos na ontologia desenvolvida.

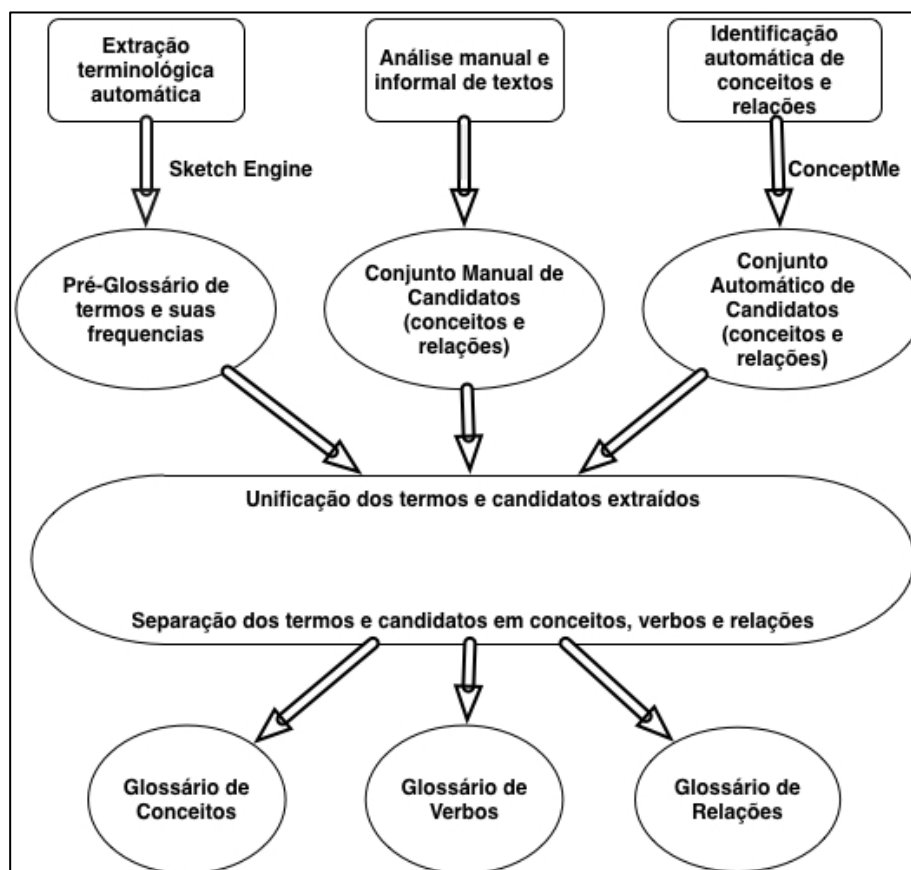
Diretrizes *OntoForInfoScience* para aquisição e extração de conhecimento

A escolha do método ou técnica mais apropriados para aquisição e extração de conhecimento do domínio que se deseja representar em ontologias é uma decisão do desenvolvedor junto com seu grupo de trabalho, que possivelmente deva envolver especialistas do domínio. Entretanto, a metodologia *OntoForInfoScience* estabelece algumas diretrizes que foram adotadas com sucesso no desenvolvimento da ontologia HEMONTO e, assim, servem de suporte ao desenvolvedor nessa etapa de construção. Os parágrafos seguintes deste tópico descrevem tais diretrizes.

O fundamento básico adotado nesse etapa é que a aquisição e extração de conhecimento combine métodos diferentes na exploração das fontes de conhecimento e que seja conduzida de maneira colaborativa, isto é, envolva diferentes visões acerca do conhecimento do domínio. Tal estratégia é assim utilizada na metodologia porque essa etapa de aquisição e extração de conhecimento é compreendida como a etapa de “pré-conceitualização” da ontologia, ou seja, uma etapa preliminar a fase de conceitualização que é capaz de gerar informações a serem utilizadas na etapa

seguinte. A Figura 34, a seguir, representa como foi conduzida tal estratégia, elucidando os recursos e ferramentas utilizadas nesse processo.

Figura 34 - A etapa de aquisição e extração do conhecimento da metodologia *OntoForInfoScience*



Fonte: elaborado pelo autor.

Afim de garantir a colaboração na aquisição e extração de termos do domínio foram adotados três métodos diferentes para realizar tal tarefa:

(i) o primeiro método corresponde ao **processo de extração terminológica** dos termos mais frequentes encontrados nos textos dos materiais de referência utilizados. Tais termos são recuperados automaticamente através da ferramenta on-line *Sketch Engine*⁴⁴, a qual é um analisador linguístico que permite construir listas de frequências das palavras dos textos analisados e obter concordâncias entre elas. Tal ferramenta realiza análises

⁴⁴ Ferramenta disponível em: <http://Sketch Engine.co.uk/>. Acesso em 01 de Abril de 2015.

linguística e morfológica dos textos processados e, além da contagem de frequência dos termos, faz a anotação linguística deles identificando-os como substantivos, verbos, nomes, adjetivos, entre outros. O resultado com o processamento automático através do *Sketch Engine* é uma lista de termos candidatos para a ontologia que forma o chamado **Pré-Glossário de Termos** (círculo mais à esquerda da Figura 34). Tal glossário é um recurso importante para o desenvolvimento de ontologias e seu uso é comum em outras metodologias de construção, tais como a *Methontology* e *Neon*, com diferenças em relação ao processo de geração do mesmo.

(ii) o segundo método empregado envolve o **uso da ferramenta ConceptME para identificação de conceitos e relações candidatos** à ontologia desenvolvida. Nota-se no emprego deste método o uso das palavras “conceito” e “relação” no lugar de “termo”, pois essa ferramenta já trata o conhecimento do domínio não apenas no nível linguístico, mas também no nível conceitual. O ConceptME utiliza algoritmos que identificam padrões linguísticos e estatísticos nos textos analisados para extrair conceitos e relações candidatos à ontologia (SOUSA et al., 2013). Um *verbo* ou um *verbo mais um pronome*, por exemplo, pode indicar uma possível *relação conceitual*, enquanto um *nome* pode indicar um *conceito*. Um padrão identificado na forma <nome1><verbo><nome2> pode indicar uma nova *estrutura conceitual*, tal que o <verbo> representa uma relação conceitual entre os conceitos representados por <nome1> e <nome2>. A extração realizada com o uso do ConceptME gera uma lista de conceitos e relações candidatos à ontologia denominada de **Conjunto Automático de Candidatos**, que será avaliada posteriormente por especialistas do domínio e desenvolvedores na etapa de conceitualização.

(iii) o terceiro e último método empregado é o mais tradicional e corresponde à **análise manual e informal dos textos** dos materiais de referência por parte do(s) desenvolvedor(es), visando identificar conceitos e relações para a ontologia em desenvolvimento. A estratégia de identificação utilizada nesse tipo de análise assemelha-se ao método automático, ou seja, *verbos* podem indicar uma *relação conceitual* e *nomes (substantivos, pronome, adjetivos)* podem representar um *conceito*. A grande diferença entre os dois métodos está no fato de que a análise humana tende a ser mais criteriosa nesse processo, devido à capacidade interpretativa inerente ao ser humano. Por outro lado, métodos automáticos apresentam maior

produtividade de análise ao processar maior quantidade de dados em um tempo bem menor que seres humanos. Combinar as duas abordagens é a melhor solução. No que diz respeito à metodologia proposta, a análise textual produzirá outra lista de conceitos e relações candidatos à ontologia, denominada de **Conjunto Manual de Candidatos**, que também deverá ser avaliada na etapa seguinte por especialistas do domínio.

A tarefa seguinte nesta etapa é a **unificação dos resultados dos três conjuntos obtidos** (pré-glossário de termos, conjunto automático de candidatos e conjunto manual de candidatos) pelos diferentes métodos empregados e, paralelamente, a **separação dos termos resultantes** em conceitos candidatos, verbos e possíveis relações formadas pela tupla conceito-verbo-conceito. As atividades executadas na tarefa de unificação incluem eliminar os rótulos (nomes) repetidos e identificar possíveis sinônimos entre os termos. Já a separação dos termos resultantes gera três novos conjuntos que servem como parâmetros de entrada na etapa de conceitualização: (i) **Glossário de Conceitos**: engloba substantivos, pronomes e adjetivos compreendidos como conceitos do domínio e candidatos a se tornarem classes na ontologia; (ii) **Glossário de Verbos**: reúne todos os verbos extraídos dos textos de referência que são candidatos a se tornarem relações na ontologia; e (iii) **Glossário de Relações**: conjunto formado por todos os padrões de relacionamento extraídos dos textos na forma <nome> <verbo> <nome>, tal que os nomes podem ser substantivos, pronomes ou adjetivos e o verbo representa um relacionamento existente entre os nomes.

7.3 Conceitualização (Etapa 3)

A fase de conceitualização é uma das mais importantes para o sucesso de uma ontologia, pois é nela que se realizam atividades como a identificação, análise e negociação dos conceitos de um determinado domínio, em um determinado contexto de aplicação, e que serão incluídos como classes na ontologia (PEREIRA e SOARES, 2008). Além disso, é nessa fase que ocorre a organização e estruturação do conhecimento por meio de artefatos de representação, que irão se transformar em relações, propriedades e restrições de uma ontologia. O resultado final dessa etapa é um modelo conceitual do domínio que englobe conceitos, suas relações e suas propriedades, tal que esse modelo atenda às especificações para a ontologia realizadas nas etapas anteriores. Considerando

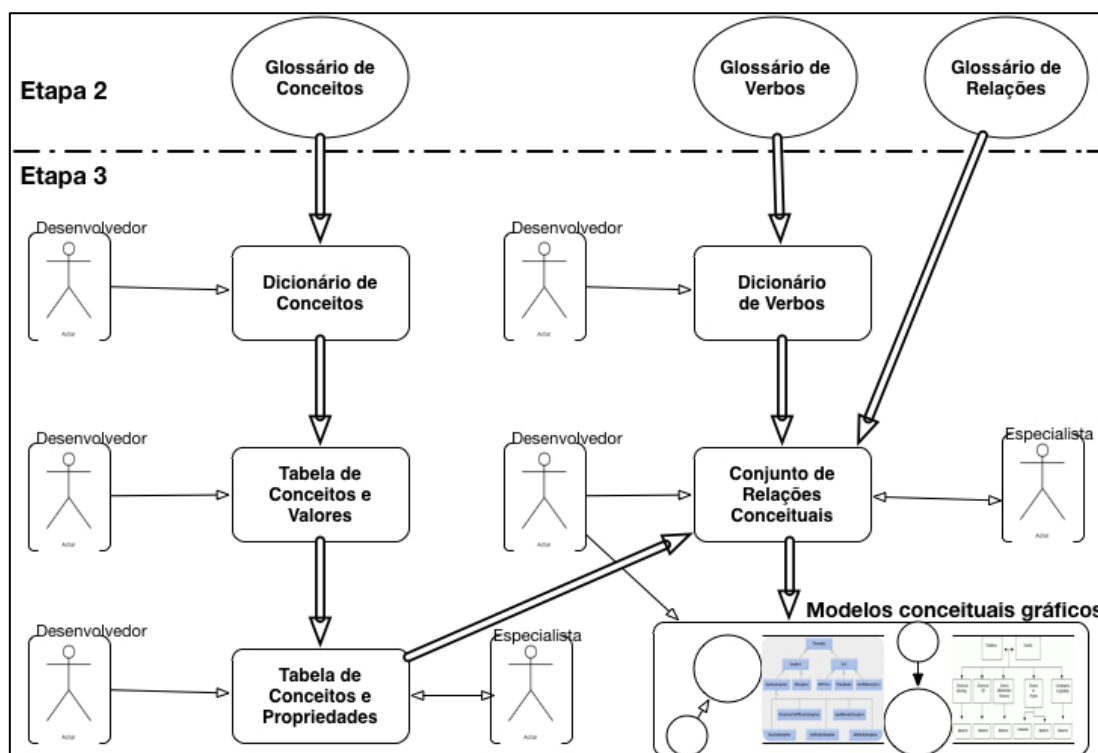
a grande importância da conceitualização, a metodologia *OntoForInfoScience* dá uma atenção especial a essa etapa.

De maneira geral, as metodologias de construção atuais sempre incluem como uma de suas etapas a conceitualização. A *Methontology* (FERNÁNDEZ et al., 1997), por exemplo, prescreve nessa etapa a identificação de conceitos, instâncias, relações verbais e propriedades do domínio, com cada termo tendo uma representação intermediária, tais como árvores de classificação de conceitos e tabelas de atributos de classes. O método 101 (NOY e GUINNESS, 2001) não denomina uma etapa especificamente como conceitualização, mas realiza-a através das etapas de enumeração de termos, consideração do reuso de termos de outras ontologias, definição de classes e organização das classes em uma taxonomia. A *NeOn Methodology* (SUÁREZ-FIGUEROA, 2008) também não denomina uma etapa específica de conceitualização, realizando-a ao longo da cadeia de atividades *NeOn*, que incluem as etapas de reuso de recursos não-ontológicos, reuso dos recursos ontológicos e reuso dos padrões de desenvolvimento ontológico (*Ontology Design Patterns – ODP's*), em uma abordagem mais direcionada para a área de Engenharia de Software.

A proposta da metodologia *OntoForInfoScience*, nessa etapa, aproxima-se mais da abordagem adotada na *Methontology*, diferenciando desta por trazer maiores detalhes do processo de conceitualização.

Antes de descrever os passos e recursos dessa etapa é importante ressaltar dois aspectos fundamentais levados em consideração pela *OntoForInfoScience*: (i) a utilização dos recursos da etapa anterior de aquisição e extração do conhecimento; e (ii) a importância do envolvimento dos especialistas na conceitualização do domínio. Sobre esse último aspecto, sugere-se ao processo de conceitualização de qualquer ontologia ser conduzido de forma a envolver os atores/especialistas do domínio em tarefas colaborativas, afim de compartilharem, negociarem e representarem um modelo conceitual comum entre as partes envolvidas.

O esquema de conceitualização da metodologia *OntoForInfoScience* é apresentado na Figura 35, a seguir, destacando os atores dessa etapa, recursos utilizados e os produtos finais a serem utilizados nas etapas posteriores.

Figura 35 - A etapa de conceitualização da metodologia *OntoForInfoScience*

Fonte: elaborado pelo autor.

A primeira observação a ser feita sobre o esquema de conceitualização da Figura 35 é a utilização dos três conjuntos de termos (glossário de conceitos, glossário de verbos e glossário de relações) obtidos da etapa de extração e aquisição de conhecimento, que estão representados como círculos para lhes diferenciar dos demais conjuntos da etapa de conceitualização. Nos parágrafos seguintes, descreve-se, detalhadamente, os passos desta etapa representados na Figura 35.

Passo 3.1: Construir o Dicionário de Conceitos

O processo de conceitualização se inicia quando o desenvolvedor da ontologia transforma o Glossário de Conceitos, que inclui apenas uma lista de conceitos do domínio extraídos dos textos analisados, em um **Dicionário de Conceitos**, o qual inclui o *conceito* associado a sua *definição textual* obtida a partir de alguma fonte de informação do domínio. Nesse passo, quando os conceitos são acrescidos de sua definição é comum a identificação de *sinônimos*, a qual é melhor realizada nessa etapa do que na anterior, porque agora as informações do domínio passam a ser tratadas a *nível semântico*. Quando sinônimos são identificados, o desenvolvedor deve eliminar a entrada (célula) no Dicionário de Conceitos e anotar o sinônimo ao lado do termo preferencial. As Tabelas 25 e 26, a seguir, apresentam exemplos que ilustram esse processo de transformação do Glossário

de Conceitos (Tabela 25) em um Dicionário de Conceitos (Tabela 26), a partir do desenvolvimento da ontologia HEMONTO.

Tabela 25 - Parte do Glossário de Conceitos da HEMONTO

ID	Termo (Conceito)	Frequência
1	body substance	56
2	blood	221
3	portion of blood	102
4	porção de sangue	28
5	sangue	46
6	arterial blood	47
7	entire portion of blood	15
8	maximum portion of blood	3
9	whole portion of blood	163
10	plasma	178
11	blood plasma	56
12	anticoagulants	68
13	anticoagulant	69
14	cell	87
15	blood cell	103
16	hemal cell	35

Fonte: elaborado pelo autor.

Tabela 26 - Parte do Dicionário de Conceitos da HEMONTO

ID	Conceito	Sinônimos	Definição
1	body substance	-	Material anatomical entity in a gaseous, liquid, semisolid or solid state, with or without admixture of cells and biological macromolecules. A portion of body substance is produced by anatomical structures or derived from inhaled or ingested substances that have been modified by anatomical structures.
2	blood	portion of blood; porção de sangue; sangue	Portion of body substance which has as its parts plasma and blood cells. Líquido principal do corpo humano composto de uma porção de plasma e dos chamados elementos formados (plasma, água, hemácias) e que representa cerca de 8% do peso corporal.
3	arterial blood	-	Arterial blood is the oxygenated blood in the circulatory system found in the lungs, the left chambers of the heart, and in the arteries. It is bright red in color, while venous blood is dark red in color (but looks purple through the opaque skin). It is the contralateral

			term to venous blood.
4	whole portion of blood	entire portion of blood; maximum portion of blood	Portion of blood collected into an anticoagulant and not further processed. A unit of blood collected into an anticoagulant and not further processed unless otherwise specified.
5	plasma	blood plasma	Portion of body substance corresponding to the liquid component of blood in which blood cells are immersed. Unless otherwise specified the product has been obtained from Whole Blood and frozen.
6	anticoagulant	anticoagulants	An agent that prevents blood clotting.
7	cell	-	Anatomical structure which has as its boundary the external surface of a maximally connected plasma membrane.
8	blood cell	hemal cell	General anatomical term which refers to nucleated and non-nucleated differentiated hemal cells.

Fonte: elaborado pelo autor.

Passo 3.2: Elaborar a Tabela de Conceitos e Valores

A tarefa seguinte da etapa é determinar *valores possíveis* para cada um dos conceitos contidos no Dicionário de Conceitos. *Valores possíveis* são exemplos, casos reais ou ocorrências de um determinado conceito do domínio. Pode-se afirmar também que os *valores possíveis* são equivalentes aos *particulares* de um determinado universal e que poderão se tornar *instâncias* das classe da ontologia.

Em uma ontologia da árvore genealógica de uma determinada família, por exemplo, podemos ter como conceitos Avô, Avó, Pai, Mãe, Idade, Sexo e Cor, e como seus valores possíveis os seguintes: Avô = {João}, Avó = {Maria}, Pai = {José}, Mãe = {Ana Maria}, Idade = {1 a 120 anos}, Sexo = {M, F} e Cor = {Branca, Negra, Parda, Mulata, Mestiça, Amarela}.

Na realização desta tarefa, o desenvolvedor deve tomar o Dicionário de Conceitos e acrescentar a cada entrada os *valores possíveis* que o conceito pode assumir, que poderão indicar *instâncias* para a ontologia ou até mesmo *classes*, dependendo do nível de especialização do domínio tratado. O Dicionário de Conceitos acrescido de valores possíveis do domínio é chamado na metodologia proposta de **Tabela de Conceitos e Valores**. Uma parte da tabela de conceitos e valores da ontologia HEMONTO é apresentada na Tabela 27, a seguir, como forma de ilustrar esta tarefa da metodologia.

Tabela 27 - Parte da Tabela de Conceitos e Valores da HEMONTO

ID	Conceito	Sinônimos	Definição	Valores
1	body	-	Material anatomical entity in a gaseous, liquid, semisolid or solid state, with or	portion of saliva, portion of semen,

	substance		without admixture of cells and biological macromolecules. A portion of body substance is produced by anatomical structures or derived from inhaled or ingested substances that have been modified by anatomical structures.	portion of cerebrospinal fluid, portion of respiratory air, portion of urine, portion of plasma, portion of lymph.
2	blood	portion of blood; porção de sangue; sangue	Portion of body substance which has as its parts plasma and blood cells. Líquido principal do corpo humano composto de uma porção de plasma e dos chamados elementos formados (plasma, água, hemácias) e que representa cerca de 8% do peso corporal.	arterial blood, venous blood.
3	arterial blood	-	Arterial blood is the oxygenated blood in the circulatory system found in the lungs, the left chambers of the heart, and in the arteries. It is bright red in color, while venous blood is dark red in color (but looks purple through the opaque skin). It is the contralateral term to venous blood.	-
4	whole portion of blood	entire portion of blood; maximum portion of blood	Portion of blood collected into an anticoagulant and not further processed. A unit of blood collected into an anticoagulant and not further processed unless otherwise specified.	-
5	plasma	blood plasma	Portion of body substance corresponding to the liquid component of blood in which blood cells are immersed. Unless otherwise specified the product has been obtained from Whole Blood and frozen.	fresh frozen plasma, plasm of 24 hours, pooled plasma, pooled fresh frozen plasma.
6	anticoagulant	anticoagulants	An agent that prevents blood clotting.	aspirin, enoxaparin, heparin, warfarin, clopidogrel, ticlopidine.
7	cell	-	Anatomical structure which has as its boundary the external surface of a maximally connected plasma membrane.	lymphocyte, fibroblast, erythrocyte, neuron.
8	blood cell	hemal cell	General anatomical term which refers to nucleated and non-nucleated differentiated hemal cells.	red blood cell (erythrocyte), white blood cell (leukocyte), thrombocyte, echinocyte, reticulocyte

Fonte: elaborado pelo autor.

A nível conceitual, a enumeração dos valores possíveis de cada conceito é suficiente para a realização deste passo da metodologia. Entretanto, a nível ontológico, torna-se necessário aplicar alguns princípios na decisão ontológica de definir se um determinado conceito é uma classe ou instância na ontologia. Além disso, a criação de uma instância de uma determinada classe deve atender as restrições (propriedades de dados) desta classe. Esses tópicos serão vistos na etapa de formalização da ontologia.

Passo 3.3: Elaborar a Tabela de Conceitos e Propriedades

O passo seguinte no processo de conceitualização corresponde à definição de *propriedades específicas* para cada conceito ou classe da tabela anterior, transformando-a na **Tabela de Conceitos e Propriedades**. Propriedades de um conceito ou classe são características que permitem diferenciar um conceito do outro no domínio sob estudo e são uma herança dos trabalhos de Aristóteles que tratava as propriedades sob o nome de “differentia”. Incluir as propriedades das classes em qualquer desenvolvimento ontológico é indispensável para uma representação fidedigna do domínio sob estudo. Para que a ontologia consiga responder as questões de competência relativas a ela, as propriedades das classes precisam estar definidas.

Nesta etapa da metodologia, que trata o conhecimento a nível conceitual e ainda não ontológico, a definição das propriedades das classes pode ser feita de forma descritiva em formato textual, não sendo necessário especificar explicitamente propriedades tais como: tipo de valores das classes (numérico, string, booleano), domínio e imagem, cardinalidade. Essas propriedades específicas de uma ontologia deverão ser definidas na etapa de formalização da ontologia, usando um editor de ontologias para tal.

O Quadro 2 e a Tabela 28, a seguir, apresentam alguns exemplos de propriedades de conceitos no domínio do sangue extraídas dos materiais de referência utilizados nesta pesquisa, as quais serão incluídas na Tabela de Conceitos e Propriedades desta etapa de conceitualização da metodologia.

Quadro 2 – Definição e forma de obtenção do componente Concentrado de Hemácias

<p>O concentrado de hemácias (CH) é obtido por meio da centrifugação de uma bolsa de sangue total (ST) e da remoção da maior parte do plasma. Seu volume varia entre 220ml e 280ml.</p> <p>Assim como o ST, o concentrado de hemácias deve ser mantido entre 2°C e 6°C e sua validade varia entre 35 e 42 dias, dependendo da solução conservadora. Os concentrados de hemácias sem solução aditiva devem ter hematócrito entre 65% e 80%. No caso de bolsas com solução aditiva, o hematócrito pode variar de 50% a 70%.</p>

Fonte: Guia para uso de hemocomponentes (BRASIL, 2008).

Tabela 28 – Alguns hemocomponentes do sangue humano e seus usos indicados

Blood Components and Indications for Use			
Component	Composition	Volume	Indications and Expected Benefit
Whole blood	RBC and plasma (approx. Hct, 40%); WBCs; platelets	500 ml	To increase red cell mass and plasma volume (plasma deficient in labile clotting factors V and VIII); for hypovolemic anemia, massive transfusion, or exchange transfusion in neonates
RBCs, adenine-saline added	RBC and 100 ml of additive solution (approx. Hct, 60%);	330 ml	To increase red cell mass in symptomatic anemia; 10 ml/kg raises Hct by 8%

	WBCs; platelets; little plasma		
RBCs, leukocytes reduced (prepared by filtration)	>85% original volume of RBCs; <5 × 10 ⁶ WBCs	>85% of original volume	To increase red cell mass; <5 × 10 ⁶ WBCs to decrease the likelihood of febrile reactions, immunization to leukocytes (HLA antigens), or CMV transmission
Platelet concentrates	Platelets (>5.5 × 10 ¹⁰ /unit); RBCs; WBCs; plasma	50 ml	Bleeding due to thrombocytopenia or thrombocytopathy; 1 unit/10 kg raises platelet count by 17–50 × 10 ⁹ /L

Fonte: Wintrobe's Clinical Hematology (GREER et al., 2009).

Na descrição textual do Concentrado de Hemácias (Quadro 2) é possível perceber algumas propriedades intrínsecas deste hemocomponente tais como: (i) seu volume entre 220 ml e 280 ml; (ii) sua temperatura de armazenamento (2 e 6 graus); (iii) sua validade entre 35 e 42 dias; e (iv) sua concentração de hematócitos variando com ou sem solução aditiva. De maneira similar, na tabela de hemocomponentes e suas indicações de uso (Tabela 28), identificam-se propriedades de cada componente, que incluem sua composição, volume e especificações nas indicações de uso. No desenvolvimento da ontologia HEMONTO essas informações foram recuperadas de documentos da área e registradas na Tabela de Conceitos e Propriedades nesta etapa de conceitualização. Tal tabela corresponde à Tabela de Conceitos e Valores alterada para inclusão das propriedades de cada conceito. Como não se tem mais nenhuma tabela a construir sobre conceitos nesta metodologia, a Tabela de Conceitos e Propriedades é apresentada aos especialistas do domínio para validação de seu conteúdo, conforme Figura 35, apresentada anteriormente. Após essa validação, que possivelmente gerará alguma alteração no conteúdo da tabela, tem-se um dos produtos finais dessa etapa de conceitualização – a própria tabela alterada – que será utilizada na fase seguinte do processo de desenvolvimento ontológico.

Passo 3.4: Construir um Dicionário de Verbos

Uma parte da conceitualização proposta na metodologia corresponde a essa definição de conceitos e seus atributos, já a outra parte corresponde à definição das relações conceituais do domínio (lado direito da Figura 35), a qual é descrita nos próximos parágrafos. Essas duas partes são complementares e podem ser conduzidas paralelamente pelo desenvolvedor. Elas apenas estão separadas no esquema de conceitualização da Figura 35 como forma de detalhar os passos realizados em cada parte.

O Glossário de Verbos da etapa anterior da metodologia é o parâmetro de entrada para definição de relações conceituais do domínio. Como ocorre com os conceitos, o desenvolvedor deve transformar o Glossário de Verbos em um **Dicionário de Verbos** do

domínio, acrescentando a cada verbo extraído dos textos analisados uma definição informal (descritiva) do significado do verbo no domínio sob estudo. Essa definição para cada verbo auxilia também na identificação de verbos que não possuem o mesmo rótulo (nome) mas possuem o mesmo significado, ou seja, são verbos sinônimos nesse contexto. Tais verbos devem ser representados em uma única entrada do Dicionários de Verbos, com o rótulo mais comum (preferencial) anotado com seus sinônimos. Como forma de ilustra a construção de tal dicionário, a Tabela 29, a seguir, apresenta parte do conteúdo do Dicionário de Verbos da HEMONTO.

Tabela 29 - Parte do Dicionário de Verbos da HEMONTO

ID	Verbo	Sinônimos	Definição	Exemplo de uso
1	is	corresponding to	Indica uma relação de gênero-espécie entre dois conceitos (generalização-especialização), que é também usada para definir um conceito.	Portion of plasma is the portion of body substance corresponding to the liquid component of blood in which blood cells are immersed.
2	part of	composed	Relação entre dois conceitos tal que um conceito é uma parte que compõe ou integra o outro conceito.	A blood collection process in which some part of the donation is returned to the donor. Blood layer composed of leukocytes and platelets, obtained after centrifugation of portion of whole blood.
3	prepared by	-	Representa a relação entre um conceito e seu processo de geração ou obtenção.	A product prepared by combining two or more single units of Platelets into one container.
4	collected into	-	Indica uma relação de extração (coleta) de um conceito através de algum meio ou processo.	A unit of blood collected into an anticoagulant and not further processed unless otherwise specified.
5	remove		Representa a relação de retirada ou remoção de um conceito que é parte de outro conceito.	The treatment of an apheresis cellular product using a compatible solution to remove most of the plasma proteins.
6	prevent	-	Indica uma relação na qual um conceito-agente evita ou previne a realização de um processo.	An agent that prevents blood clotting.
7	wash	washing	Representa um processo no qual se usa uma solução (o conceito) para produzir ou prevenir um processo.	The removal of glycerol by washing from an apheresis product.
8	has been frozen	-	Indica uma relação mandatória, na qual um conceito específico deve ser congelado para manter suas propriedades.	Plasma that has been frozen by a process and to a temperature that will maintain the activity of labile protein fractions.

Fonte: elaborado pelo autor.

Passo 3.5: Desenvolver Modelos Conceituais a partir do Conjunto de Relações

O Dicionário de Verbos contém apenas a lista de verbos candidatos às relações conceituais do domínio e suas definições, e a Tabela de Conceitos e Propriedades armazena os conceitos do domínio e suas restrições. Sendo assim, no passo seguinte desta etapa, o desenvolvedor deve correlacionar verbos candidatos a relações e conceitos do domínio para representar os relacionamentos identificados na realidade. Parte dessa tarefa já foi realizada automaticamente na etapa de extração do conhecimento (fase de pré-conceitualização) com a elaboração do Glossário de Relações, que relaciona os termos do domínio. Entretanto, tal extração automática de padrões identificados como relacionamentos precisam de ser analisados um a um manualmente, validando as relações sugeridas.

Além dessa análise, novas relações entre conceitos vão sendo criadas pelo desenvolvedor à medida que se consulta os conjuntos gerados anteriormente e os textos dos materiais de referência. Esse passo da conceitualização é representado do lado direito da Figura 35, apresentada anteriormente, no qual o desenvolvedor da ontologia deve elaborar um **Conjunto de Relações Conceituais** do domínio com base nos conjuntos já gerados: Dicionário de Verbos, Glossário de Relações e Tabela de Conceitos e Valores. Algumas regras básicas para composição desse Conjunto de Relações Conceituais são as seguintes:

- Um item na forma <nome1> <verbo> <nome2> do Glossário de Relações apenas pode ser adicionado como um item do Conjunto de Relações conceituais quando atender os seguintes critérios:
 - i) <nome1> deve estar contido na Tabela de Conceitos e Propriedades, seja como nome (rótulo) preferencial ou anotado como sinônimo. Caso <nome1> esteja anotado como sinônimo, ele deve ser substituído pelo nome preferencial na relação formada.
 - ii) <nome2> deve estar contido na Tabela de Conceitos e Propriedades, seja como nome (rótulo) preferencial ou anotado como sinônimo. Caso <nome2> esteja anotado como sinônimo, ele deve ser substituído pelo nome preferencial na relação formada.
 - iii) <verbo> deve estar contido no Dicionário de Verbos, seja como verbo (rótulo) preferencial ou anotado como sinônimo. Caso <verbo> esteja anotado como sinônimo, ele deve ser substituído pelo verbo preferencial na relação formada.

Para realizar esse procedimento, o desenvolvedor deve contar com o apoio de especialistas no domínio e utilizar uma ferramenta de modelagem para representar o Conjunto de Relações Conceituais em um formato gráfico de **Modelos Conceituais**, uma vez que relações conceituais descritas apenas textualmente não tem a capacidade de representar relacionamentos no contexto do domínio. Portanto, um dos artefatos finais gerado nessa etapa de conceitualização é o Conjunto de Modelos Conceituais do domínio, representados através de grafos conceituais, taxonomias, hierarquias e outras estruturas de representação gráfica, conforme apresentado na Figura 35, anteriormente.

Ainda sobre modelos conceituais, é importante destacar que há duas maneiras de conduzir seu desenvolvimento: (i) o próprio desenvolvedor pode modelar os grafos conceituais, taxonomias, hierarquias e validá-los posteriormente com os especialistas da área; ou (ii) o desenvolvedor pode disponibilizar a ferramenta de modelagem, o dicionário de verbos e a tabela de conceitos aos especialistas do domínio, para que eles próprios possam elaborar as estruturas de representação gráfica. Nessa segunda opção, pode ser necessário um treinamento prévio dos especialistas no uso dos recursos de modelagem utilizados no projeto.

No que diz respeito à construção da HEMONTO, optou-se pelo desenvolvimento de modelos conceituais por parte do autor desta pesquisa e desenvolvedor da ontologia, com validação pelos especialistas do domínio. As ferramentas de modelagem utilizadas na conceitualização do domínio do sangue foram o ConceptMe⁴⁵, o DiagramEditor⁴⁶ e o OmniGraffle⁴⁷.

As estruturas de representação gráficas desenvolvidas na construção da HEMONTO englobaram as **taxonomias**, as **partonomias** e **estruturas de representação gerais ou mistas**, tal que ambas se diferem pelos tipos de relação conceitual utilizadas na estrutura. **Taxonomias** são estruturas hierárquicas onde conceitos ou classes do domínio são interrelacionados unicamente através da relação conceitual *is_a* (*é_um*); nas **partonomias**, a única relação conceitual utilizada é a *part_of* (*parte_todo*); e nas **estruturas gerais ou mistas**, *diferentes tipos* de relação conceitual são utilizadas para ligar os conceitos do domínio tratado. As Figuras 36 e 37, a seguir, ilustram dois exemplos dessas estruturas criadas na etapa de conceitualização da ontologia HEMONTO. A Figura 36 apresenta a partonomia dos produtos derivados do sangue, obtidos através dos processos de separação do sangue total e que geram seus hemocomponentes e hemoderivados, já a Figura 37 descreve, em uma estrutura de representação geral ou

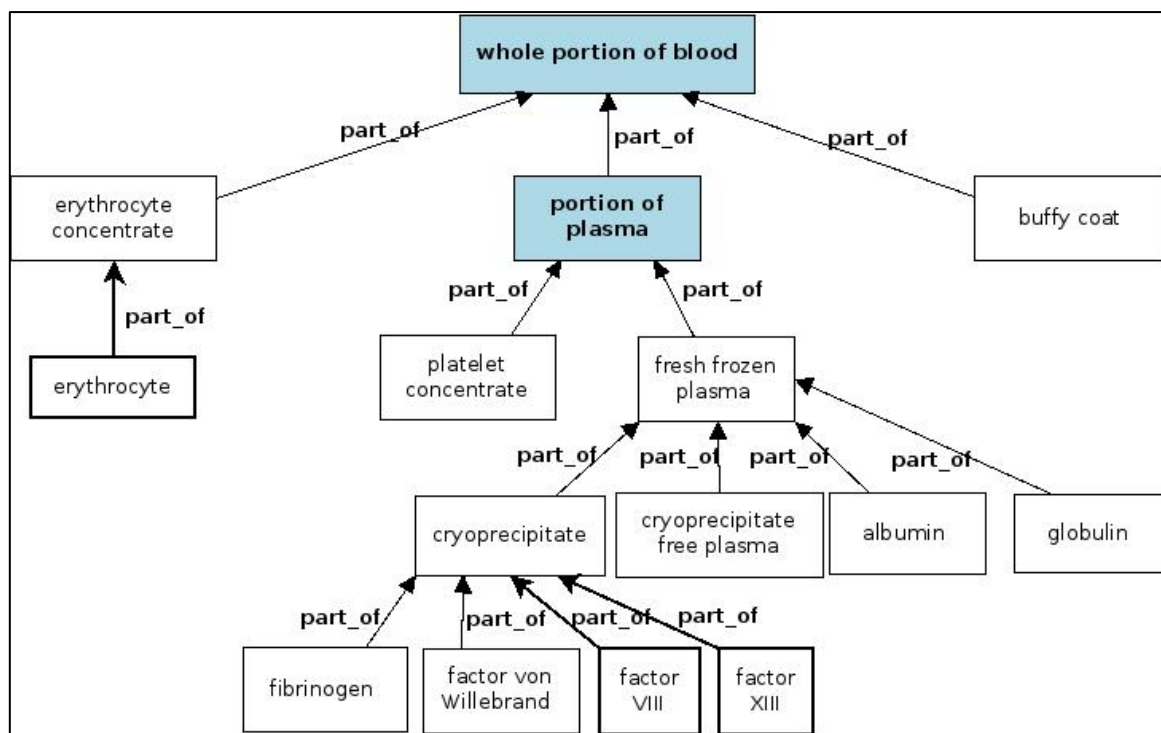
⁴⁵ Disponível em: http://www.conceptme.pt/conceptme/index.php/Main_Page. Acesso em: 10 de Novembro de 2014.

⁴⁶ Disponível em: <http://dia-installer.de/>. Acesso em: 22 de Junho de 2014.

⁴⁷ Disponível em: <https://www.omnigroup.com/omnigraffle>. Acesso em: 05 de Janeiro de 2015.

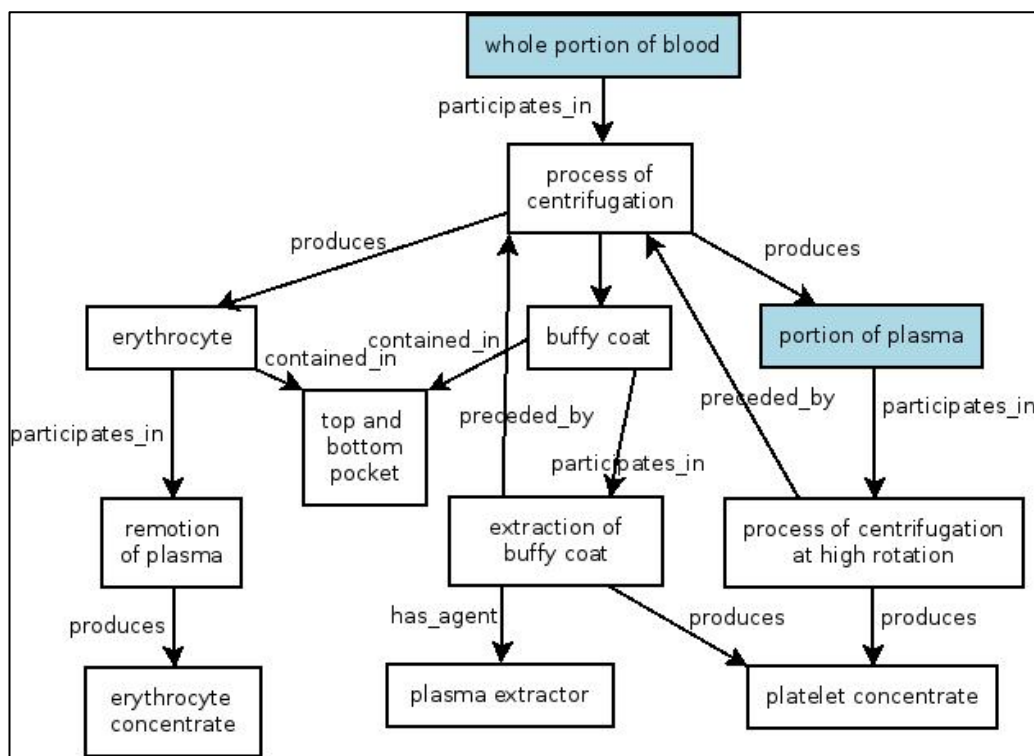
mista, os processos de obtenção de dois hemocomponentes do sangue: o *Concentrado de Hemácias (erythrocyte concentrate)* e o *Concentrado de Plaquetas (platelet concentrate)*.

Figura 36 - Partonomia dos produtos derivados do sangue.



Fonte: elaborado pelo autor.

Figura 37 - Processos de obtenção dos hemocomponentes Concentrado de Hemácias e Concentrado de Plaquetas



Fonte: elaborado pelo autor.

7.4 Fundamentação ontológica (Etapa 4)

Ao longo da presente pesquisa demonstrou-se a importância das ontologias de fundamentação, como estratégia na área de engenharia ontológica para solucionar alguns problemas das atuais terminologias e ontologias biomédicas. Ceusters et al. (2003), por exemplo, afirma ser as ontologias de fundamentação uma tentativa de quebrar com o paradigma de construção de ontologias com o mínimo compromisso ontológico.

Tomando como base esse argumento a favor do uso de ontologias de fundamentação, cuja prática é, geralmente, bem aceita e incentivada na área (DEGEN et al., 2001; MASOLO et al., 2003; GRENON e SMITH, 2004; HERRE, 2010; SMITH, 2005; BITTNER e DONNELLY, 2007), a metodologia *OntoForInfoScience* sugere, como uma das etapas no desenvolvimento ontológico, a especificação de uma fundamentação teórica-filosófica que guie tal desenvolvimento. Essa especificação deve incluir a abordagem filosófica adotada como base para a construção ontológica e a possibilidade de uso de uma ontologia de fundamentação nesse processo de construção.

A adoção de uma abordagem filosófica para ontologias, por vezes, é questionável no ponto de vista de alguns ontologistas, especialmente, aqueles que

pretendem construir ontologias leves com finalidades unicamente computacionais. Entretanto, defende-se na metodologia *OntoForInfoScience* que a adoção de princípios filosóficos de uma determinada linha auxilia nas decisões ontológicas e permite justificá-las pelo seu embasamento teórico.

A presente etapa da metodologia proposta foi denominada de “fundamentação da ontologia” e precede a etapa de formalização da ontologia, na qual o conhecimento do domínio é representado em recursos ontológicos. Em um projeto ontológico de menor porte, onde se deseja construir uma ontologia leve de domínio ou mesmo uma subontologia pode não ser necessária a realização desta etapa de fundamentação. Entretanto, na grande maioria dos projetos ontológicos, essa é uma etapa essencialmente importante para cumprir com os propósitos da ontologia desenvolvida.

Passo 4.1: Pesquisar e escolher ontologias de fundamentação

O primeiro passo, na etapa de fundamentação, consiste em considerar o uso e aplicação de uma ou mais ontologias de fundamentação no desenvolvimento da ontologia pretendida, como forma de assegurar o compromisso ontológico do projeto e tornar mais claras as decisões tomadas pelo desenvolvedor na representação do domínio sob estudo. Smith (2005) e Bittner e Donnely (2007), por exemplo, apontam que o uso dos princípios filosóficos das ontologias de fundamentação é uma forma de manter explícitas, o quanto possível, as definições formais de relações e classes incluídas na ontologia, tornando mais transparente as decisões ontológicas tomadas no processo de construção da ontologia.

Nesse sentido, uma das tarefas do desenvolvedor é a de pesquisar por ontologias de fundamentação relevantes na literatura da área e, de acordo com o contexto de aplicação e os objetivos propostos na ontologia em construção, escolher as ontologias compatíveis e mais apropriadas ao projeto ontológico desenvolvido. Tal pesquisa pode ser feita em uma simples consulta pelos termos “*foundational ontology*” ou “*top-level ontology*” pela *web*, considerando a relevância dos resultados da busca. Uma revisão de literatura sobre tais tipos de ontologia já foi apresentada nesta pesquisa e também pode servir como instrumento de consulta.

Ao optar por uma ou mais ontologias de fundamentação pesquisadas, o desenvolvedor, implicitamente, está também adotando em seu projeto uma abordagem (ou corrente) teórica-filosófica que lhe guiará em suas decisões ontológicas na modelagem e representação do domínio sob estudo. Isto ocorre porque cada ontologia de fundamentação descreve as entidades (“coisas”) do mundo a partir de uma linha de pensamento filosófico que esta segue, sendo que muitas ontologias seguem uma mesma corrente filosófica.

Ao desenvolvedor cabe outra tarefa, que é a de tornar explícita na documentação da ontologia a abordagem filosófica adotada. Essa medida serve para justificar as decisões ontológicas tomadas pelo desenvolvedor no processo de construção e compatibilizar a abordagem adotada com as ontologias de fundamentação escolhidas. Nesse último item é importante mencionar que a metodologia *OntoForInfoScience* recomenda que, na seleção de mais de uma ontologia de fundamentação, ambas ontologias escolhidas sigam uma única e comum abordagem filosófica, porque em caso contrário, pode-se incorrer em erros de justificativas das decisões tomadas, surgir questões filosóficas não solucionáveis e, por fim, o projeto ontológico se tornar inconsistente.

No que se refere ao desenvolvimento da ontologia HEMONTO foram selecionadas como: a **Basic Formal Ontology (BFO)** como ontologia de fundamentação-base para a construção da ontologia do sangue e a **Relation Ontology (RO)** como ontologia de fundamentação para as relações desta ontologia. Ambas ontologias estão inseridas no âmbito do **realismo científico** (POPPER, 1963; BUNGE, 1974; CHURCHLAND, 1985; SMITH e CEUSTERS, 2010), mais especificamente o **realismo ontológico** (GRENON e SMITH, 2004). O **realismo científico** defende o argumento de que as teorias e os modelos científicos possuam valor ontológico, isto é, eles devem ser construídos em direção à verdade, buscando captar, numa maior aproximação possível, “o que realmente existe” no mundo. Nessa abordagem, a HEMONTO é compreendida como uma teoria ontológica do sangue humano e produtos ou construções exclusivamente da mente humana não devem ser incluídos em tal ontologia desenvolvida.

A opção pelo uso das ontologias BFO e RO e o realismo ontológico justifica-se pelas características do domínio do sangue humano, que para sua descrição necessita da representação de entidades reais, tais como os produtos constituintes do sangue e seus aspectos fisiológicos. A BFO e a RO são ontologias de fundamentação direcionadas, especialmente, para a representação de domínios biomédicos e suas relevâncias na área são grandes, tal que muitas outras ontologias biomédicas de domínio foram e estão sendo desenvolvidas sob seus princípios, como por exemplo, a Cell Ontology (CL) e a Protein Ontology (PRO), utilizadas nesta pesquisa.

Passo 4.2: Aplicar a ontologia de fundamentação no desenvolvimento

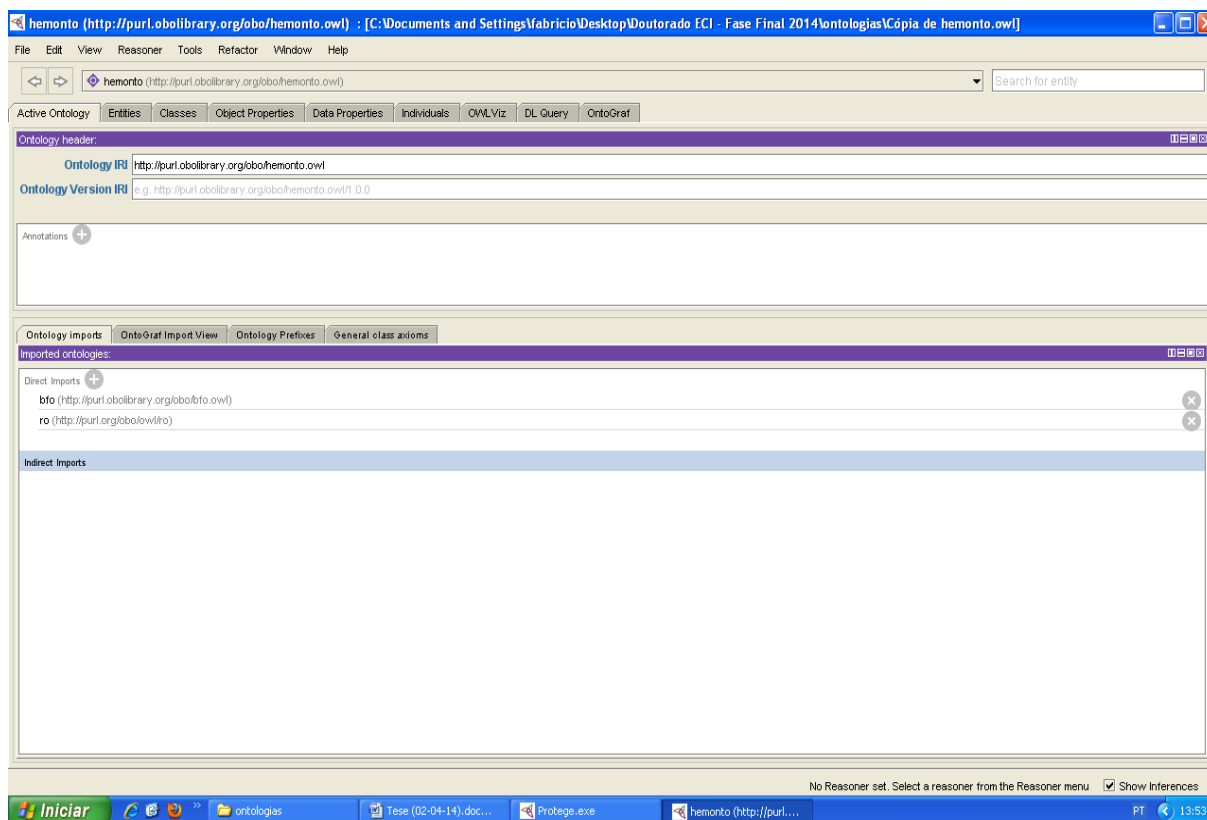
Após a escolha da(s) ontologia(s) de fundamentação e abordagem filosófica, o passo seguinte da etapa consiste na utilização da ontologia de fundamentação e seus princípios no processo de representação da ontologia de domínio em desenvolvimento.

Ontologias de fundamentação ou de alto nível são ontologias que descrevem as entidades mais gerais dos diversos domínios do conhecimento, incluindo classes como: *processo, objeto, qualidade, função, entidade imaterial*, entre outras. Em um projeto ontológico, a decisão de iniciar a construção de uma ontologia a partir de uma ontologia de fundamentação é uma medida apropriada e recomendada na literatura e pela metodologia proposta, uma vez que muitas das decisões ontológicas a serem tomadas no projeto já foram tomadas na ontologia de fundamentação e podem ser seguidas na ontologia em desenvolvimento.

Assim, o processo de representação da ontologia de domínio inicia-se com o carregamento da ontologia de fundamentação em algum editor de ontologias e continua com a inclusão de *classes, relações e propriedades específicas* do domínio sob estudo nessa estrutura. Assim, cabe ao desenvolvedor escolher um editor de ontologias apropriado (em geral, o Protégé é o mais utilizado) e importar para dentro desse editor o arquivo OWL da ontologia de fundamentação afim de iniciar o desenvolvimento. A grande maioria das ontologias de fundamentação disponibilizam na web seus arquivos OWL correspondentes.

A Figura 38, a seguir, apresenta as ontologias de fundamentação importadas no projeto de desenvolvimento da HEMONTO utilizando o editor Protégé 4.3. O menu “File” deste software possui a opção de importação de ontologias e, após realizada, a aba “Imported ontologies → Direct imports” mostra as ontologias importadas para o projeto.

Figura 38 - Ontologias importadas para a HEMONTO no Protégé 4.3.



Fonte: elaborado pelo autor.

A importação de ontologias usando o Protégé carrega no arquivo OWL da ontologia em desenvolvimento todas as informações contidas das ontologias importadas, o que inclui suas classes (aba “classes”), suas relações (aba “object properties”) e as propriedades de classe e relação, encontradas em forma de campos descritivos e também sob a forma de axiomas lógicos dentro de cada classe e relação.

Realizada tal tarefa, o desenvolvedor tem a sua disposição a estrutura-base a partir da qual pode-se iniciar o processo de formalização da ontologia, que corresponde à etapa seguinte da metodologia *OntoForInfoScience*. É também importante ressaltar que a etapa de fundamentação ontológica não se restringe a carregar as ontologias de fundamentação utilizadas em um editor e seguir com o desenvolvimento específico do domínio, isto porque, é necessário manter o compromisso ontológico com os princípios de fundamentação usados ao longo de todo processo de desenvolvimento. Na construção da HEMONTO, por exemplo, os princípios prescritos na BFO e RO foram seguidos ao longo das atividades de desenvolvimento. Alguns exemplos são:

- i) **Cada classe** da ontologia deve ter exatamente **uma única definição em linguagem natural**, de forma que essa definição descreva da maneira mais próxima da realidade possível o objeto real que ela representa na ontologia.

- ii) **Cada classe** da ontologia deve ter também exatamente **uma única definição em linguagem lógica**, que é derivada e equivalente à definição em linguagem natural.
- iii) O **nó raiz da ontologia não pode ser definido** por ser um termo primitivo, ou seja, cada definição, quando descompactada, deve nos levar de volta ao nó raiz da ontologia.
- iv) **Princípio da não-circularidade**: não se deve usar o próprio termo para definir a classe. Exemplo: hemólise =def. a causa de hemólise. Sugere-se aplicar a “regra do polegar” de Aristóteles para evitar tal circularidade.
- v) **Princípio da mono-hierarquia**: cada classe na hierarquia da ontologia deve ter exatamente uma única classe superior (nó pai), não sendo permitido o conceito de herança múltipla.

Assim, a cada passo da etapa de formalização da ontologia foram respeitados esses princípios de forma a assegurar o compromisso ontológico com as ontologias de fundamentação utilizadas (BFO e RO) e sua base filosófica (realismo ontológico). Na etapa seguinte da metodologia proposta, descreve-se em detalhes o emprego de tais princípios com exemplos ilustrativos no domínio do sangue humano.

7.5 Formalização da ontologia (Etapa 5)

Conforme citado no referencial teórico desta pesquisa, umas das características essenciais que diferencia as ontologias de outros instrumentos terminológicos para organização da informação é a sua capacidade de interpretação pelas máquinas, o que permite a realização de inferências ao conteúdo ontológico com o objetivo de deduzir algum tipo de conhecimento. Essa característica das ontologias deve-se ao uso de estruturas de representação mais completas, que incluem relações diversas, propriedades lógicas (conjuntos domínio e imagem, tipo dos valores, cardinalidade, etc.), axiomas, entre outros. Essas estruturas são representadas em alguma linguagem lógica, tais como a lógica de primeira ordem ou a lógica descritiva da OWL. Portanto, a formalização da ontologia é uma etapa indispensável no seu processo de construção e é descrita aqui como a quinta etapa da metodologia *OntoForInfoScience*.

Na presente pesquisa, compreende-se a etapa de formalização como aquela que produz uma descrição formal do domínio sob estudo baseada na conceitualização deste domínio realizada anteriormente na etapa 3 (três). Portanto, o conhecimento do domínio, tratado anteriormente apenas a nível conceitual, passa a ser tratado a nível

ontológico-formal, o que implica em uma série de adaptações nas estruturas conceituais de forma a atender as restrições ontológicas e formais desta etapa.

Os parágrafos seguintes descrevem os passos incluídos nessa etapa de formalização, apresentando explicações e recomendações de como tratar o conhecimento domínio a nível ontológico-formal.

Passo 5.1: Construir a taxonomia geral da ontologia

As taxonomias são o tipo de estrutura mais comum entre as classes de uma ontologia e permitem organizar o conhecimento desde as classes de nível superior (categorias fundamentais) até as classes mais específicas do domínio – nós internos da taxonomia.

O processo de construção de uma taxonomia inicia-se a partir das entidades mais gerais, normalmente representadas pelas classes da ontologias de fundamentação, e continua com a adição de novas entidades como nós-filhos das classes superiores correspondentes. Em geral, essas novas entidades são entidades específicas do domínio tratado.

A metodologia *OntoForInfoScience* prescreve três atividades essenciais nesse processo de construção da taxonomia geral da ontologia desenvolvida, os quais estão descritas abaixo. Essa taxonomia denominada de geral corresponde àquela que engloba todas as classes da ontologia desenvolvida e, por esse motivo, leva esse nome.

Passo 5.1.1: Tomar um conceito do Conjunto de Modelos Conceituais

A etapa de conceitualização produziu o chamado Conjunto de Modelos Conceituais, que engloba uma série de modelos para representar o conhecimento a nível conceitual do domínio tratado. Esses modelos foram desenvolvidos em algum software de modelagem contando com o apoio de especialistas e estão em formato gráfico (um grafo conceitual, por exemplo) e textual (inclui as definições dos conceitos). O papel do desenvolvedor nesse passo é buscar cada conceito do modelos desenvolvido afim de encontrar uma classe equivalente a ele para representá-lo na ontologia.

Passo 5.1.2: Pesquisar por classe equivalente em outros recursos ontológicos

De posse do conceito tomado do modelo, passa-se à pesquisa em recursos ontológicos existentes de uma classe que represente fidedignamente o significado do

conceito no domínio – sua classe equivalente. Esse passo corresponde ao princípio de reaproveitamento de termos em ontologias, que é universalmente aceito na área.

Para essa pesquisa, o desenvolvedor deve utilizar ferramentas de busca aos recursos ontológicos disponíveis na web e em repositórios de ontologias. Algumas dessas ferramentas sugeridas são: (i) ferramenta do portal *National Center for Biomedical Ontology*⁴⁸ (NCBO); (ii) *OntoBee*⁴⁹: realiza buscas em ontologias integradas ao repositório OBO; (iii) *Ontology Lookup Service*⁵⁰ (OLS); e *Foundational Model Explorer*⁵¹ (FME): realiza buscas na ontologia FMA.

O uso de tais ferramentas é um grande facilitador na procura por classes equivalentes aos conceitos ou por sinônimos desses conceitos, uma vez que elas possibilitam a busca em um número relativamente grande de ontologias e permitem o acesso às propriedades dos termos pesquisados, tais como: suas definições textuais, sinônimos, ontologia de origem e restrições. A partir dessas propriedades, o desenvolvedor pode verificar se o termo recuperado é o mesmo do procurado: por exemplo, ele pode comparar a definição do conceito elaborada na fase de conceitualização com a definição da classe pesquisada em alguma ontologia.

A Figura 39, a seguir, ilustra o resultado de busca ao conceito “*portion of body*” no domínio do sangue humano em recursos ontológicos através da ferramenta *OntoBee*. Neste caso, a *OntoBee* recuperou uma lista de ontologias que contém o termo pesquisado e, dentre elas, a ontologia selecionada para inspeção das propriedades da classe “*portion of body*” foi a FMA. Na construção da HEMONTO, foi considerada equivalente a classe da FMA ao conceito “*portion of body*” do modelo desenvolvido e, assim, a classe foi reutilizada na ontologia HEMONTO.

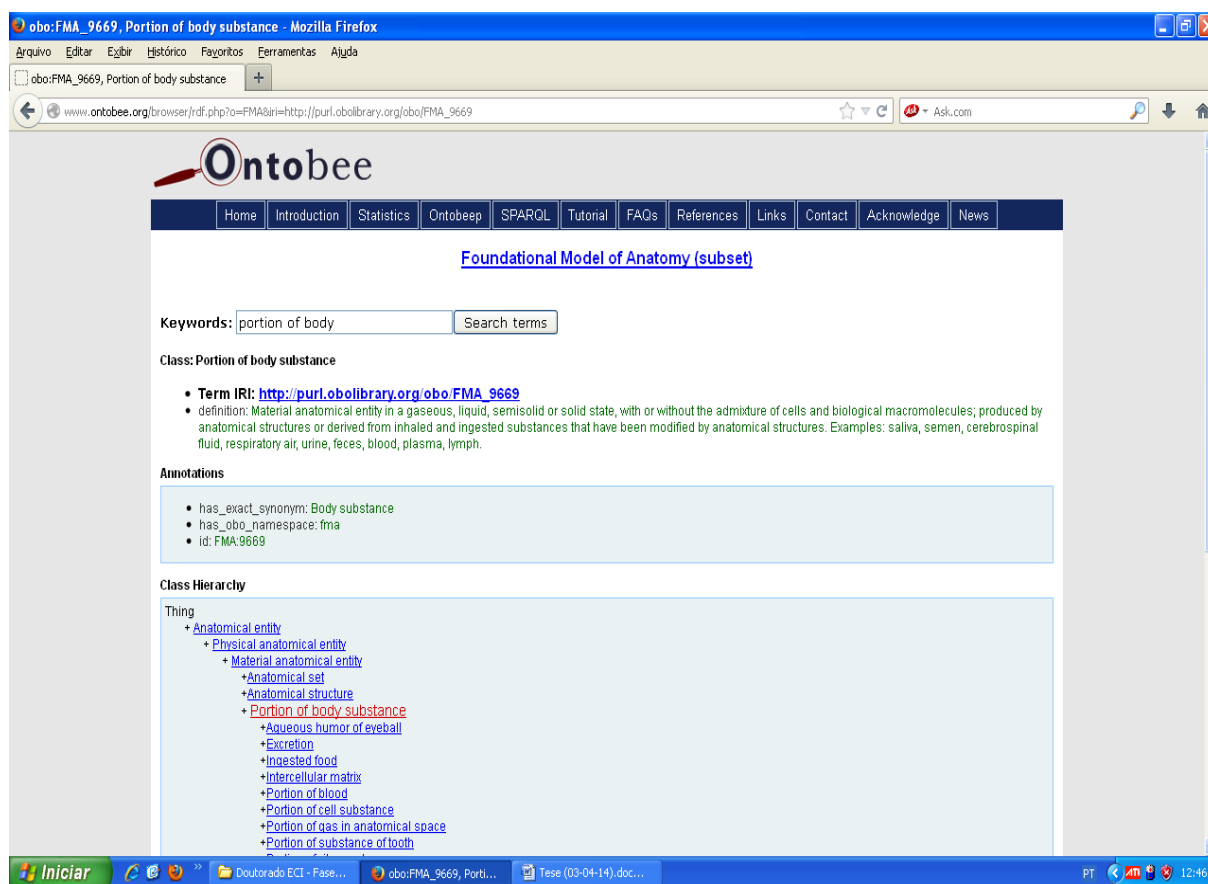
⁴⁸ Disponível em: <http://fme.biostr.washington.edu/FME/>. Acesso em: 20 de Março de 2014.

⁴⁹ Disponível em: <http://www.ontobee.org/>. Acesso em: 20 de Março de 2014.

⁵⁰ Disponível em: <http://www.ebi.ac.uk/ontology-lookup/>. Acesso em: 20 de Março de 2014.

⁵¹ Disponível em: <http://fme.biostr.washington.edu/FME/>. Acesso em: 20 de Março de 2014.

Figura 39 - Resultado da busca pelo termo "portion of body" na ferramenta OntoBee



Fonte: elaborado pelo autor.

Quando uma classe equivalente é encontrada e reutilizada na ontologia em desenvolvimento, suas propriedades como definição, sinônimos, relações, comentários, etc., podem ser também importadas da ontologia de origem caso se julguem válidas para a ontologia desenvolvida.

Por outro lado, quando nenhuma classe equivalente é encontrada ou aquelas encontradas não correspondem ao conceito pesquisado, deve-se criar uma nova classe específica do domínio na ontologia em construção para representar esse conceito.

Passo 5.1.3: Adicionar classe na taxonomia geral

Identificada a classe equivalente ao conceito ou definida a criação de uma nova classe, o passo seguinte é adicionar esta classe na taxonomia geral desenvolvida.

Foi apresentado, anteriormente, que a primeira tarefa na representação da ontologia em um editor para sua formalização é o carregamento do conteúdo de uma ontologia de fundamentação, que inclui suas classes. Assim, as classes mais gerais da

ontologia em desenvolvimento já estarão disponíveis na taxonomia geral em função da importação da ontologia de fundamentação.

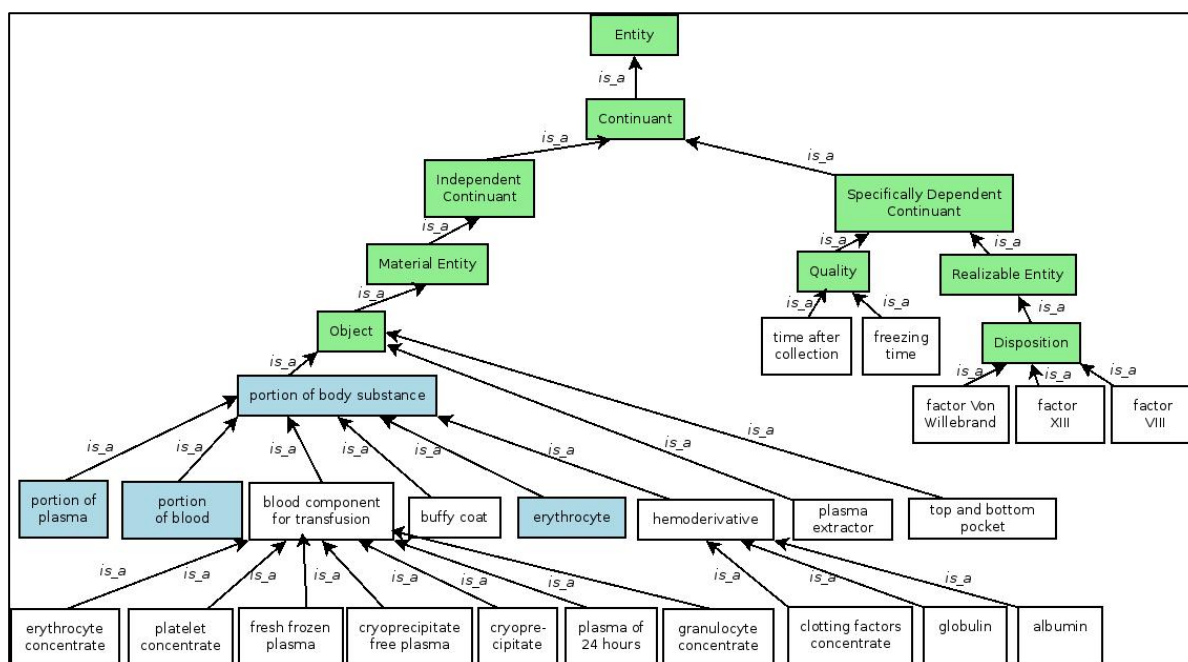
Partindo da ontologia de fundamentação importada, para o desenvolvedor adicionar uma nova classe na taxonomia geral é necessário categorizar ou classificar a classe a ser adicionada como um tipo das classes da ontologia de fundamentação utilizada. Então, quando se determina a classe superior (da ontologia de fundamentação) correspondente à nova classe, esta é inserida como um nó-filho ou subclasse da classe superior na taxonomia.

Nesse processo de construção da taxonomia é fundamental e indispensável seguir o *princípio da mono-hierarquia*, sob o qual cada classe na hierarquia da ontologia deve ter exatamente uma única classe superior (nó-pai), não sendo permitido o conceito de herança múltipla (MUNN e SMITH, 2008).

Na construção da ontologia HEMONTO, cada classe específica do domínio do sangue é classificada como um tipo de classe BFO e adicionada como um nó filho dessa classe na taxonomia geral. Assim, por exemplo, as classes “anatomical structure” e “portion of body substance” são adicionadas na taxonomia como classes “object” (entity => continuant => independent continuant => material entity => object) e as classes “blood coagulation” e “basic blood analysis” são adicionadas como classes “process” (entity => occurrent => process).

Esse processo se repete para cada nova classe adicionada à taxonomia geral da ontologia, tal que novas classes mais específicas vão sendo adicionadas como subclasses das classes específicas do domínio já adicionadas. Isso ocorre, por exemplo, com as classes “portion of blood”, “portion of plasma”, “leukocytes analysis” e “platelets analysis”, tal que as duas primeiras são do tipo “object” e adicionadas como subclasses da classe “portion of body substance”; e as duas últimas são do tipo “process” e adicionadas como subclasses da classe “basic blood analysis”. Ao final da inclusão de todas as classes do domínio forma-se a taxonomia geral da ontologia. Como ilustração, a Figura 40, a seguir, apresenta uma parte da taxonomia geral da ontologia HEMONTO, construída a partir dessa abordagem.

Figura 40 - Parte da taxonomia geral da ontologia HEMONTO



Fonte: elaborado pelo autor.

Uma última consideração sobre a taxonomia geral de uma ontologia é a necessidade de se especificar o tipo da relação formal **is_a** usado na ligação das classes da taxonomia, já que existe mais de um tipo da relação **is_a** do ponto de vista ontológico: (i) *Subsunção sob um gênero (espécie-gênero) ou Instanciação*: neste caso a relação **is_a** é usada como sinônimo de **instance_of** e representa a relação entre um uma instância ou particular e uma classe nessa ordem; (ii) *Especialização ou Especificação (tipo-subtipo)*: neste caso a relação **is_a** é usada como sinônimo de **subclass_of** e representa a relação entre duas classes ou dois tipos, tal que uma é subclasse (ou subtipo) da outra; (iii) *Sinonímia, igualdade ou equivalência*: neste caso a relação **is_a** é usada como sinônimo de **same_as**, a qual representa entre duas entidades que são idênticas.

Ao utilizar o editor de ontologias Protégé, implicitamente, a taxonomia geral da ontologia é desenvolvida usando a relação **is_a** como sinônimo de **subclass_of**. Porém, o desenvolvedor precisa ter conhecimento dessa informação de forma a especificar corretamente as relações entre as classes nesse editor.

No desenvolvimento da ontologia HEMONTO, a relação **is_a** na taxonomia geral é usada como sinônimo de **subclass_of**, usando o editor de ontologias Protégé. Além dessa especificação, outra distinção incluída para a relação **is_a** no desenvolvimento da HEMONTO é se a relação ocorre entre entidades *continuentes ou ocorrentes* do domínio, conforme definição a seguir:

- I. C **is_a** $C1$: $\forall c, \forall t$, se c **instance_of** C at t então c **instance_of** $C1$ at t , tal que $C, C1$ suportam apenas entidades continuantes e c **instance_of** C at t é uma relação primitiva de instanciação, na qual a entidade continuante particular c instancia o universal C num dado tempo t .
- II. P **is_a** $P1$: $\forall p$, se p **instance_of** P então p **instance_of** $P1$, tal que $P, P1$ suportam apenas entidades ocorrentes e p **instance_of** P é uma relação primitiva de instanciação, na qual a entidade ocorrente particular p instancia o universal P .

São dadas duas definições semi-formais para a relação ontológica **is_a**, porque a HEMONTO segue a estrutura taxonômica da ontologia de fundamentação BFO, a qual faz a distinção entre *entidades continuantes* e *ocorrentes*. Isso é importante para a construção de relações ontológicas e sua validação, uma vez que a relação **is_a** pode ser usada entre dois continuantes (relação ontológica I) ou entre dois ocorrentes (relação ontológica II), mas nunca entre uma entidade continuante e outra ocorrente – isso geraria um erro ontológico.

Passo 5.2: Definir propriedades descritivas das classes

Com a classe adicionada à taxonomia geral da ontologia, a tarefa seguinte é definir suas propriedades do ponto de vista ontológico. Na etapa de conceitualização da metodologia *OntoForInfoScience* construiu-se uma Tabela de Conceitos e Propriedades, as quais englobam propriedades extraídas dos documentos da área utilizados e definidas por especialistas do domínio. Essa tabela deve ser usada como base nesse passo, porém é necessário adequar suas propriedades ao nível ontológico, já que elas foram definidas de forma livre e textual na etapa de conceitualização.

Nesse passo serão tratadas apenas as propriedades das classes denominadas na metodologia como “propriedades descritivas”, isto é, propriedades que estão no formato textual (não lógico), mas que seguem as diretrizes de anotação em uma ontologia. No Protégé 4.3 estas propriedades são denominadas de “Annotations” e correspondem a atributos essenciais das classes que foram sendo incorporados nesse editor por serem os principais atributos comuns em ontologias bem conceituadas na área, tais como a *Gene Ontology*, a BFO e a RO, além do padrão de representação RDF. Tais atributos ontológicos ou, simplesmente, propriedades descritivas são de preenchimento obrigatório na metodologia proposta e suas descrições estão apresentadas na Tabela 30, a seguir.

Tabela 30 - Propriedades descritivas das classes de uma ontologia

Propriedade descritiva	Atributo no Protégé	Forma de preenchimento
Identificador da classe	<i>ID</i>	<ul style="list-style-type: none"> Como esse campo é numérico, deve se usar um número inteiro sequencial para preenchê-lo. Caso a classe seja importada de outra ontologia sugere-se manter o número do ID da classe de origem.
Nome da classe	<i>label</i>	<ul style="list-style-type: none"> Preenchimento do atributo no formato <i>string</i>. Utilizar apenas labels com letra minúscula e sem acento, podendo conter espaços ou hifens separadores entre as palavras de uma classe.
Ontologia de origem	<i>imported from</i>	<ul style="list-style-type: none"> Deve-se preenchê-lo com o nome do arquivo da ontologia implementada. Exemplos: GO.owl; FMA.owl.
	<i>hasOBONamespace</i>	<ul style="list-style-type: none"> Deve-se preenchê-lo unicamente com o nome da ontologia, em minúsculo.
URI	<i>hasURI</i>	<ul style="list-style-type: none"> Preenchimento do atributo com o formato padrão de nomes URI's. Exemplo: http://purl.obolibrary.org/obo/FMA_68646
Sinônimos	<i>hasSynonym</i>	<ul style="list-style-type: none"> Preenchimento do atributo no formato <i>string</i>. Quando se tem mais de um sinônimo para o termo, pode-se enumerar todos os sinônimos em um mesmo atributo separados por vírgula (,).
	<i>hasExactSynonym</i>	
	<i>hasRelatedSynonym</i>	
Termo alternativo	<i>alternative term</i>	<ul style="list-style-type: none"> Preenchimento do atributo no formato <i>string</i>.
Definição textual (em linguagem natural)	<i>definition</i>	<ul style="list-style-type: none"> Preenchimento do atributo no formato <i>string</i>. A definição textual deve ser elaborada seguindo algum método que facilite a transição para a definição dessa mesma classe em uma linguagem formal. Seguindo o método Aristotélico, por exemplo, constrói-se uma definição textual da classe simples, com a não inclusão de informações ontológicas (estas são anotadas como comentários) e com uso de nomes (termos) incluídos na própria ontologia ou referenciados por um namespace da ontologia de origem.
Elucidação	<i>elucidaton</i>	<ul style="list-style-type: none"> Preenchimento do atributo no formato <i>string</i>, com possibilidade de uso de símbolos da lógica, tais como variáveis e quantificadores.
Exemplo de uso	<i>example of usage</i>	<ul style="list-style-type: none"> Preenchimento do atributo no formato <i>string</i>. Quando se tem mais de um exemplo de uso, pode-se enumerar todos esses exemplos em um mesmo atributo separados por vírgula (,).
Anotações (Comentários)	<i>comments</i>	<ul style="list-style-type: none"> Preenchimento do atributo no formato <i>string</i>. Quando se tem mais de um comentário da classe, sugere-se que cada comentário possa fazer parte de um novo atributo 'comentário' não utilizando vírgulas para separar vários comentários.

Fonte: elaborado pelo autor.

Há duas situações possíveis no preenchimento dessas propriedades descritivas por parte do desenvolvedor:

- i) quando a classe adicionada é uma classe importada de outra ontologia, o desenvolvedor deve importar as propriedades desta outra ontologia e adaptá-las para o formato adequado de cada propriedade descritiva. Quando necessário, a Tabela de Conceitos e Propriedades pode ser consultada pelo desenvolvedor para complementar as informações sobre propriedades. Um

exemplo clássico de adaptação é no preenchimento da definição textual da classe (atributo “definition”): na maioria dos casos, a definição textual é longa e inclui mais informações do que as realmente necessárias para definir a classe. Assim, neste caso, as informações excedentes para definição devem ser retiradas deste atributo e incluídas na propriedade de anotações (atributo “comments”). Como exemplo, pode-se citar a reutilização da classe “portion of body substance” da ontologia FMA no desenvolvimento da HEMONTO:

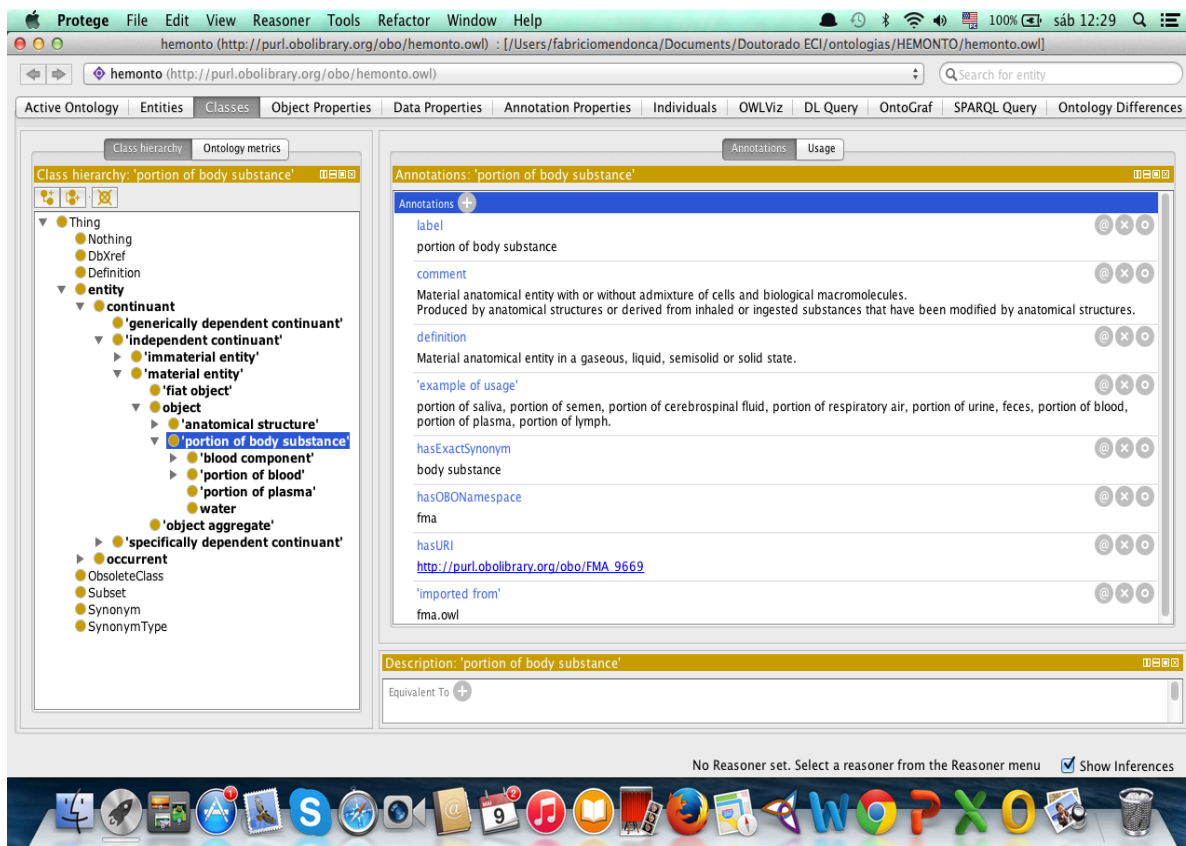
- No FMA, “portion of body substance” é definida como “material anatomical entity in a gaseous, liquid, semisolid or solid state, with or without admixture of cells and biological macromolecules. A portion of body substance is produced by anatomical structures or derived from inhaled or ingested substances that have been modified by anatomical structures”.

- Na HEMONTO, a definição textual (atributo “definition”) de “portion of body substance” fica assim: “material anatomical entity in a gaseous, liquid, semisolid or solid state” e as demais informações são registradas como comentários (atributo “comments”): “material anatomical with or without admixture of cells and biological macromolecules. Produced by anatomical structures or derived from inhaled or ingested substances that have been modified by anatomical structures”.

ii) quando a classe adicionada é um classe nova criada pelo desenvolvedor e não existente em outras ontologias, as propriedades devem ser consultadas na Tabela de Conceitos e Propriedades da etapa de conceitualização e adaptadas para o formato de cada propriedade descritiva. Neste caso, por não haver importação, os atributos “*imported from*” e “*hasOBONamespace*” não precisam de ser preenchidos.

A Figura 41, a seguir, mostra um exemplo de preenchimento dessas propriedades descritivas no Protégé 4.3 para a classe “portion of body substance” da ontologia HEMONTO, a qual foi importada da ontologia FMA através de pesquisa pela ferramenta OntoBee. É possível notar diferenças existentes entre as propriedades contidas para esta classe no FMA e as propriedades preenchidas no Protégé, onde foi necessário complementar as informações requeridas de acordo com a metodologia proposta.

Figura 41 - Preenchimento das propriedades da classe "portion of body substance" no Protégé 4.3



Fonte: elaborado pelo autor.

Dentre as propriedades descritivas destaca-se como uma das mais importantes a definição textual (atributo “definition”) da classe. Nesse sentido, dedicou-se uma atenção especial a ela, conforme apresentado nos parágrafos seguintes.

Em conformidade com o aspecto ontológico essencial de criteriosidade na transição da descrição informal da ontologia para sua descrição formal, as definições textuais das classes devem ser mais próximas o possível das definições formais, de forma a facilitar e suavizar a transição entre estes dois tipos de definições. Para auxiliar nesse propósito, a metodologia *OntoForInfoScience* utiliza alguns princípios ontológicos relacionados à teoria que fundamenta a BFO (MUNN e SMITH, 2008). São eles:

- i) Cada classe da ontologia deve ter exatamente uma **única definição** em linguagem natural, de forma que essa definição descreva da maneira mais próxima da realidade possível o objeto real que ela representa na ontologia.

ii) **Princípio da não-circularidade**: não se deve usar o próprio termo para definir a classe. Sugere-se aplicar a “regra do polegar” de Aristóteles para evitar tal circularidade.

Exemplo: *FMA: anatomical structure (definition)* = Material anatomical entity which is generated by coordinated expression of the organism's own genes that guide its morphogenesis. *Anatomical structure* has inherent 3D shape; its parts are connected and spatially related to one another in patterns determined by coordinated gene expression.

Exemplo adaptado: *HEMONTTO/FMA: anatomical structure (definition)* = Material anatomical entity which is generated by coordinated expression of the organism's own genes that guide its morphogenesis.

HEMONTTO/FMA: anatomical structure (comments) = Anatomical structure has inherent 3D shape; its parts are connected and spatially related to one another in patterns determined by coordinated gene expression.

iii) Deve-se incluir na definição da classe **apenas as condições necessárias e suficientes** que a compõem. Qualquer outra informação (informação complementar) sobre a classe deve ser incluída na ontologia como comentários ou anotações.

Exemplo: *FMA: portion of plasma (definition)* = Portion of body substance corresponding to the liquid component of blood in which blood cells are immersed. Unless otherwise specified the product has been obtained from Whole Blood and frozen.

Exemplo adaptado: *portion of plasma (condição necessária)* = liquid component of blood, *portion of plasma (condição suficiente)* = in which blood cells are immersed.

HEMONTTO/FMA: portion of plasma (definition) = Portion of body substance corresponding to the liquid component of blood in which blood cells are immersed.

HEMONTTO/FMA: portion of plasma (comments) = Unless otherwise specified the product has been obtained from Whole Blood and frozen.

iv) **Termos primitivos** da ontologia não devem ser definidos, mas sim **elucidados**, através de exemplos ilustrativos, declarações do seu uso recomendado ou axiomas lógicos. Termos primitivos correspondem às entidades mais gerais da ontologia que, em geral, são representadas pelas classes das ontologias de fundamentação.

Exemplo: **fiat object** (elucidation) = b is a fiat object part, b is a material entity which is such that for all times t, if b exists at t then there is some object c such that b

proper continuant_part of c at t and c is demarcated from the remainder of c by a two-dimensional continuant fiat boundary.

fiat object (example of usage) = a divisão do cérebro em regiões; superfícies dorsal e ventral do corpo humano.

fiat object (definition) = não possui.

De acordo com esses princípios ontológicos e exemplos apresentados, o papel do desenvolvedor quanto à definição textual de classes é adaptar as definições textuais encontradas na Tabela de Conceitos e Valores ou nas classes importadas de outras ontologias, de forma a atender as restrições incluídas para esta tarefa na metodologia *OntoForInfoScience*. Tais restrições são necessárias e úteis para o próximo passo da formalização da ontologia, que corresponde à criação de definições formais das classes.

Passo 5.3: Criar uma definição formal para cada classe da ontologia

Outro princípio ontológico fundamental na etapa de formalização de uma ontologia é: *Cada classe da ontologia deve ter também, exatamente, uma única definição em linguagem lógica, que seja derivada e equivalente à definição em linguagem natural.* (MUNN e SMITH, 2008).

Vamos chamar essa definição em linguagem lógica de **definição formal** da classe na ontologia e, nesse passo, temos a descrição dos procedimentos a serem adotados pelo desenvolvedor para a construção de definições formais da ontologia desenvolvida.

Antes da descrição de tais procedimentos, é importante destacar que a criação de definições formais das classes é de fundamental importância para possibilitar processos de inferência à base de conhecimento da ontologia desenvolvida, garantindo, assim, que seu conteúdo possa ser interpretado pelas máquinas. A maioria das metodologias de construção atuais não incluem ou não detalham bem essa tarefa.

Como já mencionado, a construção de definições formais das classes em alguma linguagem lógica depende necessariamente de como foram elaboradas as definições textuais desses termos na ontologia. Nesse sentido, o passo anterior da metodologia *OntoForInfoScience* é imprescindível para o sucesso do passo atual, uma vez que os princípios e restrições para as definições textuais facilitam a transição entre a definição textual e a formal.

As restrições impostas para a criação da definição formal da classe são listadas, a seguir:

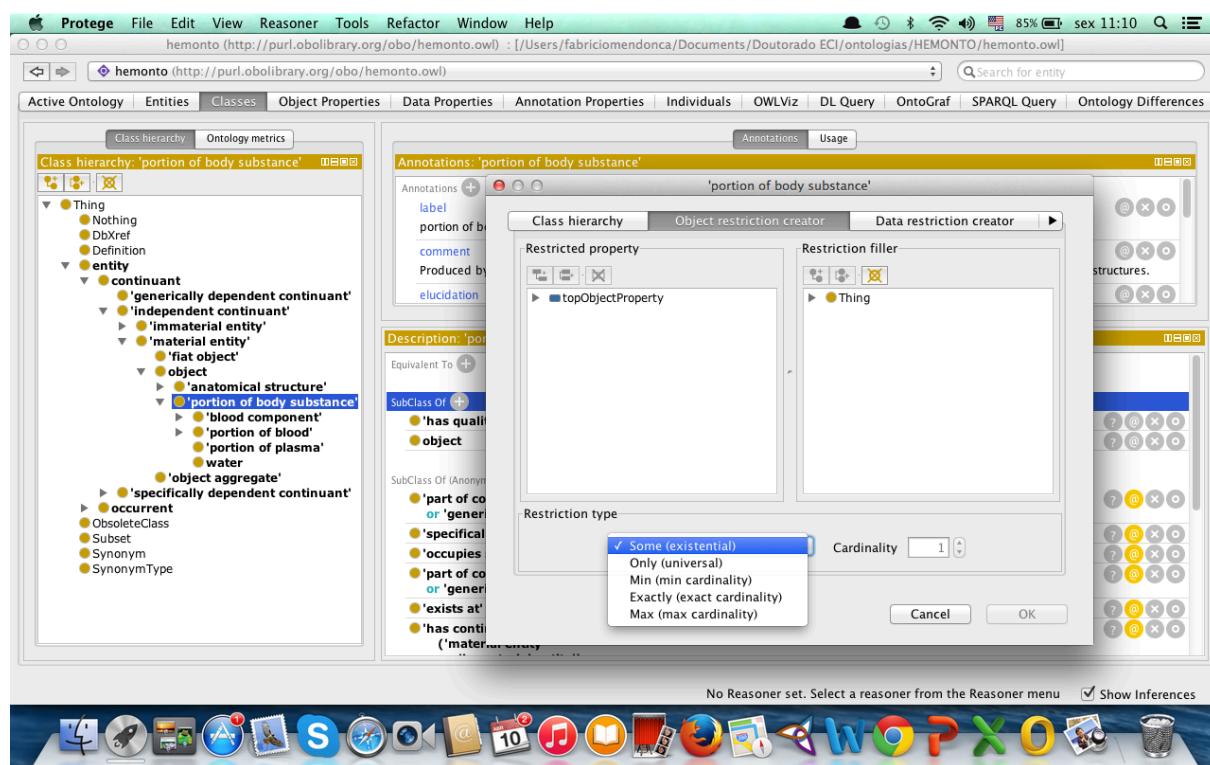
- i) Escolher uma *linguagem lógica* a ser usada para a representação das definições formais. Caso o desenvolvedor esteja usando o editor de ontologias Protégé, a linguagem usada nesse software é a OWL-DL, que é a mais comum para este tipo de representação.
- ii) No momento da derivação da descrição textual para a formal, deve-se utilizar apenas *nomes (termos) que estejam incluídos na própria ontologia* ou referenciados por um namespace da ontologia de origem.
- iii) *Substantivos isolados ou expressões nominais* formadas por um substantivo seguido de outra classe gramatical (adjetivo, preposição, pronome ou outro substantivo) indicam uma possível *classe* da ontologia e, por esse motivo, na descrição formal devem ser substituídos unicamente por uma classe da ontologia desenvolvida ou por um namespace da ontologia de origem.
- iv) *Verbos isolados ou expressões verbais*, formadas por um verbo seguido de outra uma preposição ou outro verbo, indicam uma possível *relação* da ontologia e, por esse motivo, na descrição formal devem ser substituídos unicamente por relações ontológicas de alguma outra ontologia ou da própria ontologia em desenvolvimento.
- v) No momento da construção da definição lógica, ao se relacionar duas classes da ontologia deve-se definir as *restrições de tipo para as condições necessárias e suficientes* das classes, as quais são equivalentes aos quantificadores das linguagens lógicas: \forall (quantificador universal – para todo) e \exists (quantificador existencial – existe pelo menos um). No Protégé, que usa a linguagem OWL, essas restrições são representadas por: *Only (universal)* e *Some (existential)*, respectivamente.
- vi) Ao se relacionar duas classes da ontologia deve-se definir também a restrição de cardinalidade, a qual deve informar a faixa de valores possíveis entre as classes da relação. No Protégé, que usa a linguagem OWL, essas restrições são representadas por: *Min (cardinalidade mínima)*, *Exactly (cardinalidade exata)* e *Max (cardinalidade máxima)*.

Para uma melhor compreensão da aplicação destas regras na construção de definições formais na ontologia, apresenta-se a seguir dois exemplos de definições na ontologia HEMONTO para os conceitos “*portion of blood*” e “*portion of plasma*”.

Primeiramente é necessário apresentar a interface do Protégé para definições formais. O Protégé 4.3 disponibiliza ao usuário o campo “SubClass Of” para inserção da

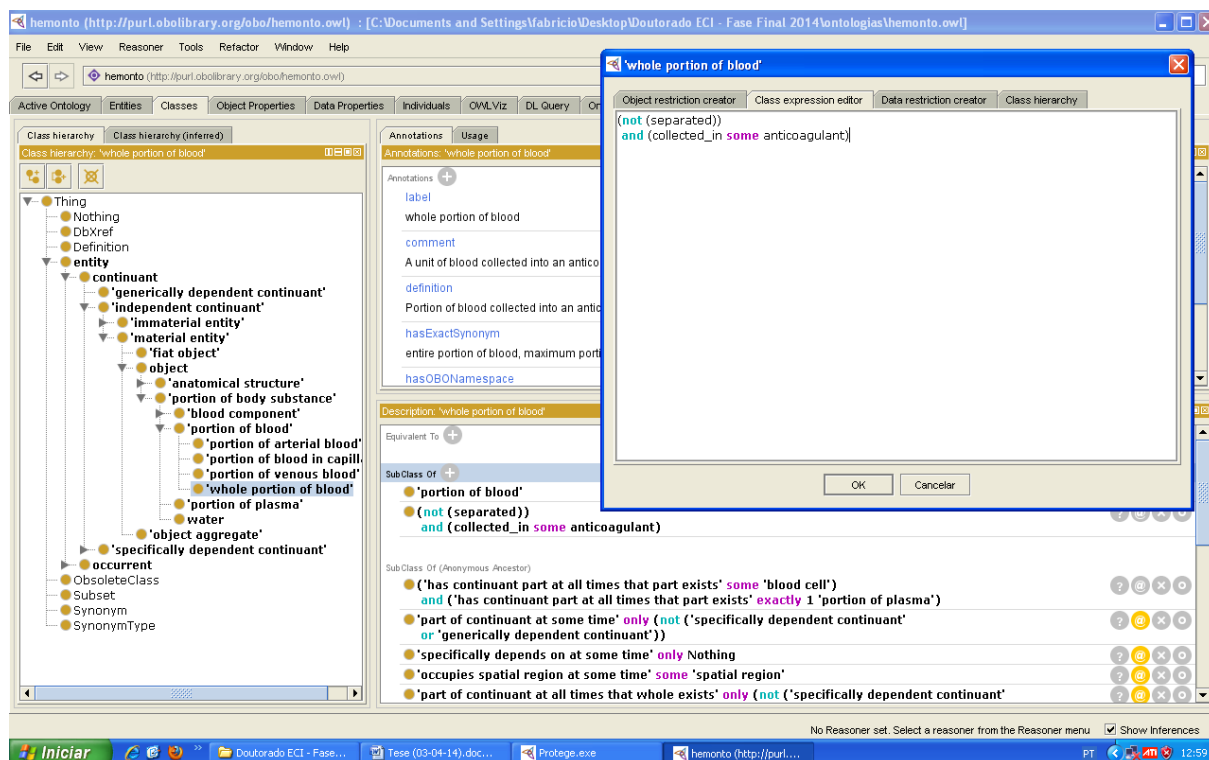
definição formal em cada classe da ontologia. Tal campo recebe esse nome, porque cada classe da ontologia é definida em função de sua superclasse correspondente, exceto o nó-raiz da taxonomia. Para validação da sintaxe das declarações formais da classe implementadas pelo desenvolvedor, o Protégé fornece duas opções ao usuário: (i) a aba “*Object restriction creator*”, onde o usuário pode selecionar as classes e a relação da ontologia que serão usadas na definição formal, além das restrições de tipo e cardinalidade (veja Figura 42); (ii) a aba “*Class expression editor*”, onde o usuário dispõe de um editor de expressões para classes, no qual pode-se especificar classes, relações e restrições da ontologia para construir a definição formal desde que as declarações estejam escritas corretamente de acordo com a sintaxe da linguagem OWL usada neste software (veja Figura 43). Qualquer uma dessas duas opções (ou caminhos) do Protégé permite a criação da definição formal da classe, seguindo as regras elucidadas na metodologia proposta.

Figura 42 - Preenchimento das restrições de tipo e cardinalidade no Protégé 4.3



Fonte: elaborado pelo autor.

Figura 43 - Criação da definição formal de uma classe no Protégé 4.3



Fonte: elaborado pelo autor.

Compreendida a interface do Protégé nesta parte, passe-se ao processo de definição formal das classes, neste caso dos exemplos “*portion of blood*” e “*portion of plasma*” da ontologia do sangue proposta.

A definição textual da classe “*portion of blood*” foi extraída da ontologia FMA, de onde essa classe foi importada para a ontologia HEMONTO. Tal definição na FMA já estava em um formato textual adequado (conforme princípios mencionados anteriormente) para a transição para a forma lógica. Assim, a essa definição textual foram aplicadas as restrições de i) a vi), para a criação de uma definição formal da classe, conforme mostrado a seguir.

Portion of blood: definição textual

Da FMA - Portion of blood: portion of body substance, which has as its parts plasma and blood cells.

Anotação da definição textual

Da FMA - Portion of blood: portion of body substance [FMA/HEMONTO: *portion of body substance*], which **has as its parts** [RO: *has_part*] plasma [FMA/HEMONTO: *portion of plasma*] and blood cells [FMA/HEMONTO: *blood cell*].



- i) uso da linguagem lógica OWL-DL do Protégé;
- ii) expressões nominais e verbais foram destacadas na definição textual afim de assegurar que sejam usadas na definição formal apenas termos contidas na ontologia ou em ontologias importadas.
- iii) expressões nominais foram sublinhadas na definição textual e para cada uma realizou-se sua associação com a classe da ontologia identificada, colocada entre colchetes. Por exemplo: plasma [FMA/HEMONTO: *portion of plasma*].
- iv) de maneira similar, expressões verbais foram destacadas em negrito e para cada uma realizou-se sua associação com uma relação ontológica, colocada entre colchetes. Por exemplo: **has as its parts** [RO: *has_part*]. Além destas, é sempre incluída para cada classe a relação **is_a** entre esta classe e sua superclasse, que está implícita na definição textual apresentada: *Portion of blood: (is_a) portion of body substance*.
- v) analisando elementos textuais como: artigos, pronomes, plural/singular; é possível determinar as restrições de tipo para as condições necessárias e suficientes da classe. Exemplo: (some) blood cells.
- vi) de maneira semelhante, identifica-se as restrições de cardinalidade. Exemplo: (exactly 1) plasma.



Portion of blood: definição formal (em lógica OWL-DL)

portion of blood **is_a** *portion of body substance* and (**has_part** exactly 1 *portion of plasma*)
and (**has_part** some *blood cell*)

O mesmo procedimento é adotado para a elaboração de todas as outras classes da ontologia HEMONTO e o resultado final é uma ontologia axiomatizada a partir das definições formais de suas classes. A seguir, outro exemplo de definição formal, agora para a classe “portion of plasma”.

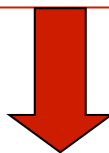
Portion of plasma: definição textual

Da FMA - *Portion of plasma*: portion of body substance corresponding to the liquid component of blood in which blood cells are immersed.



Anotação da definição textual

Da FMA - Portion of plasma: portion of body substance [FMA/HEMONTO: *portion of body substance*] **corresponding to** [BFO: *has_quality*] the liquid [FMA/HEMONTO: *liquid*] **component of** [RO: *part_of*] blood [FMA/HEMONTO: *portion of blood*] in which blood cells [FMA/HEMONTO: *blood cell*] **are immersed** [RO: *contains*].



Portion of plasma: definição formal (em lógica OWL-DL)

Portion of plasma **is_a** *portion of body substance* and (**has_quality** *liquid*) and (**part_of** *portion of blood*) and (**contains** *blood cell*).

Passo 5.4: Definir propriedades de dados das classes

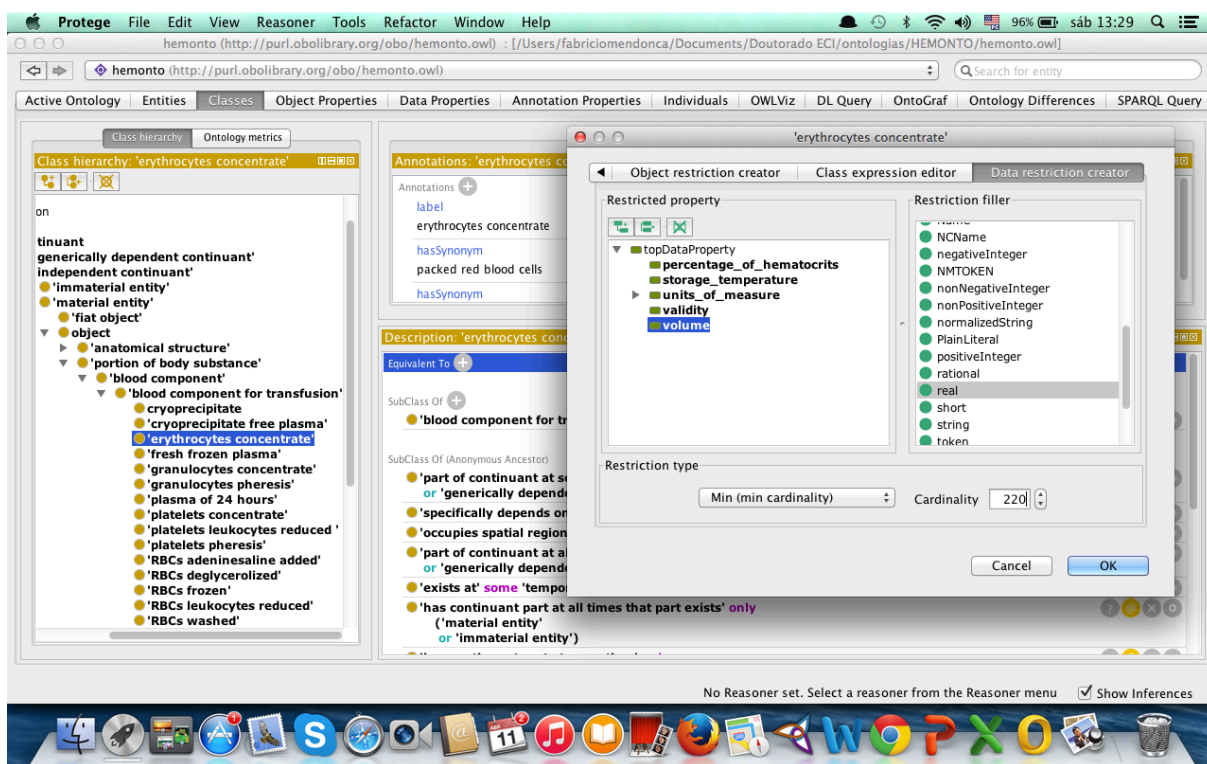
Além das propriedades descritivas e das restrições lógicas impostas pela definição formal da classe, na formalização da ontologia também não pode faltar o preenchimento das chamadas *propriedades de dados* (*data properties*) das classes. Propriedades de dados são utilizadas em uma ontologia para especificar restrições aos valores (dados) possíveis que podem ser atribuídos às classes da ontologia. Assim, tais propriedades servem para restringir as instâncias que uma classe pode conter. Tipos de dados (valores), cardinalidade e quantificadores universais ou existenciais são alguns exemplos de propriedades de dados. No método 101 (NOY e MCGUINNESS, 2001) as propriedades dos dados são chamadas de *facet*s (*facets*) das propriedades gerais (*slots*).

Na construção da ontologia HEMONTO, a definição de propriedades de dados para as classes foi utilizada para representar as características específicas de obtenção de cada hemocomponente do sangue, tais como: o volume necessário do hemocomponente, sua validade, a temperatura de armazenamento para sua conservação e a porcentagem de hematócritos em seu volume total. No editor de ontologias Protégé, as propriedades de dados são definidas na aba “Data Properties” e, após sua criação, elas devem ser associadas às classes que as possuem. Um exemplo dessa definição e associação é apresentado na Figura 44, a seguir, para a classe “erythrocytes concentrate” (concentrado de hemácias) na ontologia HEMONTO com o uso do Protégé.

O concentrado de hemácias (CH) é obtido por meio da centrifugação de uma bolsa de sangue total (ST) e da remoção da maior parte do plasma. Seu volume varia entre 220ml e 280ml.

Assim como o ST, o concentrado de hemácias deve ser mantido entre 2°C e 6°C e sua validade varia entre 35 e 42 dias, dependendo da solução conservadora. Os concentrados de hemácias sem solução aditiva devem ter hematócrito entre 65% e 80%. No caso de bolsas com solução aditiva, o hematócrito pode variar de 50% a 70%.

Figura 44 - Preenchimento das propriedades de dados de uma classe no Protégé 4.3



Fonte: elaborado pelo autor.

Após realizado tal procedimento para cada propriedade da classe “erythrocytes concentrate”, o conjunto de final de “data properties” da classe na ontologia é composto conforme dados da Tabela 31, a seguir.

Tabela 31 - Propriedades de dados (data properties) da classe "erythrocytes concentrate"

Concentrado de Hemácias (<i>Erythrocytes Concentrate</i>)	
Volume (<i>volume</i>)	Value = real and (volume >= 220 and volume <= 280) and (unit_of_measure = milliliters)
% de hematócritos (<i>percentage_of_hematocrits</i>)	- sem solução aditiva: value = real and (percentage_of_hematocrits >= 65 and percentage_of_hematocrits <= 80) - com solução aditiva: value = real and (percentage_of_hematocrits >= 50 and percentage_of_hematocrits <= 70)
Temperatura de armazenamento (<i>storage_temperature</i>)	Value = real and (storage_temperature >= 220 and storage_temperature <= 280) and (unit_of_measure = Celsius)
Validade (<i>validity</i>)	Value = integer and (validity >= 35 and validity <= 42) and (unit_of_measure = days)

Fonte: elaborado pelo autor.

Passo 5.5: Criar instâncias das classes

O passo seguinte na etapa de formalização da ontologia consiste na criação de instâncias para as classes. Essa também é uma tarefa indispensável no desenvolvimento ontológico e auxilia nas respostas às questões de competência propostas à ontologia.

Instâncias são exemplos, ocorrências, casos particulares e reais das classes usadas para representar objetos específicos da realidade de um dado domínio. Na literatura da área, o termo “instância” é equivalente aos termos “particular” e “individual”, quando uma instância é usada para se referir a um objeto particular ou específico da realidade.

A grande dificuldade dessa tarefa para o desenvolvedor, é que ele precisa decidir se um determinado conceito representa uma classe ou uma instância na ontologia, de acordo com o domínio tratado e com as aplicações potenciais da ontologia. Um dos recursos que pode ajudar o desenvolvedor nessas decisões é a Tabela de Conceitos e Valores construída na etapa de conceitualização, já que ela enumera os valores possíveis de cada conceito do domínio e que podem se tornar as instâncias da classe que representa este conceito. Entretanto, apenas essa enumeração não é suficiente, sendo para isso necessário alguns princípios básicos para ajudar o desenvolvedor na decisão ontológica de definir o que é classe e o que é instância. Para essa tarefa, a metodologia *OntoForInfoScience* sugere algumas regras básicas, inspiradas no método 101 e na metodologia NeOn, conforme descrito a seguir:

- i) como orientação geral, deve-se consultar o propósito geral da ontologia, seus objetivos específicos e seus casos de aplicação;
- ii) deve-se determinar o menor nível de granularidade da representação. Esse nível deve representar a divisão entre conceitos que são classes e outros que são instâncias;
- iii) as questões de competência da especificação da ontologia devem ser consultadas como suporte para determinação de conceitos como classes ou instâncias: nomes ou adjetivos das perguntas das questões de competência indicam classes da ontologia; nomes ou adjetivos das respostas às questões de competência indicam instâncias da ontologia.

Exemplo:

QC2. Quais proteínas são encontradas numa porção de plasma sanguíneo?

Resposta: albumin, fibrinogen, globulin.

Candidatos a classes: proteínas (protein), porção de plasma sanguíneo (portion of plasma). **Candidatos a instâncias:** albumin, fibrinogen, globulin.

QC3. Quais os tipos de leucócitos podem ser encontrados numa porção de sangue?

Resposta: granular leukocyte: basophil, eosinophil, neutrophil; nongranular leukocyte: lymphoblast, lymphocyte, monocyte, peripheral blood mononuclear.

Candidatos a classes: leucócitos (leukocyte), porção de sangue (portion of blood), granular leukocyte, nongranular leukocyte. **Candidatos a instâncias:** basophil, eosinophil, neutrophil, lymphoblast, lymphocyte, monocyte, peripheral blood mononuclear.

QC7 (alterada). Quais as condições de temperatura e armazenamento o plasma sanguíneo deve ser submetido para obtenção do hemocomponente plasma fresco congelado?

Resposta: fresh frozen plasma: $-25^{\circ} \text{C} \leq \text{storage_temperature} \leq -18^{\circ} \text{C}$ and volume $\geq 180 \text{ ml}$ and validity = 12 months.

Candidatos a classes: temperatura, armazenamento, plasma sanguíneo (portion of plasma), plasma fresco congelado (fresh frozen plasma).

Candidatos a instâncias: -25°C , -18°C , 180 ml, 12 months.

iv) conceitos devem ser representados como instâncias quando são os itens mais específicos da árvore de conceitos (nós-folha) e abaixo deles não há a continuidade da ontologia – a ontologia pára;

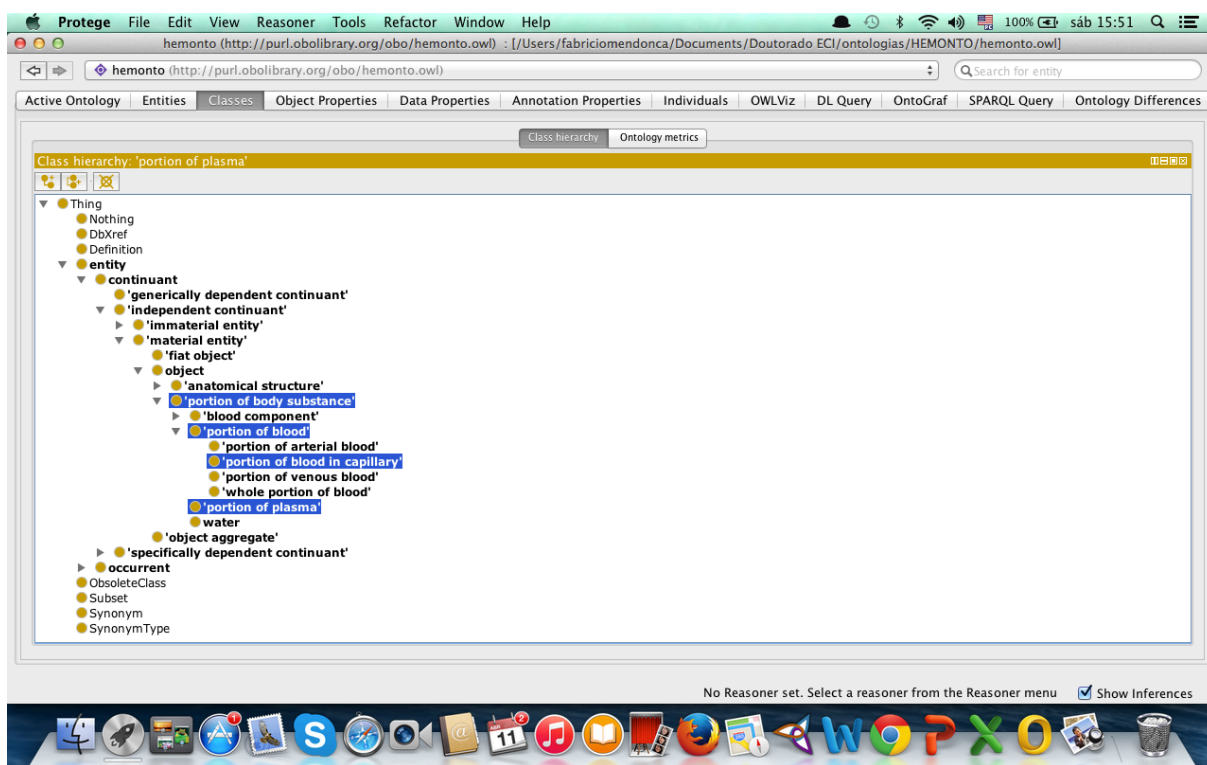
v) conceitos que possuem propriedades específicas diferentes de outros conceitos devem ser representados como classes. Exemplo: os diferentes hemocomponentes do sangue humano na ontologia HEMONTO, conforme mostrado na Tabela 32, a seguir.

Tabela 32 - Exemplos de conceitos com propriedades específicas que são classes

Componente (conceito)	Volume	% de hematócritos	Temperatura	Validade
Concentrado de hemácias	$\geq 220 \text{ ml}$ $\leq 280 \text{ ml}$	Sem solução aditiva entre 65% e 80%. Com solução aditiva entre 50% e 70%.	$\geq 2^{\circ} \text{C}$ $\leq 6^{\circ} \text{C}$	$\geq 35 \text{ dias}$ $\leq 42 \text{ dias}$
Plasma Fresco Congelado	$\geq 180 \text{ ml}$	0%	$\geq -25^{\circ} \text{C}$ $\leq -18^{\circ} \text{C}$	12 meses
Plasma 24 horas	$\geq 200 \text{ ml}$ $\leq 250 \text{ ml}$	0%	$\geq -30^{\circ} \text{C}$ $\leq -18^{\circ} \text{C}$	12 meses
Crioprecipitado	15 ml	0%	$\geq 1^{\circ} \text{C}$ $\leq 6^{\circ} \text{C}$	1 ano

vi) se determinados conceitos formam uma hierarquia natural no domínio, estes devem ser representados como classes. Exemplo: no domínio do sangue humano, os conceitos “*portion of body substance*” – “*portion of blood*” – “*portion of blood in capillary*” e “*portion of plasma*” formam uma hierarquia natural, conforme representado na Figura 45 abaixo.

Figura 45 - Exemplo de hierarquia natural no domínio do sangue



Fonte: elaborado pelo autor.

Uma última observação sobre a criação de instâncias das classes é que elas devem ser criadas respeitando as restrições (propriedades de dados) da classe, tais como: o tipo de valor (numérico, string, booleano, etc.), a cardinalidade, o domínio e imagem da classe. Essa medida possibilita a realização de inferências corretas para as instâncias da ontologia. Um editor de ontologias, como o Protégé 4.3, por exemplo, realiza essa checagem automática das restrições da classe para suas instâncias.

Passo 5.6: Especificar relações ontológicas

Após a construção da taxonomia geral da ontologia, da definição de classes, suas propriedades e da criação de instâncias destas classes, o passo seguinte na metodologia *OntoForInfoScience* compreende a especificação das relações entre as

classes da ontologia, chamadas de *relações ontológicas* devido a sua natureza estritamente ontológica, isto é, relações identificadas apenas entre entidades da realidade do domínio sob estudo.

Já na fase de conceitualização da ontologia, o processo de especificação de relações conceituais do domínio é considerado um dos problemas mais difíceis (AUGER e BARRIERE, 2010) porque é necessário encontrar uma relação que represente da maneira mais aproximada possível o relacionamento identificado na realidade entre as entidades relacionadas. Na maioria dos casos, é bastante difícil de se atingir um consenso entre especialistas do domínio sobre o tipo mais adequado de relação a ser utilizado.

O uso de softwares para conceitualizações colaborativas, tal como o ConceptMe (SOUSA et al., 2013), é extremamente útil para essa tarefa de especificação de relações conceituais e, por esse motivo, a metodologia proposta utilizou o ConceptMe na etapa de conceitualização. Entretanto, quando é necessário tratar do conhecimento do domínio a nível ontológico-formal, é preciso ir além da colaboração a nível conceitual e realizar a transição das relações conceituais para relações estritamente ontológicas. Este é o propósito principal deste passo da metodologia *OntoForInfoScience*, cujos resultados já foram aceitos para publicação em (MENDONÇA et al., 2015).

Antes de descrever propriamente o conjunto de regras e princípios para a transição das relações conceituais para o nível formal-ontológico, é importante citar que nos passos anteriores da metodologia proposta, o desenvolvedor já se deparou com relações do domínio e, por necessidades de especificação, já reutilizou alguma relação de uma ontologia importada ou mesmo criou uma relação específica do domínio para concluir sua tarefa. A criação de definições formais para as classes da ontologia (passo 5.3), por exemplo, apenas pode ser concretizada com a utilização de relações ontológicas para se definir o termo. Nesse sentido, algumas relações ontológicas já foram criadas, mas é preciso validá-las e definir suas propriedades através das regras e princípios descritos nesse passo.

Na tarefa de especificação de relações ontológicas do domínio, a metodologia *OntoForInfoScience* estabelece as seguintes regras e princípios básicos:

- i) Pesquisar e reutilizar, sempre que possível, relações ontológicas já definidas em outras ontologias desenvolvidas, sejam elas ontologias de fundamentação ou ontologias de domínio relacionadas à área do conhecimento sob estudo. A pesquisa por relações já incluídas em outras ontologias segue o mesmo princípio da pesquisa por classes, ou seja, sugere-se a utilização de

ferramentas de busca aos repositórios ontológicos, tais como a *OntoBee*⁵² e o *Ontology Lookup Service*⁵³, seguida da avaliação por parte do desenvolvedor das propriedades da relação a ser importada para a ontologia em desenvolvimento, por exemplo, se a relação tem correspondência com o relacionamento pesquisado ou se a relação possui propriedades suficientes que permitam sua utilização na ontologia.

ii) Considerar o uso de relações da ontologia de fundamentação importada ou importar uma ontologia específica de relações ontológicas, tal como a *Relation Ontology (RO)* para relações em domínios biomédicos. Ontologias de fundamentação, inclusive aquelas específicas em relações ontológicas, seguem princípios ontológicos fundamentais das teorias em que baseiam e são reutilizadas na construção de muitas ontologias de domínio, tal como o uso da RO no repositório de ontologias da OBO. Essas características das ontologias de fundamentação facilitam a integração entre ontologias e o trabalho do desenvolvedor na especificação de propriedades descritivas e formais das relações em uma ontologia. Como exemplo, é feita novamente referência à construção da ontologia HEMONTO, na qual a base de relações ontológicas é composta das relações incluídas nas duas ontologias de fundamentação importadas: BFO e RO. Foram utilizadas relações de ambas ontologias, tal que algumas relações redundantes foram excluídas para garantir a consistência ontológica. Isso ocorreu porque em suas versões mais recentes, a BFO tem incorporada em seu conteúdo as relações ontológicas da RO. A Tabela 33, a seguir, traz alguns exemplos de relações da BFO e RO utilizadas no projeto de desenvolvimento da HEMONTO.

Tabela 33 - Relações ontológicas da RO e BFO utilizadas na construção da HEMONTO

Relação	Ontologia de origem	Definição semi-formal
P is_a P1	RO	$\forall p$, se p instance_of P então p instance_of P1, tal que P, P1 suportam apenas entidades ocorrentes e p instance_of P é uma relação primitiva de instanciação, na qual a entidade ocorrente particular p instancia o universal P. Exemplos de uso: - Process of freeze is_a Process. - Process of centrifugation at high rotation is_a Process of centrifugation.
C part_of C1	RO	$\forall c, \forall t$, se c instance_of C at t então há algum c1 tal que c1 instance_of C1 at t e c part_of c1 at t, onde c part_of c1 at t é uma relação primitiva entre

⁵² Disponível em: <http://www.ontobee.org/>. Acesso em: 20 de Março de 2014.

⁵³ Disponível em: <http://www.ebi.ac.uk/ontology-lookup/>. Acesso em: 20 de Março de 2014.

		<p>dois particulares continuantes, na qual um é parte do outro no tempo t mencionado.</p> <p>Exemplos de uso:</p> <ul style="list-style-type: none"> - Plasma part_of Whole blood. - Fibrinogen part_of Cryoprecipitate
C part of continuant at all times C1	BFO 2.0	<p>$b \text{ continuant_part_of } c \text{ at } t = \text{Def. } b \text{ is a part of } c \text{ at } t \ \& \ t \text{ is a time } \& \ b \text{ and } c \text{ are continuants. (axiom label in BFO2 Reference: [002-001])}$</p> <p>Exemplos de uso:</p> <ul style="list-style-type: none"> - portion of blood 'has continuant part at all times that part exists' some 'blood cell' and 'has continuant part at all times that part exists' exactly 1 'portion of plasma'.
P produces C	RO	<p>$\forall p$, se p instance_of P então há algum c e algum t; tal que se c instance_of C1 at t and p produces c at t, onde p produces c at t é uma relação entre o processo p, o continuante c e um tempo t, no qual p produces c se algum processo que occurs_in p has_output c.</p> <p>Exemplos de uso:</p> <ul style="list-style-type: none"> - Process of centrifugation produces Buffy coat. - Cryoprecipitate extraction produces Cryoprecipitate free plasm.
P preceded_by P1	RO	<p>Definição:</p> <p>$\forall p$, se p instance_of P então há algum p1 tal que p1 instance_of P1 and p preceded_by p1, onde p preceded_by p1 = $\forall t, t1$, se p occurring_at t and p1 occurring_at t1, então t1 earlier t, onde t earlier t1 é uma relação primitiva entre dois tempos tal que t ocorre antes de t1 e p occurring_at t = for some c, p has_participant c at t.</p> <p>Exemplos de uso:</p> <ul style="list-style-type: none"> - Process of centrifugation at high rotation preceded_by Process of centrifugation. - Process of collection preceded_by Process of centrifugation.

Fonte: elaborado pelo autor.

iii) Quando não for encontrada nenhuma relação ontológica equivalente em ontologias existentes, cabe ao desenvolvedor criar uma nova relação na ontologia que, em geral, é específica do domínio tratado. Essa é uma prática comum quando a ontologia desenvolvida representa um domínio complexo de conhecimento. Nesses casos, o desenvolvedor deve criar uma relação assegurando seu compromisso ontológico e, para tanto, a metodologia *OntoForInfoScience* também prescreve alguns princípios básicos para esta tarefa específica:

iii.i) a relação criada deve ser genuinamente ontológica, isto é, uma relação identificada entre entidades na realidade, independente da nossa experiência ou métodos de aprendizagem.

iii.ii) a relação criada deve ser obtida universalmente, ou seja, uma declaração na forma *A relation B* deve ser obtida para todas as instâncias

de A e não apenas obtida, por exemplo, a partir de algumas instâncias representativas estatisticamente de A.

iii.iii) a relação criada deve ser definida de maneira única e rigorosa, o que significa dizer que é necessária uma definição formal ou semi-formal clara para a relação e que definições intuitivas não devem ser feitas.

Com base nos princípios e regras de criação de uma nova relação ontológica, foram definidas algumas relações específicas no domínio do sangue humano, incorporadas na ontologia HEMONTO. A Tabela 34, a seguir, apresenta três exemplos.

Tabela 34 - Algumas relações específicas do domínio do sangue na HEMONTO

Relação	Ontologia de origem	Definição informal
P prevent P1	HEMONTO	Relação ontológica identificada entre dois ocorrentes, P e P1, sendo que o ocorrente P é uma entidade realizável que evita ou previne a realização do processo (ocorrente) P. Exemplos de uso: - anticoagulant is_a realizable_entity and prevent some 'blood coagulation'
C collected_into P	HEMONTO	Relação ontológica identificada entre um continuante C e um ocorrente P, que é uma entidade realizável, tal que o continuante C é extraído ou coletado através de algum meio ou processo, que corresponde à entidade realizável P. Exemplos de uso: whole portion of blood is_a 'portion of blood' collected_in some anticoagulant and not (separated)
C prepared_by P	HEMONTO	Relação ontológica identificada entre um continuante C e um ocorrente P, tal que o continuante C é obtido ou gerado a partir da realização do processo (ocorrente) P. Exemplos de uso: pooled platelet is_a 'blood component for transfusion' prepared by combining some platelets into one container.

iv) Caracterizar ou definir o tipo da relação ontológica criada ou importada, quando esta relação possui mais de um tipo. Em um dos tópicos do Capítulo 5 da presente pesquisa, discorreu-se sobre relações que possuem mais de um tipo específico sob o mesmo rótulo (*label*), como é o caso das relações *is_a* e *part_of*. Para estas relações, o desenvolvedor deve deixar claro em sua ontologia qual o tipo específico desta relação está sendo usado em um determinado relacionamento entre duas classes da ontologia. Como já citado anteriormente, essa é uma medida imprescindível em domínios biomédicos e também em muitos domínios especializados. Para esta caracterização dos

tipos de relação ontológica, a metodologia *OntoForInfoScience* se baseou em uma revisão dos tipos identificados na literatura da área (WINSTON et al., 1987) (VARZI, 1996) (KEET e ARTALE, 2008) e, assim, construiu uma lista dos tipos possíveis para as relações *is_a* e *part_of*. O papel do desenvolvedor nessa tarefa de caracterização é, com base na identificação dos diferentes tipos dessas relações, criar no editor de ontologias uma relação (Object Property no Protégé) para cada tipo e também sua definição semi-formal ou formal que o diferencie dos outros tipos. A Tabela 35, a seguir, apresenta os tipos específicos das relações *is_a* e *part_of*, e a Figura 46 mostra a hierarquia de tipos da relação *part_of* construída no Protégé 4.3. É importante destacar que na tabela X, os tipos de relações especificados estão sob a forma *A relation B* e *B relation C* para classes ou universais e também *a relation b* e *b relation c* para instâncias ou particulares.

Tabela 35 - Tipos ontológicos das relações *is_a* e *part_of* utilizados na HEMONTO

Tipo da Relação	Descrição	Exemplos	Preenchimento
instance-of (is_a)	Tipo de relação <i>is_a</i> verificada quando se tem um particular a exemplificando um universal A. Quando a é um continuante A também deve ser, se a é um ocorrente A deve também ser.	fresh frozen plasma instance_of blood component albumin instance_of protein	Utilizar a parte do editor de ontologias onde são feitos os preenchimentos de instâncias da classes para indicar esse tipo de relação. No Protégé, utilizar a aba "Individuals"
subclass-of (is_a)	Tipo de relação <i>is_a</i> verificada entre dois universais (A e B), tal que uma entidade é um tipo da outra. Quando A é um continuante B também deve ser, se A é um ocorrente B deve também ser.	cryoprecipitate subclass-of blood component for transfusion	Indicar esse tipo de relação na construção da taxonomia geral da ontologia. No Protégé, utilizar "Add Class" na aba Classes.
same-as (is_a)	Tipo de relação <i>is_a</i> verificada entre duas entidades que são idênticas, podendo ocorrer entre particulares e universais, continuantes e ocorrentes, tal quando a, A são continuantes b, B também devem ser; e se a, A são ocorrentes b, B também devem ser.	leukocyte same_as white blood cell portion of venous blood same_as blood in vein	Indicar esse tipo de relação na propriedade descritiva da classe correspondente aos seus sinônimos. No Protégé apontar os sinônimos das classes nos atributos <i>hasSynonym</i> , <i>hasExactSynonym</i> ou <i>hasRelatedSynonym</i> .
structural-part-of (part_of)	Tipo de relação <i>part-of</i> verificada entre dois continuantes (universal-universal ou particular-particular) na qual a parte compõem a estrutura do todo, funcionalmente ou estruturalmente.	circulatory system structural_part_of human body	Incluir esse tipo como uma subclasse na hierarquia de relações <i>part_of</i> da ontologia. Seu uso na ontologia é feito como qualquer outra relação, no Protégé, como um <i>Object Property</i> .

contained-in (part_of)	Tipo de relação part-of verificada entre dois continuantes (universal-universal ou particular-particular) na qual a parte ocupa uma região espacial 2D que está inserida (dentro) na região espacial 3D ocupada pelo todo.	blood in coronary artery contained_in heart blood cells contained_in portion of plasma	Incluir esse tipo como uma subclasse na hierarquia de relações <i>part_of</i> da ontologia. Seu uso na ontologia é feito como qualquer outra relação, no Protégé, como um <i>Object Property</i> .
located-in (part_of)	Tipo de relação part-of verificada entre dois continuantes (universal-universal ou particular-particular) na qual a parte ocupa uma porção (ou limite menor) do espaço ocupado pelo todo (porção inteira ou todo limite).	portion of blood in capillary located_in portion of blood of human body	Incluir esse tipo como uma subclasse na hierarquia de relações <i>part_of</i> da ontologia. Seu uso na ontologia é feito como qualquer outra relação, no Protégé, como um <i>Object Property</i> .
involved-in (part_of)	Tipo de relação part-of verificada entre dois ocorrentes (universal-universal ou particular-particular) na qual a parte representa uma etapa do todo (processo inteiro).	centrifugation involved_in process for obtaining erythrocytes concentrate	Incluir esse tipo como uma subclasse na hierarquia de relações <i>part_of</i> da ontologia. Seu uso na ontologia é feito como qualquer outra relação, no Protégé, como um <i>Object Property</i> .
member-of (part_of)	Tipo de relação part-of verificada entre dois continuantes tal que a parte é um objeto físico que compõem o todo, o qual é um objeto social não-físico.	erythrocytes member_part_of whole portion of blood	Incluir esse tipo como uma subclasse na hierarquia de relações <i>part_of</i> da ontologia. Seu uso na ontologia é feito como qualquer outra relação, no Protégé, como um <i>Object Property</i> .
constitutes (part_of)	Tipo de relação part-of verificada entre dois continuantes na qual a parte é uma quantidade de matéria que constitui o todo, o qual é um objeto físico.	nutrients constitutes portion of blood water constitutes portion of blood	Incluir esse tipo como uma subclasse na hierarquia de relações <i>part_of</i> da ontologia. Seu uso na ontologia é feito como qualquer outra relação, no Protégé, como um <i>Object Property</i> .
sub-quantity-of (part_of)	Tipo de relação part-of verificada entre dois continuantes, tal que ambos são quantidades de matéria e a parte é uma porção menor de toda a matéria.	platelet subquantity_of platelets concentrate	Incluir esse tipo como uma subclasse na hierarquia de relações <i>part_of</i> da ontologia. Seu uso na ontologia é feito como qualquer outra relação, no Protégé, como um <i>Object Property</i> .
temporary-part-of (part_of)	Tipo de relação part-of verificada entre continuantes ou entre ocorrentes a ser usada quando a parte em algum instante de tempo está localizada como parte do todo, mas não durante toda sua existência.	blood component for transfusion temporary_part_of portion of body substance portion of blood collected by venipuncture temporary-part-of portion of body substance	Incluir esse tipo como uma subclasse na hierarquia de relações <i>part_of</i> da ontologia. Seu uso na ontologia é feito como qualquer outra relação, no Protégé, como um <i>Object Property</i> .
immaterial-part-of (part_of)	Tipo de relação part-of verificada entre dois continuantes, tal que a	lumen of coronary artery immaterial_part_of heart	Incluir esse tipo como uma subclasse na hierarquia de relações

	parte é um objeto imaterial e está conectada ao todo, o qual representa a entidade material.	cavity of ventricle immaterial_part_of heart	part_of da ontologia. Seu uso na ontologia é feito como qualquer outra relação, no Protégé, como um Object Property.
--	--	---	--

Fonte: elaborado pelo autor.

Figura 46 - Hierarquia de tipos da relação part_of na ontologia HEMONTO

The screenshot displays the Protege software interface for the HEMONTO ontology. The left pane shows the 'Object property hierarchy: part_of' with various sub-properties like 'has spatiotemporal occupant', 'has boundary', and 'part_of'. The right pane shows the 'Annotations: part_of' section, including a label 'part_of' and a detailed comment explaining the 'parthood' relation. Below the annotations, the 'Characteristics' section shows that the relation is 'Transitive'. The 'Description: part_of' section shows it is 'Equivalent To' and 'SubProperty Of' 'has_part'.

Fonte: elaborado pelo autor.

Passo 5.7: Definir propriedades das relações ontológicas

De maneira similar ao preenchimento das propriedades das classes, outra tarefa essencial na formalização da ontologia é definir as propriedades das relações ontológicas criadas ou importadas no passo anterior desta metodologia. Tais propriedades incluem tanto aquelas descritivas (por exemplo, o nome da relação, sua definição em linguagem natural, exemplos de uso, anotações) quanto as propriedades lógicas (por exemplo, conjuntos domínio e imagem, relação inversa, propriedades lógicas básicas: simetria, reflexividade, transitividade e funcionalidade).

Portanto, após a criação de uma nova relação ou importação de uma já existente, o desenvolvedor deve preencher todos atributos relativos às propriedades de

uma relação ontológica. Afim de orientar essa tarefa, a metodologia *OntoForInfoScience* explica o preenchimento de cada atributo, descritivo ou lógico, a partir dos atributos para relação incluídos no editor de ontologias Protégé. A Tabela 36, a seguir, descreve tais atributos.

Tabela 36 - Propriedades descritivas e lógicas de uma relação ontológica

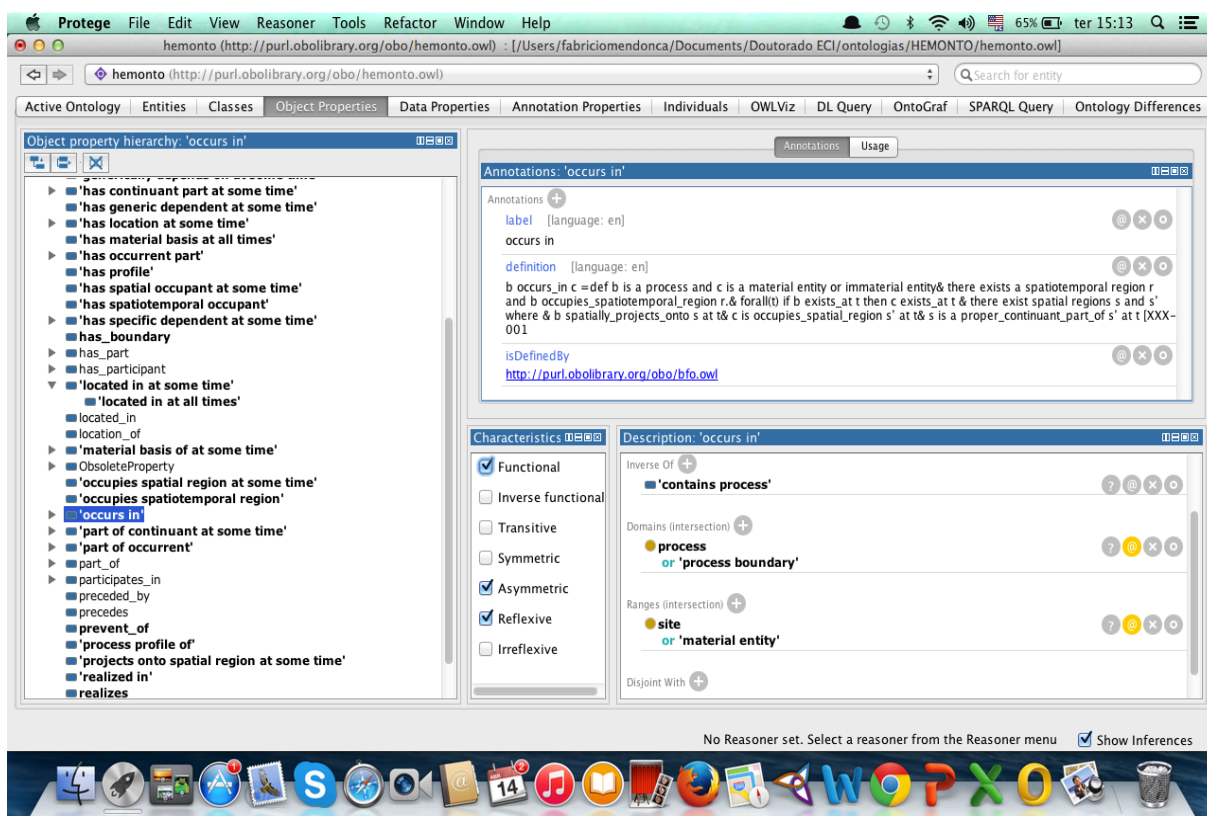
Propriedades Descritivas de uma relação ontológica		
Propriedade	Atributo no Protégé	Forma de preenchimento
Identificador da relação	<i>ID</i>	<ul style="list-style-type: none"> Como esse campo é numérico, deve se usar um número inteiro sequencial para preenchê-lo. Caso a relação seja importada de outra ontologia sugere-se manter o número do ID da relação de origem.
Nome da relação	<i>label</i>	<ul style="list-style-type: none"> Preenchimento do atributo no formato <i>string</i>.
Definição textual ou semi-formal	<i>definition</i>	<ul style="list-style-type: none"> Preenchimento do atributo no formato <i>string</i>, com possibilidade de uso de símbolos da lógica, tais como variáveis e quantificadores.
Elucidação	<i>elucidation</i>	<ul style="list-style-type: none"> Preenchimento do atributo no formato <i>string</i>, com possibilidade de uso de símbolos da lógica, tais como variáveis e quantificadores.
Exemplos de uso	<i>example of usage</i>	<ul style="list-style-type: none"> Preenchimento do atributo no formato <i>string</i>. Quando se tem mais de um exemplo de uso, pode-se enumerar todos esses exemplos em um mesmo atributo separados por vírgula (,).
Ontologia de origem	<i>isDefinedBy</i>	<ul style="list-style-type: none"> Deve-se preenchê-lo com o nome do arquivo da ontologia implementada. Exemplos: GO.owl; FMA.owl.
	<i>hasOBONamespace</i>	<ul style="list-style-type: none"> Deve-se preenchê-lo unicamente com o nome da ontologia, em minúsculo.
Anotações (Comentários)	<i>editor note</i>	<ul style="list-style-type: none"> Preenchimento do atributo no formato <i>string</i>. Quando se tem mais de um comentário da classe, sugere-se que cada comentário possa fazer parte de um novo atributo 'comentário' não utilizando vírgulas para separar vários comentários.
	<i>comment</i>	
Propriedades Lógicas de uma relação ontológica		
Propriedade	Atributo no Protégé	Forma de preenchimento
Propriedades lógicas básicas: - Funcionalidade - Reflexividade - Transitividade - Simetria	<i>Characteristics:</i> - <i>Functional</i> - <i>Inverse functional</i> - <i>Transitive</i> - <i>Symmetric</i> - <i>Asymmetric</i> - <i>Reflexive</i> - <i>Irreflexive</i>	<ul style="list-style-type: none"> Definição: corresponde às características lógicas ou propriedades básicas da lógica, que permite caracterizar o tipo da relação de acordo com as instâncias que ela mapeia. Tipos de propriedades suportados: Funcional, Inversa Funcional, Transitiva, Simétrica, Assimétrica, Reflexiva e Irreflexiva. Deve-se analisar a relação que está sendo definida e, a partir dessa análise, marcar (no editor de ontologias) os tipos de propriedades que a caracteriza.
Funcionalidade		<ul style="list-style-type: none"> Uma relação é funcional para um determinado indivíduo (particular) quando só pode haver no máximo um indivíduo a ser relacionado através desta relação.
Reflexividade		<ul style="list-style-type: none"> Uma relação é reflexiva se as entidades relacionadas são idênticas e a relação é válida entre elas.
Transitividade		<ul style="list-style-type: none"> Uma relação é transitiva quando é possível inferir essa mesma relação em uma cadeia de entidades relacionadas.
Simetria		<ul style="list-style-type: none"> Uma relação é simétrica quando ela é verdadeira entre as entidades relacionadas nos dois sentidos (direita-esquerda, esquerda-direita).
Definição Formal	<i>definition</i>	<ul style="list-style-type: none"> Definição formal da relação, especificando suas características próprias com o uso da lógica da OWL.

		<ul style="list-style-type: none"> • Preenchimento do atributo usado a lógica OWL com as variáveis e quantificadores próprios.
Relação inversa	<i>Inverse Of</i>	<ul style="list-style-type: none"> • Esse atributo deve ser preenchido com o nome (<i>label</i>) da relação que é inversa à relação que está se definindo. • Preenchimento do atributo no formato <i>string</i>.
Domínio	<i>Domains</i>	<ul style="list-style-type: none"> • Definição: propriedade que mapeia instâncias de uma classe de domínio para instâncias de uma classe de imagem. • Correspondem às classes da ontologia que podem ser relacionadas através desta relação, desde que tais classes façam parte do domínio da relação. • Deve-se preencher esse campo unicamente com o(s) nome(s) da(s) classe(s) (<i>label</i>) do domínio.
Imagens	<i>Ranges</i>	<ul style="list-style-type: none"> • Definição: propriedade que mapeia instâncias de uma classe de domínio para instâncias de uma classe de imagem. • Correspondem às classes da ontologia que podem ser relacionadas através desta relação, desde que tais classes façam parte da imagem da relação. • Deve-se preencher esse campo unicamente com o(s) nome(s) da(s) classe(s) (<i>label</i>) da imagem.

Fonte: elaborado pelo autor.

Considerando a construção da ontologia HEMONTO pela metodologia proposta, apresenta-se, a seguir, um exemplo de definição das propriedades da relação ontológica “occurs in”, através do software Protégé 4.3 (veja Figura 47). As demais relações desta ontologia e suas propriedades descritivas e lógicas são apresentadas no capítulo de documentação da ontologia.

Figura 47 - Preenchimento das propriedades da relação "occurs_in" no Protégé 4.3



Fonte: elaborado pelo autor.

7.6 Avaliação da ontologia (Etapa 6)

Após o desenvolvimento ontológico em si, a etapa seguinte do projeto ontológico dentro da metodologia *OntoForInfoScience* corresponde à avaliação da ontologia segundo um conjunto de parâmetros e critérios avaliativos. O objetivo desta etapa é checar as estruturas ontológicas desenvolvidas e sugerir melhorias que a torne uma representação mais aproximada possível da realidade do domínio que ela representa.

Relembrando a definição de Gómez-Perez (2004), a avaliação ontológica é uma técnica de julgamento do conteúdo da ontologia quanto a sua adequação à especificação de requisitos, questões de competência e compromisso com o mundo real, durante cada fase de seu desenvolvimento e entre as fases de seu ciclo de vida. Nesse processo de julgamento do conteúdo ontológico identifica-se uma *fase de validação*, na qual a ontologia é avaliada em função de sua adequação à realidade do domínio que representa, e uma *fase de verificação*, na qual se avalia se a ontologia está sendo construída de maneira correta de acordo com suas restrições e princípios ontológicos básicos.

Na metodologia de construção proposta aos cientistas da informação, foram especificados um conjunto de critérios avaliativos do conteúdo da ontologia, que envolvem tanto a validação quanto a verificação ontológicas. A definição desses critérios foi feita tendo como base algumas abordagens relevantes na literatura da área sobre avaliação de ontologias, as quais envolvem: (i) os princípios ontológicos da teoria que fundamenta a BFO (MUNN e SMITH, 2008); (ii) critérios de avaliação ontológica para caracterizar os tipos de relação e validar suas propriedades lógicas básicas (MENDONÇA et al., 2015); (iii) critérios para identificação de pitfalls no conteúdo da ontologia (POVEDA-VILLALÓN, 2010); e (iv) regras das etapas de avaliação ontológica das metodologias *NeOn Methodology*, *Methontology* e *método 101*.

Antes de passar à descrição dos critérios avaliativos da metodologia *OntoForInfoScience*, é importante registrar que o propósito desta etapa não envolve a concepção de um framework formal e também um ambiente de desenvolvimento para avaliação ontológica, como ocorre em algumas outras metodologias e métodos (*OntoClean*, *NeOn*, *OntoLingua*⁵⁴ (FARQUHAR, FIKES e RICE, 1996), *OntoKEM*⁵⁵, *Methontology*). Na proposta apresentada, a especificação de um conjunto de critérios consistentes para avaliação da ontologia mostrou-se suficiente para auxiliar desenvolvedores e especialistas do domínio nessa tarefa.

Considerando as *fases de validação e verificação* na avaliação da ontologia, foram definidos dois subconjuntos de *critérios avaliativos*, cada um pertencente a uma destas duas fases. Além disso, os critérios de cada um dos subconjuntos foram agrupados segundo uma característica a ser avaliada na ontologia, como, por exemplo, a consistência de seus termos, aspectos de documentação, completude, expandibilidade da ontologia, entre outras. Essas características avaliadas foram chamadas na metodologia de *parâmetros de avaliação*. Tais parâmetros e seus critérios avaliativos correspondentes são apresentados na Tabela 37, a seguir, que também inclui uma descrição explicativa de cada critério para seu uso na metodologia proposta.

⁵⁴ Disponível em: <http://www.ksl.stanford.edu/software/ontolingua/>. Acesso em: 29 de Abril de 2015.

⁵⁵ *OntoKEM* é uma ferramenta para engenharia do conhecimento que suporta um processo de construção e documentação de ontologias, baseado no processo de desenvolvimento do método 101 e nos artefatos documentais das metodologias para ontologias *On-to-Knowledge* e *Methontology*. Disponível em: <http://ontokem.egc.ufsc.br/ProjetoOntologias/>. Acesso em 29 de Abril de 2015.

Tabela 37 - Parâmetros e critérios avaliativos da metodologia OntoForInfoScience

Critérios de Validação: adequação ao domínio (mundo real)		
Parâmetro de avaliação	Critério avaliativo	Descrição
Compromisso ontológico	Fundamentação ontológica	Analisar se a ontologia desenvolvida faz uso de alguma ontologia de fundamentação e, em caso positivo, verificar se as classes e relações da ontologia de fundamentação foram importadas e estão sendo utilizadas na ontologia.
	Natureza ontológica	Avaliar se as classes e relações criadas são genuinamente ontológicas, isto é, apenas devem existir classes e relações identificadas entre entidades na realidade, independente da nossa experiência ou métodos de aprendizagem humanos.
	Não intuitividade	Analisar se todas classes e relações da ontologia foram definidas de maneira única e rigorosa, tal que as definições criadas não podem envolver a intuitividade humana.
	Universalidade de classes	Validar se as classes criadas são obtidas universalmente do domínio, isto é, uma determinada classe C representa ou engloba todas as instâncias possíveis do conceito C que ela representa e não apenas algumas instâncias representativas estatisticamente de C.
	Universalidade de relações	Validar se as relações criadas são obtidas universalmente do domínio, isto é, uma declaração na forma <i>A relation B</i> deve ser obtida para todas as instâncias de A e não obtida apenas a partir de algumas instâncias representativas estatisticamente de A.
	Formalização das definições das entidades	Analisar se todas as classes e relações da ontologia possuem definições formais para elas em alguma linguagem decidível – interpretada pelas máquinas.
	Definição de propriedades	Checar para cada classe e relação da ontologia se foram definidas propriedades descritivas suficientes para seu entendimento e propriedades lógicas suficientes para a realização de inferências.
Especificação	Resposta às questões de competência	Checar se todas questões de competência propostas na especificação da ontologia podem respondidas a partir do conteúdo ontológico desenvolvido, ou seja, consultar as respostas das questões de competência. Este critério é um forte indicativo de que a ontologia foi capaz de apreender o conhecimento no domínio explorado.
	Grau de representatividade do domínio	Este critério é um indicativo para mensurar a relevância da representação do domínio na ontologia construída. Uma das formas de medi-lo é avaliar a abrangência dos documentos de referência utilizados como fonte de conhecimento e outra forma é avaliar a utilização prática da ontologia por usuários e especialistas do domínio.
Validação especializada	Consulta aos especialistas	A participação de especialistas do domínio, em geral, ocorre em algumas etapas anteriores do desenvolvimento ontológico e, de alguma forma, deve acontecer também na fase de validação da ontologia. Assim, deve ser verificada na documentação se houve o envolvimento de especialistas no processo de validação.
	Grau de participação	Mensurar o grau de participação dos especialistas na validação ontológica e, caso necessário, equilibrar ou intermediar essa participação. O grau de participação pode ser medido pelas estruturas ontológicas avaliadas pelos especialistas (conjunto de termos, taxonomias, propriedades, etc.) e pelo tempo gasto com a validação. Se esse tempo for julgado excessivo, na maioria das vezes provocado pelo não consenso entre os

		especialistas, o desenvolvedor deve intervir e regular essa participação para se atingir algum consenso.
Expandibilidade	Capacidade de atualização	Analisar, através da documentação da ontologia, se são previstas formas de atualização da mesma, para que seu conteúdo reflita as alterações ocorridas na realidade do domínio.
	Capacidade de integração	Essa é uma medida das possibilidades ou facilidades de integração da ontologia desenvolvida. A capacidade de integração está diretamente relacionada com o grau de reutilização da ontologia. Assim, quanto mais classes e relações reutilizadas de outras ontologias relevantes no domínio, maiores possibilidades de integração. No caso de ontologias biomédicas, são recomendadas reutilizações das ontologias vinculadas ao repositório OBO, o que facilita a integração com este repositório e muitas das ontologias biomédicas existentes.
Critérios de Verificação: correteza ontológica		
Parâmetro de avaliação	Critério avaliativo	Descrição
Completeza	Única definição textual	Verificar se cada classe e cada relação da ontologia possuem uma única definição textual.
	Única definição formal	Verificar se cada classe da ontologia possui uma única definição formal derivada de sua definição textual.
	Definição de propriedades descritivas	Checar em cada classe e relação da ontologia se suas propriedades descritivas, mencionadas na metodologia proposta, foram definidas.
	Caracterização das propriedades básicas da relação	Para cada relação ontológica, verificar se suas propriedades básicas (reflexividade, funcionalidade, simetria e transitividade) foram caracterizadas.
Integridade	Conectividade com o domínio e contexto	Checar se as classes, relações ou propriedades da ontologia possuem algum tipo de relação com o restante (conteúdo) da ontologia, isto é, estão conectadas ao domínio do conhecimento tratado e ao contexto de aplicação.
	Definição de equivalência	Verificar se foram definidas como equivalentes classes e relações desta ontologia que possuem tal propriedade. Esse critério deve ser aplicado após a importação de todas as classes e relações para a ontologia, já que a equivalência costuma aparecer após a importação de termos de outras ontologias.
	Definição de disjunção	Analisar se foram definidos axiomas de disjunção para classes que possuem tal propriedade.
Integridade e Completeza	Definição de domínio e imagem da relação	Verificar para cada relação ontológica se foram definidas quais classes são suportadas no domínio e quais são suportadas na imagem.
	Definição de relação inversa	Checar se em toda relação da ontologia foi definida sua relação inversa correspondente.
	Definição de tipos	Analisar nas propriedades de dados ("data properties") de cada classe da ontologia se foi definido o tipo de valor suportado pela classe (string, numérico, booleano, etc.)
	Criação de cardinalidade	Verificar se foram definidas as cardinalidades de cada relação ontológica e também as cardinalidades dos valores possíveis de uma classe, em suas propriedades de dados.
Consistência	Não polissemia	Verificar se não foram incluídos termos polissêmicos na ontologia, isto é, se não há uma classe, relação ou propriedade com o mesmo nome e que tem mais de um significado ou conceito.
	Ausência de ciclos na hierarquia	Analisar na taxonomia geral (hierarquia) da ontologia se não existem duas classes A e B, tal que A tem B como

		subclasse e ao mesmo tempo B é também superclasse de A.
	Ausência de classes variadas ou genéricas	Avaliar se não existe na ontologia nenhuma classe que represente dois ou mais conceitos diferentes sob o mesmo rótulo. Esse tipo de classe, chamada de genérica ou variada, que mescla dois ou mais conceitos não pode ser criado em uma ontologia.
	Não recursividade na definição	Checar em cada termo (classe ou relação) da ontologia se o próprio termo não é usado em sua própria definição, caracterizando uma recursividade.
	Consistência na definição da relação inversa	Para cada relação inversa da ontologia, verificar se de fato ela é inversa a outra relação, ou seja, não foi feita uma definição errada de relação inversa; e também verificar se ela não foi definida como uma relação inversa a si mesmo.
Precisão	Definição de sinônimos nas classes	Analisar as classes da ontologia, tal que conceitos sinônimos do mundo real devam ser representados através de uma única classe com o sinônimo indicado como atributo da classe e não como uma classe diferente.
	Unicidade dos rótulos (identificadores) das classes	Checar se cada classe da ontologia possui um rótulo (label) ou identificador único. Rótulos repetidos não são aceitos em uma ontologia.
	Não criação de instâncias como classes	Analizando o domínio e o contexto de aplicação da ontologia, verificar se não existem instâncias (particulares) criadas como classes. Isso faz com que a hierarquia se torna tão especializada, que as classes que são nós-folha não tem instâncias, pois elas são as próprias instâncias, e não é permitido.
	Especificação dos diferentes tipos da relação "is_a"	Analisar se os tipos "subClassOf", "instanceOf" e "sameAs" da relação "is_a" estão sendo tratados de forma distinta na ontologia, tal que o primeiro é indicado na hierarquia de classes, o segundo na definição de instâncias e o terceiro na definição de sinônimos.
	Especificação dos diferentes tipos da relação "part_of"	Checar na hierarquia de relações da ontologia se os diferentes tipos da relação "part_of", tais como: temporary_part_of; member_of; contained_in; entre outros, estão incluídos de alguma forma, sejam como subrelações "part_of" ou como uma relação na hierarquia.
Documentação	Existência de um documento formal da ontologia	Verificar se existe algum documento formal que descreva a ontologia.
	Único critério de nomeação dos termos da ontologia	Avaliar se foi adotado um critério para nomeação dos termos da ontologia e se ele é o único adotado.
	Anotação de comentários	Verificar se existem comentários anotados para algumas classes ou relações da ontologia e se eles foram anotados no atributo correspondente. Em algumas ontologias já desenvolvidas, comentários foram inseridos como rótulos das classes, o que é extremamente incorreto.

Fonte: elaborado pelo autor.

O exemplo de aplicação de tais critérios na avaliação de uma ontologia é apresentado na documentação formal da ontologia HEMONTO, no capítulo 8 (oito) desta tese. Além disso, é importante registrar que os atuais editores de ontologia, tal como o Protégé 4.3, fornecem recursos que auxiliam na identificação de problemas de integridade, imprecisão e inconsistências no conteúdo ontológico. Tais editores, utilizam seus classificadores (*reasoners*) para identificação destes problemas, como é o caso do

FaCT++ e o HerMiT do Protégé 4.3. Por isso, uma última recomendação da metodologia *OntoForInfoScience* aos desenvolvedores para a avaliação de suas ontologias é o uso dos classificadores (*reasoners*) nos editores de ontologia.

7.7 Documentação da ontologia (Etapa 7)

Concluídas as seis etapas anteriores da metodologia proposta, a etapa seguinte consiste em documentar todas as informações contidas na ontologia através de um documento formal em linguagem natural. Essa etapa é indispensável em um projeto ontológico, pois registra tudo aquilo que foi realizado, como foi e porquê foi, facilitando a consulta e a posterior atualização da ontologia. Schiessl e Bräscher (2011), por exemplo, destacam que a documentação associada aos termos representados na ontologia é particularmente importante não apenas para melhorar sua clareza, mas também para facilitar sua manutenção, utilização e reutilização.

O aspecto mais importante nessa etapa da metodologia é mencionar os elementos essenciais ao documento formal da ontologia desenvolvida. Nesse sentido, a metodologia *OntoForInfoScience* propôs um template com os elementos essenciais na documentação da ontologia, conforme mostrado no Quadro 3, a seguir. No template proposto é importante destacar que cada uma das etapas anteriores produz um resultado que se torna um elemento na documentação. Artefatos de representação preliminares ou parciais no processo de desenvolvimento ontológico, tais como o Pré-Glossário de Termos, Conjuntos de Candidatos à ontologia e Tabela de Conceitos e Valores, não são incluídos no template e na documentação da ontologia, porque eles se transformam, ao longo do processo de desenvolvimento, nos artefatos de representação finais.

Quadro 3 - Template de Documentação de uma ontologia

Etapa 1: Documento de Especificação
<ul style="list-style-type: none"> - Domínio/Escoço da ontologia - Propósito geral - Classes de usuários - Uso pretendido - Tipo da ontologia - Grau de formalidade - Questões de competência
Etapa 2: Documentos de referência
<ul style="list-style-type: none"> - Relação dos documentos do domínio tratado utilizados como materiais de referência para estudo e conceitualização do domínio.

Etapa 3: Modelos conceituais
- Conjunto de modelos conceituais desenvolvidos do domínio, que tenham sido aceitos e compartilhados pelos grupos envolvidos nesta etapa.
Etapa 4: Ontologias reutilizadas
- Relação e breve descrição das ontologias reutilizadas no desenvolvimento, que inclua a(s) ontologia(s) de fundamentação usada(s) como ponto de partida na construção.
Etapa 5: Conteúdo ontológico
<ul style="list-style-type: none"> - Taxonomia geral da ontologia - Dicionário de classes, que inclua como elementos: as definições textual e formal de cada classe, suas propriedades descritivas e lógicas, e referências às classes importadas de outras ontologias. - Dicionário de relações ontológicas, que inclua como elementos: as definições textual e semi-formal de cada relação, suas propriedades descritivas e lógicas, e referências às relações importadas de outras ontologias. - Estruturas gráficas de representação da ontologia, que apresentem em modo gráfico (visual) os relacionamentos existentes entre as classes da ontologia, tais como taxonomias, partonomias e outras estruturas.
Etapa 6: Métricas de avaliação
<ul style="list-style-type: none"> - Conjunto de critérios, métodos e técnicas usados na avaliação (validação e verificação) do conteúdo ontológico. - Conjunto de respostas às questões de competência propostas à ontologia.

Fonte: elaborado pelo autor.

Como exemplo de documentação formal com os elementos citados, a presente pesquisa apresenta no capítulo seguinte (capítulo 8), a documentação da ontologia desenvolvida sobre hemocomponentes e hemoderivados do sangue humano – HEMONTO.

7.8 Disponibilização da ontologia (Etapa 8)

Como última etapa na metodologia de construção proposta, tem-se a disponibilização da ontologia em um algum meio eletrônico de fácil acesso e visualização aos usuários desta ontologia. Esse é uma medida fundamental para disseminar o conhecimento representado na ontologia e torná-la um instrumento útil para organização e uso da informação do domínio que ela representa.

Foram enumeradas duas atividades principais nesta etapa de disponibilização da ontologia, as quais envolvem: (i) a forma de geração e representação da ontologia em alguma linguagem lógica que suporte seus recursos; e (ii) a apresentação da ontologia em

formato e meio eletrônicos acessíveis aos seus usuários. Considerando estas duas atividades principais, os parágrafos seguintes descrevem cada uma delas como passos da etapa de disponibilização da ontologia na metodologia *OntoForInfoScience*.

8.1 - Gerar a representação da ontologia em uma linguagem lógica

Conforme mencionado anteriormente, um dos aspectos importantes em um projeto ontológico diz respeito ao uso de uma linguagem formal para implementação da ontologia desenvolvida. Em geral, utilizam-se linguagens lógicas para a representação formal da ontologia, uma vez que elas permitem a interpretação pelas máquinas e a realização de inferências ao conteúdo ontológico. Esse passo da metodologia trata da representação formal da ontologia em alguma linguagem lógica e, assim, computável, de forma que o conteúdo ontológico possa ser interpretado pelas máquinas.

Dentre as linguagens lógicas existentes, o desenvolvedor deve optar pela linguagem que mais se adequa aos objetivos pretendidos com a ontologia desenvolvida. Linguagens lógicas descritivas, como a OWL, são mais favoráveis a aplicações computacionais, devido ao seu maior poder de decidibilidade e menor expressividade; enquanto linguagens lógicas mais complexas, como a lógica de primeira-ordem (*First Order Logical-FOL*), são mais expressivas, permitindo a representação de praticamente todas as informações ontológicas, porém são menos decidíveis do ponto de vista de implementação. Diante dessa diferença, cabe ao desenvolvedor selecionar qual o tipo de linguagem lógica mais adequada aos objetivos e aplicações da ontologia, dando mais ênfase à decidibilidade ou à expressividade.

Uma boa dica ao ontologista nessa escolha é pesquisar em projetos ontológicos similares e na literatura da área do domínio tratado quais linguagens têm sido mais usadas. Nos domínios biomédicos, por exemplo, as linguagens de lógica descritiva tem sido preferíveis em relação às linguagens de primeira-ordem. Além disso, linguagens formais como a OWL possuem diferentes versões (OWL-Full, OWL-DL, OWL-Lite) para atender necessidades diferentes de cada projeto ontológico.

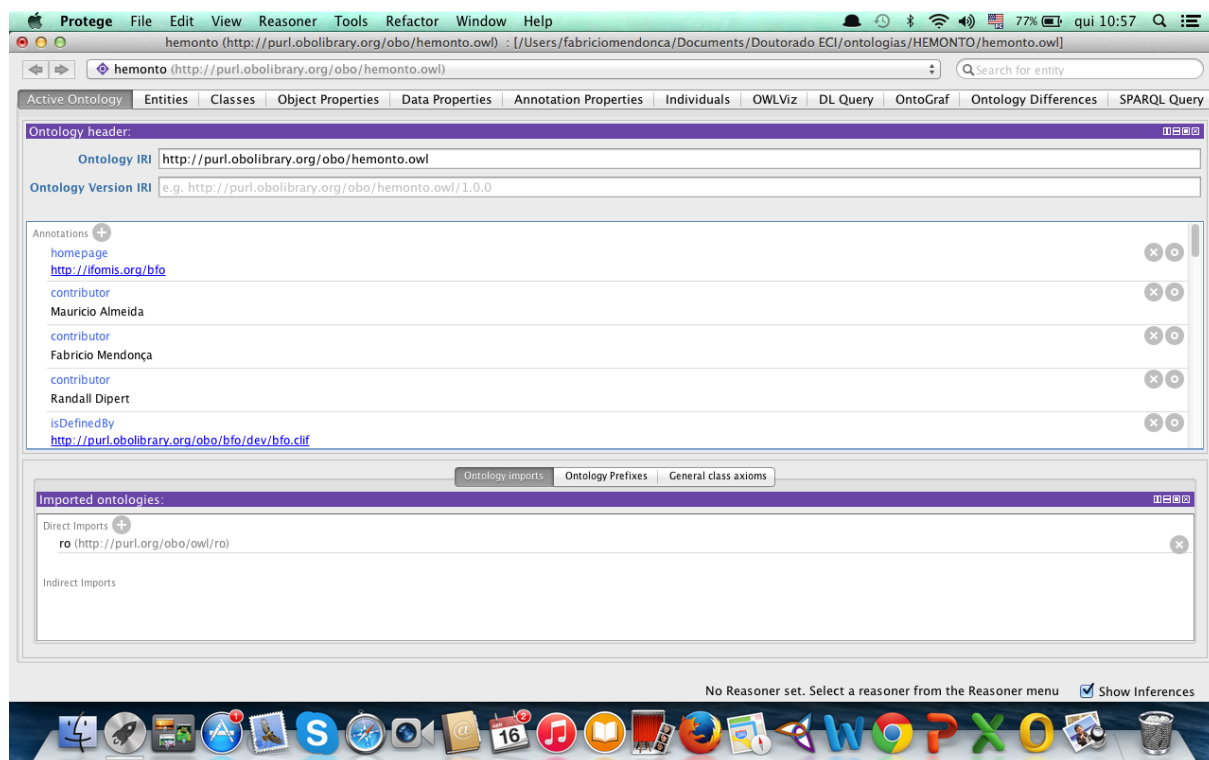
No desenvolvimento da ontologia sobre o sangue humano (HEMONTO) optou-se pelo uso da linguagem OWL/XML para sua representação formal. Essa escolha foi motivada, em primeiro lugar, pela aceitação desta linguagem em domínios biomédicos e, em segundo lugar, pelas facilidades propiciadas pelo editor de ontologias Protégé, no qual já há um direcionamento para a representação da ontologia nessa linguagem à medida que vai se construindo a ontologia. OWL/XML é o formato (extensão) padrão do arquivo gerado das ontologias desenvolvidas no Protégé 4.3. Além deste, são também suportados os formatos: RDF/XML, OWL Functional Syntax, Manchester OWL Syntax e OBO Format.

8.2 – Apresentar a ontologia em formato e meio eletrônico

No passo anterior foi citado como os atuais editores de ontologias, tais como: o Protégé, Onto-Edit, OBO-Edit, entre outros, possibilitam ao desenvolvedor, com relativa facilidade, gerar uma representação formal da ontologia desenvolvida em um formato computável, como o das linguagens lógicas. Esses formatos de possível interpretação pelas máquinas são atualmente utilizados como recursos na *web*. Portanto, nesse passo, a metodologia *OntoForInfoScience* busca alertar o desenvolvedor no sentido de publicar sua ontologia em um meio eletrônico de fácil acesso como a *web*.

Usando algum editor de ontologias, como o próprio Protégé, o desenvolvedor pode criar um identificador de acesso à ontologia – um *Internationalized Resource Identifier* (IRI) - e através dele é possível fazer referência ao conteúdo desta ontologia pela *web*. Além disso, o desenvolvedor pode criar uma homepage de acesso à ontologia na *web* e disponibilizar aos seus usuários. Em termos práticos, o desenvolvedor deve usar a aba “Active Ontology” e a função “Ontology header” do Protégé 4.3, conforme mostrado na Figura 48 abaixo, para preencher essas informações: IRI da ontologia, homepage, colaboradores da ontologia – atributo “contributor”, entre outras.

Figura 48 - Preenchimento do IRI de acesso à ontologia, homepage e anotações gerais



Fonte: elaborado pelo autor.

Incrementado um pouco mais a homepage da ontologia desenvolvida, uma sugestão recomendada pela metodologia é a criação de uma interface de busca aos termos da ontologia, que permita ao seu usuário navegar pela base de conhecimento (taxonomias, classes e relações, propriedades), pesquisando por termos de seu interesse. Essa tarefa é uma maneira de facilitar ainda mais o acesso ao conteúdo ontológico, permitindo uma maior disseminação da ontologia, avaliação de sua aplicabilidade prática e reutilização de seus termos em outros projetos ontológicos.

No projeto de construção da ontologia HEMONTO, por questões de tempo e escopo, não foram criadas uma interface de busca ao conteúdo ontológico e nem outro sistema informatizado de acesso a sua base de conhecimento. Entretanto, essa tarefa está incluída como trabalhos futuros da presente pesquisa.

Ainda sobre a HEMONTO, no que diz respeito a sua disponibilização, o recurso utilizado para acesso dinâmico de seus usuários ao conteúdo da ontologia foi o uso do plug-in do Protégé OWLDoc⁵⁶. Através do OWLDoc foi possível obter uma visão geral e dinâmica de todas as estruturas ontológicas incluídas na HEMONTO e, através do uso do formato *.doc tornou-se mais fácil a visualização por parte de especialistas do domínio e também de outros usuários. O uso do plug-in OWLDoc é uma solução recomendada pela metodologia OntoForInfoScience para documentação e apresentação da ontologia desenvolvida.

⁵⁶ Disponível em: <http://protegewiki.stanford.edu/wiki/OWLDoc>. Acesso em: 16 de Abril de 2015.

8 HEMONTO: ontologia sobre os componentes do sangue humano

De acordo com a metodologia de construção proposta e as demais metodologias pesquisadas, uma etapa fundamental e indispensável em um projeto ontológico corresponde à documentação formal da ontologia, onde se deve registrar tudo aquilo que foi realizado no seu processo de construção, de forma a melhorar sua clareza e facilitar sua utilização, manutenção e reutilização.

Para cumprir com tais propósitos, optou-se por separar em um capítulo único toda a documentação formal da ontologia sobre os componentes do sangue humano (HEMONTO). Portanto, o presente capítulo traz a documentação formal da HEMONTO, dividindo-o de acordo com o template de documentação proposto na metodologia *OntoForInfoScience*, o qual inclui as seguintes partes: (i) documento de especificação (resultado da etapa 1); (ii) documentos de referência (etapa 2); (iii) modelos conceituais (etapa 3); (iv) ontologias reutilizadas (etapa 4); (v) conteúdo ontológico (etapa 5); e (vi) métricas de avaliação (etapa 6). Para cada uma das partes da documentação tem-se uma seção correspondente no presente capítulo, as quais são descritas nos parágrafos seguintes.

8.1 Documento de Especificação

A especificação dos requisitos da ontologia HEMONTO é dada pelo template de especificação (Quadro 4), abaixo, que reúne os elementos obrigatórios nessa etapa da metodologia *OntoForInfoScience*.

Quadro 4 - Template de Especificação da ontologia HEMONTO

Domínio/Esopo Geral
A HEMONTO é uma ontologia biomédica de domínio que representa o conhecimento relativo ao sangue humano, especificamente os componentes do sangue que são utilizados para fins terapêuticos, denominados de hemocomponentes e hemoderivados. Seu escopo geral abrange os elementos constituintes dos hemocomponentes e hemoderivados do sangue humano, bem como os processos de obtenção de tais produtos, além de aspectos fisiológicos do sangue.
Propósito Geral
A HEMONTO tem como propósito geral ser um instrumento de apoio a profissionais das áreas de hematologia e hemoterapia no exercício de sua prática clínica, explicitando-lhes as características de cada um dos produtos derivados do sangue, bem como dos processos utilizados na obtenção de tais produtos. Esse apoio visa a informar esses profissionais sobre os procedimentos adequados na extração, manipulação e armazenamento dos componentes do sangue utilizados no tratamento

hemoterapêutico dos pacientes.
Classes de usuários
A ontologia HEMONTO é destinada a profissionais que trabalham nas áreas de hematologia e hemoterapia, incluindo médicos, técnicos, laboratoristas e profissionais de sistemas de informação, além de cientistas e pesquisadores que atuam em domínios biomédicos.
Uso pretendido
O uso geral pretendido com a ontologia HEMONTO é no suporte à separação dos componentes físicos do sangue humano de uma forma apropriada, tal que esses componentes possam ser extraídos do sangue das pessoas, manipulados e armazenados de acordo com as normas e especificações da área, garantindo assim a eficácia de seu uso para tratamentos hemoterapêuticos.
Tipo da ontologia
A HEMONTO foi classificada como uma <i>ontologia de domínio (biomédico)</i> , <i>ontologia com médio rigor formal</i> e <i>ontologia para sistemas de informação</i> .
Grau de formalidade
A HEMONTO possui um grau de formalidade médio (médio rigor formal), incluindo um conjunto de axiomas criados na definição de cada classe, relação e propriedade da ontologia. A HEMONTO está representada na linguagem OWL-DL, também considerada de médio rigor formal.
Delimitação do escopo de cobertura
<p>- Ponto de partida: entidades do mundo real, categorizados por meio das classes da ontologia BFO, e o termo “sangue” (FMA: <i>portion of blood</i>). A partir do termo FMA: <i>portion of blood</i> são obtidas as especializações do sangue (ou seja, seus componentes e tipos) e as generalizações (ou seja, o “sangue” compreendido como parte do corpo humano).</p> <p>- Limite do domínio coberto: parte-se do termo “sangue” e, a partir de uma abordagem híbrida que combina a estratégia <i>top-down</i> com a estratégia <i>bottom-up</i>, define-se todos os componentes, tipos e derivados do sangue humano até atingir o limite de componentes como proteínas e enzimas. Daí para frente, onde temos enzimas, genes e componentes celulares, faz-se referência a outras ontologias que definem tais termos, como a GO, a ChEBI e a Cell Ontology, não os definindo na ontologia HEMONTO.</p> <p>Questões de competência:</p> <p>QC1. Quais os elementos constituintes de uma porção de sangue e em que proporção são encontrados?</p> <p>QC2. Quais proteínas são encontradas numa porção de plasma sanguíneo?</p> <p>QC3. Quais os tipos de leucócitos podem ser encontrados numa porção de sangue?</p> <p>QC4. Quais parâmetros básicos do sangue são relevantes para o estudo de suas características fisiológicas?</p> <p>QC5. Quais diferenças existentes entre os hemocomponentes e hemoderivados do sangue humano?</p>

QC6. Quais os tipos de hemocomponentes e hemoderivados podem ser obtidos a partir de uma unidade de sangue total?

QC7. Quais as condições de temperatura e armazenamento o plasma sanguíneo deve ser submetido para obtenção de hemocomponentes plasmáticos?

QC8. Quais elementos do sangue, quando processados, dão origem ao hemocomponente crioprecipitado?

QC9. Quais glicoproteínas, que cumprem o papel de fatores de coagulação, fazem parte do crioprecipitado?

QC10. Quais elementos estão envolvidos num processo de punção venosa?

QC11. Quais são os tipos de processos de análise do sangue?

QC12. Que tipos de células patológicas existem no sangue humano?

Fonte: elaborado pelo autor.

8.2 Documentos de Referência

Os documentos de referência sobre o sangue utilizados no desenvolvimento da ontologia HEMONTO compreendem normas e manuais técnicos, além de publicações científicas atuais, contendo os avanços recentes na área. São eles:

- ✓ **Guia para o uso hemocomponentes** do Ministério da Saúde (BRASIL, 2008);
- ✓ **ISBT 28 Standard Terminology for Blood, Cellular Therapy and Tissue Product Descriptions** - versão 3.33 (ICCBBA, 2010);
- ✓ **AABB's Technical Manual** 17ª edição (AABB, 2005);
- ✓ **Wintrobe's Clinical Hematology** 12ª edição (GREER et al., 2009).

O “Guia para o uso de Hemocomponentes” aborda diretrizes na visão do médico, funcionando como um instrumento de apoio para a prescrição médica referente à escolha do momento e do hemocomponente mais adequado para a transfusão (BRASIL, 2008). Especificamente, segundo seus autores, tal documento engloba o arsenal terapêutico comumente usado no suporte hemoterápico, as indicações e contra-indicações dos hemocomponentes, a transfusão pediátrica, as principais condutas frente às reações transfusionais imediatas e o papel dos comitês transfusionais na prática hemoterápica. Dentre os conteúdos incluídos nesse guia optou-se, na presente pesquisa, por abordar os produtos que compõem os hemocomponentes e hemoderivados, bem como os

procedimentos empregados na obtenção de tais componentes do sangue humano, para a construção da ontologia proposta.

O “Padrão Terminológico ISBT 28 para o sangue, terapia celular e descrição dos produtos dos tecidos” foi desenvolvido para permitir uma distinção entre produtos, que são requeridos nas práticas clínicas, no cuidado à saúde e grupos de gerenciamento (ICCBBA, 2010). Nesse sentido, tal terminologia padrão engloba um conjunto de termos especializados de entendimento internacional comum nos domínios do sangue e de terapia celular.

O “17º Manual Técnico da associação AABB” é uma fonte de informação sobre banco de sangue, transfusão e terapia celular. A AABB é uma associação internacional que desempenha o papel de desenvolver e compartilhar padrões e programas educacionais visando otimizar as ações e prover segurança aos pacientes e doadores no contexto das áreas de hematologia e hemoterapia (AABB, 2011). Segundo dados atuais do seu site⁵⁷, a AABB conta com cerca de 2000 instituições de saúde e 8000 profissionais (físicos, enfermeiros, cientistas, pesquisadores, administradores, médicos, entre outros) vinculados, os quais são originados de cerca de 80 diferentes países do mundo, e que corroboram para a produção e atualização de informações nos manuais técnicos desta associação. No que se diz respeito a presente pesquisa, foi utilizado do manual técnico da AABB a parte relativa à prática da transfusão de sangue, que, nesta 17ª edição publicada, corresponde ao capítulo 21.

O livro-texto “*Wintrobe’s Clinical Hematology*” engloba um conteúdo rico e bem detalhado na área de hematologia, que vai desde o histórico passado da área até os desenvolvimentos recentes em hematologia, ligados com seu uso na internet e com as tendências futuras (GREER et al., 2009). Cada capítulo do livro-texto trata de um tópico específico da área de hematologia, tais como: abordagem terapêutica para problemas hematológicos, biologia molecular e hematologia, coagulação do sangue, transfusão de sangue, terapia genética para doenças hematológicas, entre outros. Dedicado aos profissionais médicos e cientistas da área de hematologia, o *Wintrobe’s Clinical Hematology* possibilitou o entendimento dos produtos e processos envolvidos em transfusões de sangue e, conseqüentemente, na obtenção e composição dos hemocomponentes do sangue humano, que são descritos no capítulo 23 deste livro-texto.

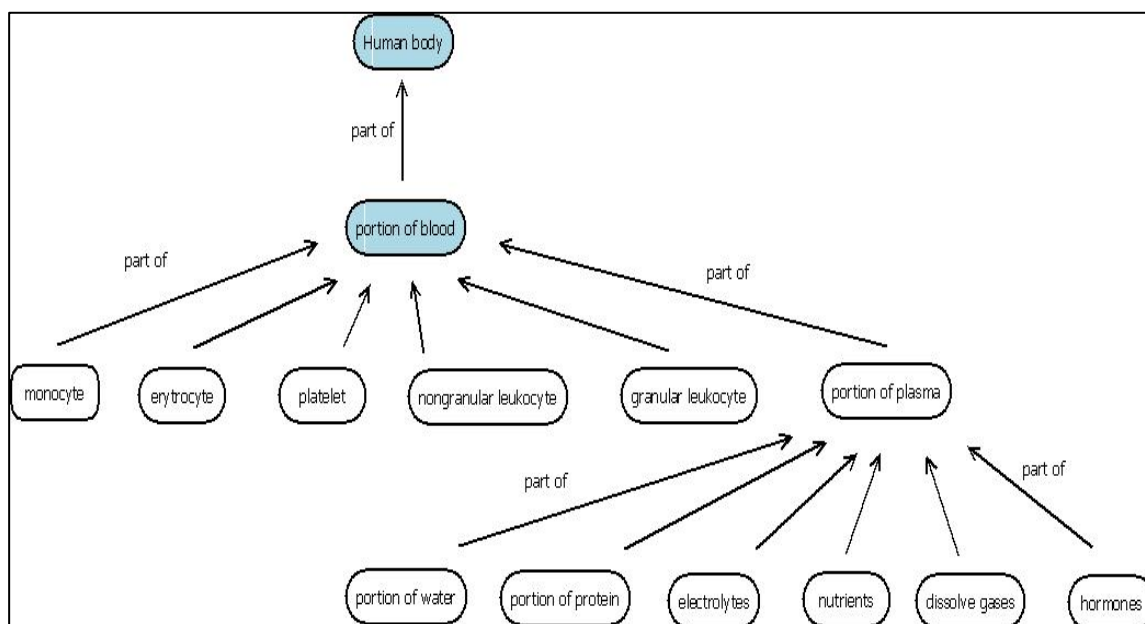
⁵⁷ O site desta associação está disponível em: <http://www.aabb.org>. Acesso em: 18 de Março de 2014.

8.3 Modelos Conceituais

O resultado da etapa de conceitualização do domínio do sangue produziu um conjunto de modelos conceituais aceitos e compartilhados pelos grupos envolvidos nesta etapa a serem utilizados no desenvolvimento do conteúdo ontológico da HEMONTO. Cada modelo conceitual produzido é acompanhado de uma descrição textual, que tem por objetivo explicar a representação gráfica do modelo.

Os modelos conceituais desenvolvidos englobam taxonomias, partonomias e estruturas gerais de representação, as quais incluem mais de um tipo de relação conceitual. Por questões de espaço, apresentam-se apenas alguns exemplos de modelos desenvolvidos. Os demais estão disponíveis na documentação da ontologia no formato OWL-Doc.

Figura 49 - Partonomia de uma porção de sangue.



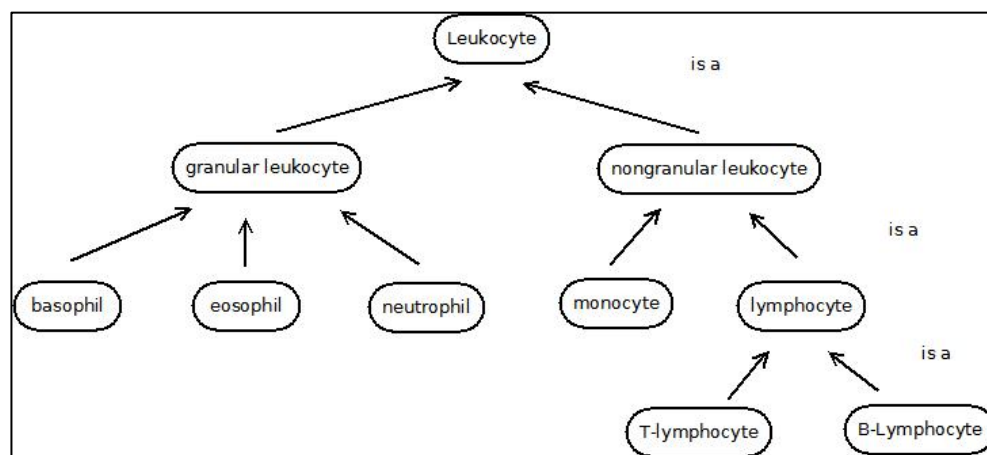
Fonte: elaborado pelo autor.

Na partonomia de uma porção de sangue - Figura 49 -, o sangue (*portion of blood*) é definido como o líquido principal do corpo humano (*human body*) e que representa cerca de 8% do peso corporal. Este volume é composto de uma porção de plasma (*portion of plasm*) e dos chamados elementos formados (*monocytes*, *erythrocytes*, *platelets* e *leukocytes*). Uma porção de plasma é constituída de uma porção de água (*portion of water*), uma porção de proteína (*portion of protein*) e das seguintes soluções: eletrólitos (*electrolytes*), nutrientes (*nutrients*), gases insolúveis (*dissolve gases*) e hormônios (*hormones*). Elementos formados correspondem às plaquetas (*platelets*), conhecidas

também como trombócitos, os leucócitos (*leukocytes*), também chamados de glóbulos brancos e que podem ser granulares (*granular leukocytes*) ou não granulares (*non granular leukocytes*), e os eritrócitos (*erythrocytes*), conhecidos também como glóbulos vermelhos ou hemácias.

Ainda em relação aos elementos constituintes de uma porção de sangue, exemplos de eletrólitos podem ser: K, Ca², Mg², Cl, HCO₃, HPO₄ e SO₄; exemplos de proteínas presentes no plasma são: albumina, globulina, fator H, Icg total, transferrina, fibrinogênio, para citar apenas algumas; exemplos de nutrientes são: glicose, aminoácidos, lipídios, colesterol e vitaminas; exemplos de gases dissolvidos são: dióxido de carbono, oxigênio e nitrogênio; exemplos de produtos não aproveitados são: uréia, ácido úrico, creatinina e bilirrubina.

Figura 50 - Taxonomia dos tipos de leucócitos.

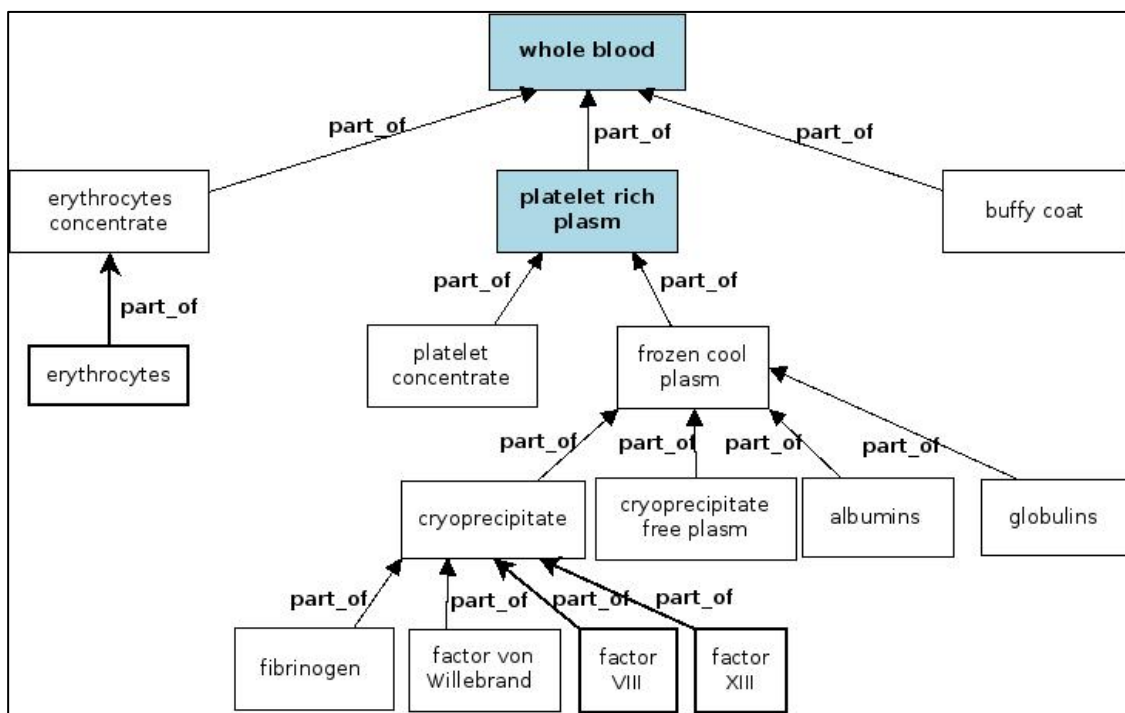


Fonte: elaborado pelo autor.

No que se refere aos leucócitos (*leukocytes*), existem também especificações para seus dois tipos mais gerais: os granulares e os não granulares, e esse conhecimento específico deu origem à taxonomia dos tipos de leucócitos, apresentado na Figura 50. Os leucócitos granulares (*granular leukocytes*) podem ser de três tipos: basófilos (*basophils*), neutrófilos (*neutrophils*) e eosinófilos (*eosophils*); e os leucócitos não granulares (*nongranular leukocytes*) são dois: os monócitos (*monocytes*) e os linfócitos (*lymphocytes*).

A partonomia de uma porção de sangue e a taxonomia dos tipos de leucócitos, apresentadas anteriormente, fazem parte dos aspectos fisiológicos do sangue humano. Um outro tipo de informação representado na ontologia HEMONTO diz respeito aos produtos derivados do sangue que funcionam como hemocomponentes e hemoderivados, os quais estão representados na partonomia, a seguir, da Figura 51.

Figura 51 - Partonomia dos produtos derivados do sangue.



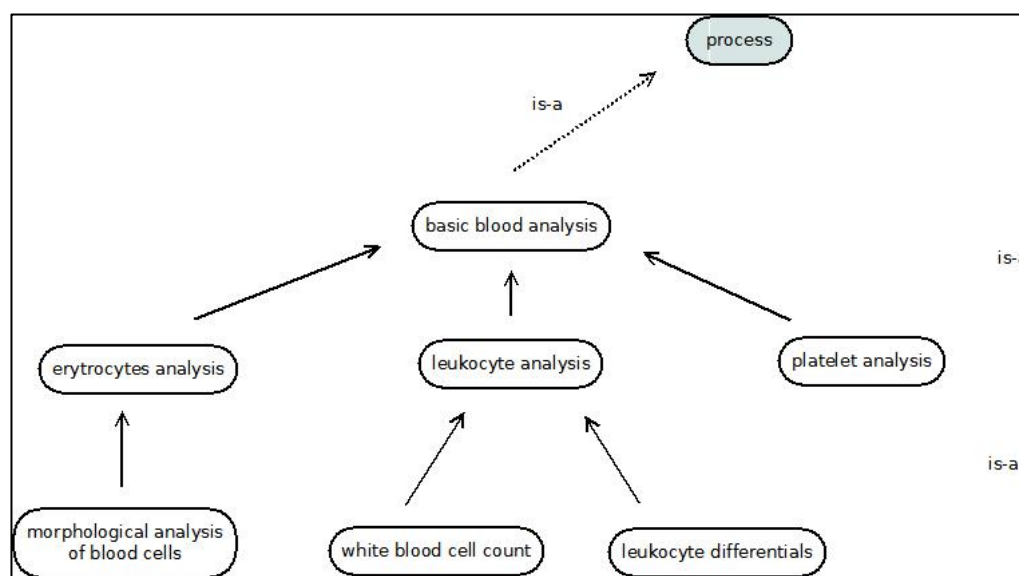
Fonte: elaborado pelo autor.

De acordo com o FMA, o sangue total (*FMA:whole blood*) é a substância e líquido principal do corpo humano, formado de plasma e células sanguíneas. Para a obtenção dos hemocomponentes e hemoderivados do sangue é necessário submeter uma unidade de sangue total a processos específicos, tais como centrifugação e ocongelamento. O sangue total, quando submetido ao primeiro processo de centrifugação, é separado, inicialmente, em três produtos: (i) hemácias (*erythrocytes*), que ficam no fundo da bolsa coletora de sangue e cuja porção quando armazenada em condições específicas de temperatura e armazenamento gera o hemocomponente *erythrocytes concentrate*; (ii) o plasma rico em plaquetas (*platelet rich plasm*), que corresponde ao plasma em sua situação natural (bruta) que fica na camada superior da bolsa; e (iii) o *buffy coat* ou camada leucoplaquetária, porção do sangue formada por leucócitos e plaquetas, que é a camada intermediária entre o plasma e as hemácias. Na sequência, temos que após um novo processo de centrifugação do sangue, realizado em alta rotação, o *plasma rico em plaquetas* é separado em dois produtos: (i) o hemocomponente concentrado de plaquetas (*platelet concentrate*) e (ii) o plasma fresco (*cool plasm*): um plasma com baixa porcentagem de plaquetas. Por sua vez, o plasma fresco pode ser submetido a um processo de extração de um de seus componentes – o crioprecipitado – dando origem a dois outros hemocomponentes: (i) o próprio crioprecipitado (*cryoprecipitate*) e (ii) o crioprecipitado livre de plasma (*cryoprecipitate-free plasm*). Do plasma fresco ainda é

possível extrair dois hemoderivados: albuminas (*albumins*) e globulinas (*globulins*), a partir do fracionamento desse plasma por meio de processos industriais. Por fim, o hemocomponente crioprecipitado contém uma série de glicoproteínas de alto peso molecular (*fibrinogen*, *factor Von Willebrand*, *factor VIII* e *factor XIII*) que cumprem o papel de fatores de coagulação na hemoterapia. Por meio de processos industriais é possível obter tais proteínas e gerar um outro hemoderivado importante conhecido como concentrado de fatores de coagulação (*clotting factors concentrate*), que engloba essas proteínas elucidadas.

Outro tipo de informação sobre o sangue humano representado na ontologia HEMONTO refere-se aos processos de obtenção dos componentes do sangue e também os processos de análise básica do sangue. Para a análise do processo básico de amostras de sangue foi construída uma partonomia, conforme apresentado a seguir na Figura 52.

Figura 52 - Análise do processo básico de amostras de sangue.

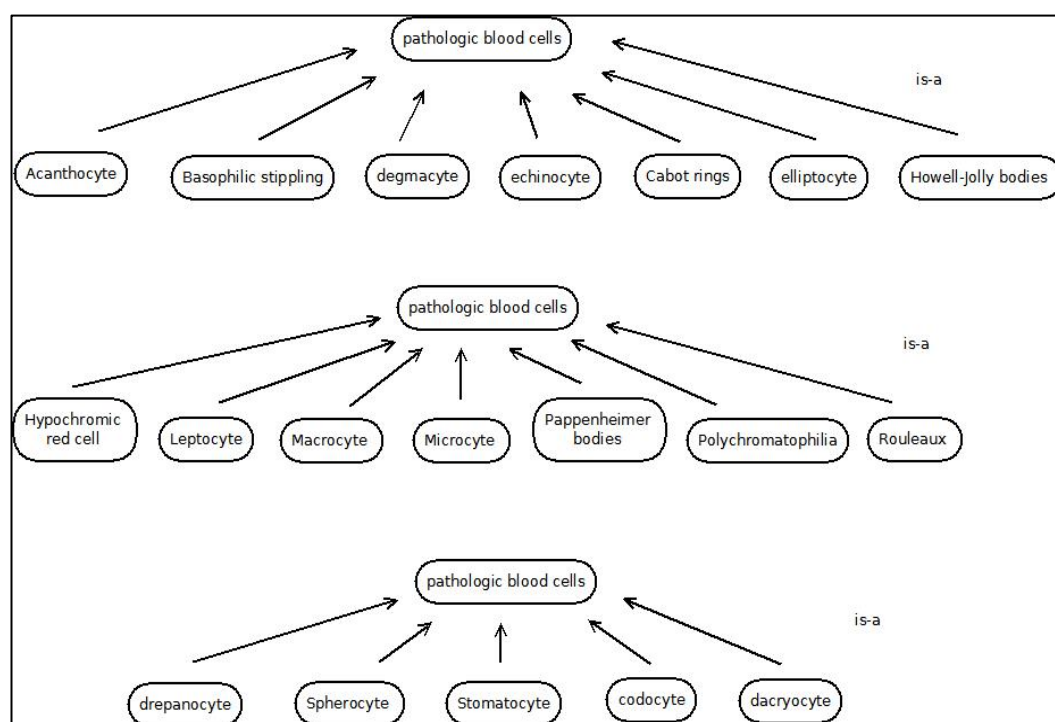


Fonte: elaborado pelo autor.

Na partonomia da análise do processo básico de amostras de sangue foram considerados três tipos de processos de análise: análise de eritrócitos (*erythrocytes analysis*), análise de leucócitos (*leukocytes analysis*) e análise de plaquetas (*platelet analysis*). Abaixo da análise de eritrócitos está a *análise morfológica das células sanguíneas (morphological analysis of blood cells)*, que são processos pelos quais os tipos patológicos das células sanguíneas são identificados. Há também a contagem dos glóbulos brancos (*white blood cell count*) e processos diferenciais de leucócitos (*leukocyte differentials*), abaixo da análise de leucócitos.

Por sua vez, a análise morfológica das células sanguíneas identifica as células patológicas do sangue (*pathologic blood cells*), que foram representadas através de taxonomias, conforme mostra a Figura 53, a seguir. A avaliação dos esfregaços de sangue é um importante estágio da avaliação da doença hematológica, porque muitas doenças podem ter contagem normal de sangue, porém uma morfologia celular anormal.

Figura 53 - Tipos de células sanguíneas patológicas.



Fonte: elaborado pelo autor.

8.4 Ontologias reutilizadas

No desenvolvimento da teoria ontológica do sangue humano (HEMONTO) foram utilizadas as ontologias de fundamentação *Basic Formal Ontology (BFO)* e *Relation Ontology (RO)*, e também as ontologias biomédicas de domínio: *Gene Ontology (GO)*; *Foundational Model Anatomy (FMA)*; *Cell Ontology (CL)*; *Protein Ontology (PRO)* e *Chemical Entities of Biological Interest (ChEBI)*. De maneira mais geral e resumida, pode-se explicar da seguinte forma o emprego das ontologias selecionadas no desenvolvimento da HEMONTO:

- A BFO foi tomada como ponto de partida para a construção da ontologia HEMONTO. Especificamente, a BFO foi utilizada para a definição das classes mais gerais da ontologia construída, ou seja, suas categorias fundamentais.
- As relações da RO e da BFO foram reutilizadas no desenvolvimento da ontologia HEMONTO para relacionar as classes do domínio. Evidentemente, por ser um domínio biomédico específico, foi necessário criar relações específicas, além daquelas da RO e BFO, para relacionar todas as classes da ontologia.
- Classes das ontologias biomédicas de domínio (GO, FMA, CL, PRO e ChEBI) e suas propriedades foram reaproveitadas na especificação de classes específicas do domínio do sangue, incluindo, por exemplo, aspectos fisiológicos do sangue reaproveitados do *Foundational Model Anatomy* (FMA), células e tipos de células da *Cell Ontology* (CL), e proteínas presentes no sangue humano da *PRotein Ontology* (PRO).

8.5 Conteúdo ontológico

O conteúdo ontológico em si da HEMONTO está representado por meio da taxonomia geral da ontologia, das estruturas de representação gráfica entre as classes e relações da ontologia, um dicionário geral de classes e outro dicionário de relações da ontologia. Esses elementos de representação são descritos nos parágrafos seguintes desta seção.

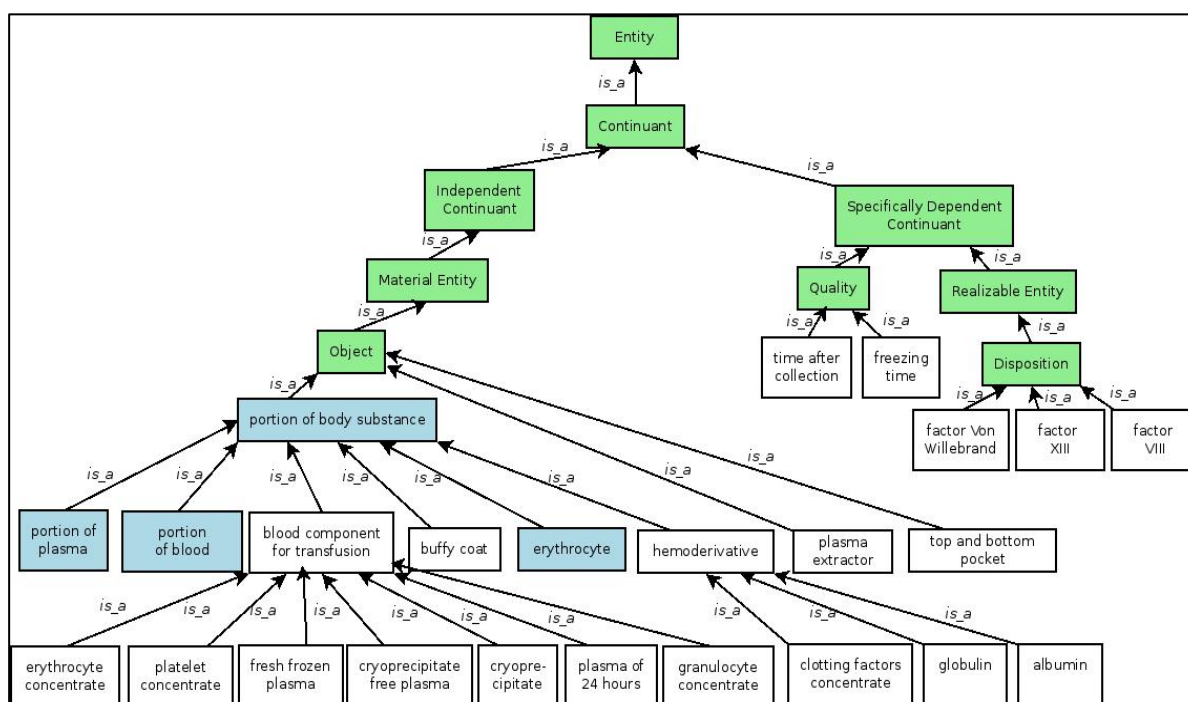
A HEMONTO é uma ontologia biomédica de domínio que, em sua versão atual, engloba um total de 209 termos, dos quais 155 são classes e outros 54 são relações. Dentre as classes incluídas na HEMONTO, 113 são entidades continuantes, 41 são ocorrentes e 1 é apenas entidade. Entre as classes, temos 62 classes específicas da HEMONTO, 34 classes importadas da Basic Formal Ontology (BFO), 22 classes do Foundational Model Anatomy (FMA), 14 da PRotein Ontology (PRO), 9 da Cell Ontology (CL), 4 da Chemical Entities of Biological Interest (ChEBI), 4 da Measurement method ontology (MMO), 3 da Gene Ontology (GO), 1 da Ontology for Biomedical Investigations (OBI), 1 da Phenotypic Trait Ontology (PATO) e 1 da National Drug File Reference Terminology (NDF-RT). Das 54 relações da HEMONTO, as quais incluem também as relações inversas, tem-se 20 relações importadas da Basic Formal Ontology (BFO), 18 da Relation Ontology (RO) e 2 da Gene Ontology (GO), além de 14 novas relações específicas da HEMONTO.

Taxonomia geral da ontologia

A taxonomia geral da ontologia HEMONTO está representada através da hierarquia de classes desenvolvida no Protégé 4.3. Para sua apresentação nesta parte da documentação da ontologia foi feita uma abreviação do número de classes por questões de espaço. As principais classes estão representadas.

Seguindo a estrutura taxonômica da ontologia BFO, que considera dois grandes grupos de entidades: *continuentes* e *ocorrentes*, a taxonomia geral da HEMONTO foi separada em duas taxonomias contendo esses dois tipos de entidades, conforme apresentado nas Figuras 54 e 55, a seguir.

Figura 54 - Parte da taxonomia de entidades continuantes da HEMONTO



Fonte: elaborado pelo autor.

A taxonomia da Figura 54 representa parte do conjunto de *entidades continuantes* da ontologia HEMONTO, na qual as entidades da BFO estão destacadas em verde, aquelas importadas da ontologia FMA estão em azul e as demais entidades são específicas da HEMONTO.

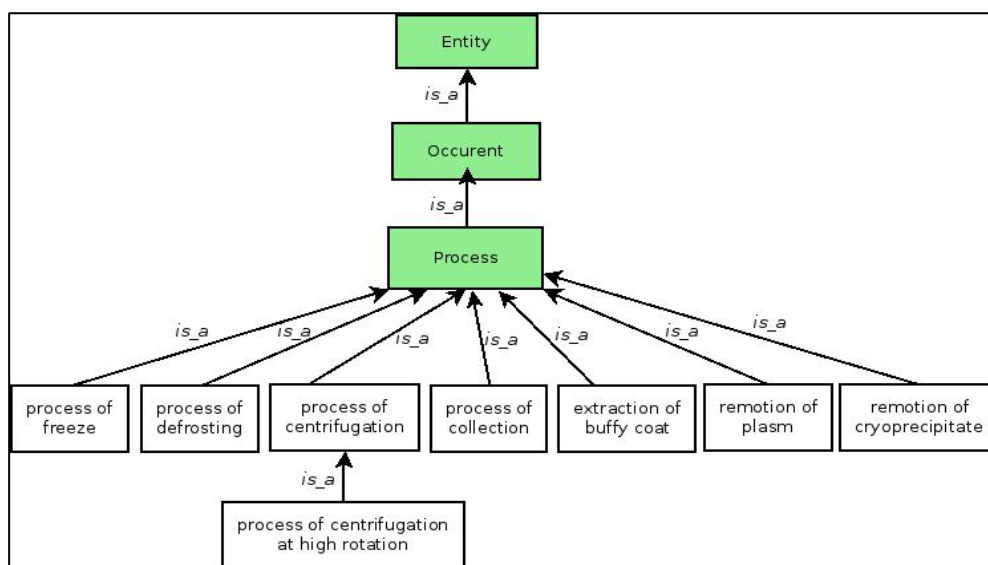
Na taxonomia de entidades continuantes da HEMONTO, porção de substância do corpo (*FMA: portion of body substance*) é uma entidade material e um objeto de acordo com a BFO (*BFO: object*). Uma porção de sangue (*FMA: portion of blood*) é uma subclasse

de porção de substância do corpo e, por consequência, também um *BFO: object*. O mesmo raciocínio é aplicado na inserção de outras subclasses da entidade porção de substância do corpo: hemocomponentes (*HEMONTO: blood components for transfusion*), hemoderivados (*HEMONTO: hemoderivatives*), porção de plasma (*FMA: portion of plasma*), buffy coat (*HEMONTO: buffy coat*) e eritrócitos (*FMA: erythrocytes*). Os tipos de hemocomponentes são também *BFO: object* e subclasses da entidade *HEMONTO: blood components for transfusion*: concentrado de eritrócitos (*HEMONTO: erythrocytes concentrate*), concentrado de plaquetas (*HEMONTO: platelet concentrate*); plasma fresco congelado (*HEMONTO: frozen cool plasm*); crioprecipitado livre de plasma (*HEMONTO: cryoprecipitate free plasm*), crioprecipitado (*HEMONTO: cryoprecipitate*), plasma de 24 horas (*HEMONTO: plasm of 24 hours*) e concentrado de granulócitos (*HEMONTO: granulocytes concentrate*). Os tipos de hemoderivados são também *BFO: object* e subclasses da entidade *HEMONTO: hemoderivatives*: concentrado de fatores de coagulação (*HEMONTO: clotting factors concentrate*), globulinas (*HEMONTO: globulins*) e albuminas (*HEMONTO: albumins*). São também objetos de acordo com a BFO, o extrator de plasma (*HEMONTO: plasm extractor*) e as bolsas top and bottom (*HEMONTO: top and bottom pockets*), mesmo que sejam de natureza diferente das porções de substância do corpo.

Já no ramo direito da taxonomia de entidades continuantes temos entidades especificamente dependentes (*BFO: specifically dependent continuant*), as quais incluem outros importantes termos no domínio do sangue: os fatores de coagulação do sangue (*HEMONTO: factor Von Willebrand*; *HEMONTO: factor VIII*; *HEMONTO: factor XIII*) são entidades realizáveis e disposição (*BFO:disposition*) de acordo com a BFO. Já as entidades tempo após a coleta (*HEMONTO: time after collection*) e tempo de congelamento (*HEMONTO: freezing time*) são qualidades (*BFO:quality*) também de acordo com os princípios da BFO.

A outra taxonomia de classes da HEMONTO engloba o conjunto de *entidades ocorrentes* e está representado na taxonomia da Figura 55, a seguir.

Figura 55 - Extrato da taxonomia de entidades ocorrentes da HEMONTO



Fonte: elaborado pelo autor.

Na taxonomia de *entidades ocorrentes* (Figura 55), mais uma vez as entidades extraídas da BFO são destacadas em verde e as demais entidades são específicas da HEMONTO. Em tal taxonomia foram incluídos todos os processos que estão envolvidos na obtenção dos hemocomponentes e hemoderivados do sangue humano: processo de congelamento (HEMONTO: *process of freeze*); processo de descongelamento (HEMONTO: *process of defrosting*); processo de centrifugação (HEMONTO: *process of centrifugation*) e um tipo específico dele conhecido como *processo de centrifugação em alta rotação* (HEMONTO: *process of centrifugation at high rotation*); processo de coleta (HEMONTO: *process of collection*); extração de buffy coat (HEMONTO: *extraction of buffy coat*); remoção de plasma (HEMONTO: *remotion of plasm*) e remoção de crioprecipitado (HEMONTO: *remotion of cryoprecipitate*). Todas essas entidades foram classificadas como um *BFO:process*, pelo motivo de terem sua existência vinculada a um acontecimento ou ocorrência e por possuírem partes temporais próprias e dependência de uma ou mais entidades materiais, conforme definido em Grenon and Smith (2004).

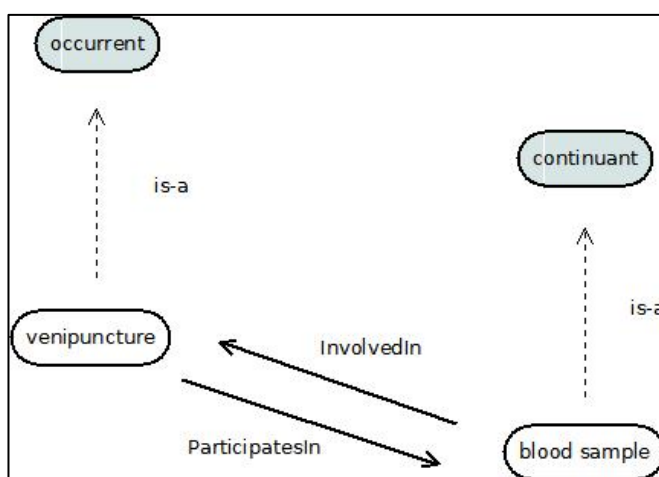
Representação gráfica das classes e relações da ontologia

Estruturas de representação gráfica da ontologia HEMONTO incluem taxonomias específicas, partonomias e estruturas de representação gerais ou mistas, que contém mais de um tipo de relação ontológica. A seguir, apresentam-se as principais de estruturas de representação desta ontologia acompanhadas de sua descrição textual. O enfoque foi dado aos processos de obtenção dos hemocomponentes e hemoderivados do sangue.

O passo inicial para compreensão dos processos de obtenção dos hemocomponentes e hemoderivados do sangue é uma análise dos processos básicos e parâmetros dos elementos sanguíneos com base no entendimento do procedimento de coleta de sangue.

A coleta de sangue, muitas vezes, permite determinar os parâmetros básicos necessários para as funções hematológicas e os diagnósticos. O sangue é coletado através do processo de punção venosa, em tubos de coleta contendo anticoagulante. Na BFO, processos são entidades que representam eventos dos quais os continuantes participam. Punção venosa (HEMONTO: *venipuncture*) é um processo (um ocorrente BFO), especificamente um *planned process* (processo planejado) que envolve uma amostra de sangue (HEMONTO: *blood sample*), classificado como um continuante BFO. Essas informações estão representadas na estrutura da Figura 56, a seguir, onde as entidades extraídas da BFO estão representadas em azul.

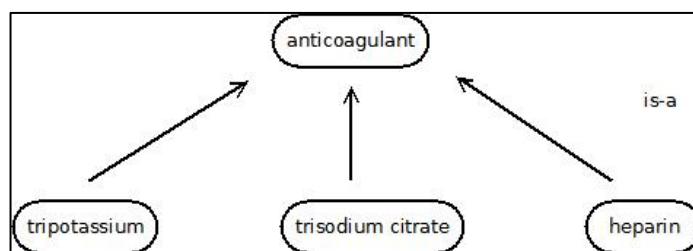
Figura 56 – Representação do processo de punção venosa.



Fonte: elaborado pelo autor.

Além disso, no processo de punção venosa temos três anticoagulantes (HEMONTO: *anticoagulant*) mais comumente usados: tripotássio (HEMONTO: *tripotassium*), citrato trisódico (HEMONTO: *trisodium citrate*) e heparina (HEMONTO: *heparin*), que estão representados na taxonomia da figura 57, a seguir.

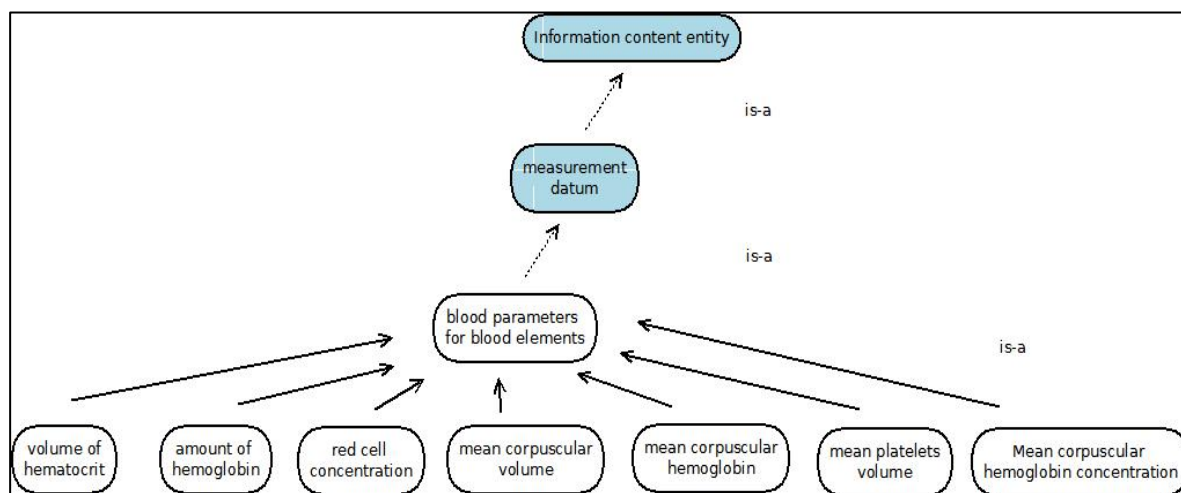
Figura 57 – Taxonomia dos tipos de anticoagulantes.



Fonte: elaborado pelo autor.

Outra parte do conhecimento representada na ontologia *HEMONTO* refere-se aos parâmetros básicos do sangue humano, os quais foram representados a partir das entidades incluídas na ontologia IAO. Os parâmetros básicos para análise do sangue foram classificados como dados mensuráveis (IAO: *measurement datum*) e definidos como entidades de conteúdo informacional (IAO: *information content entities*) (veja Figura 58). Na taxonomia da Figura 58, as entidades extraídas da IAO estão representadas em azul.

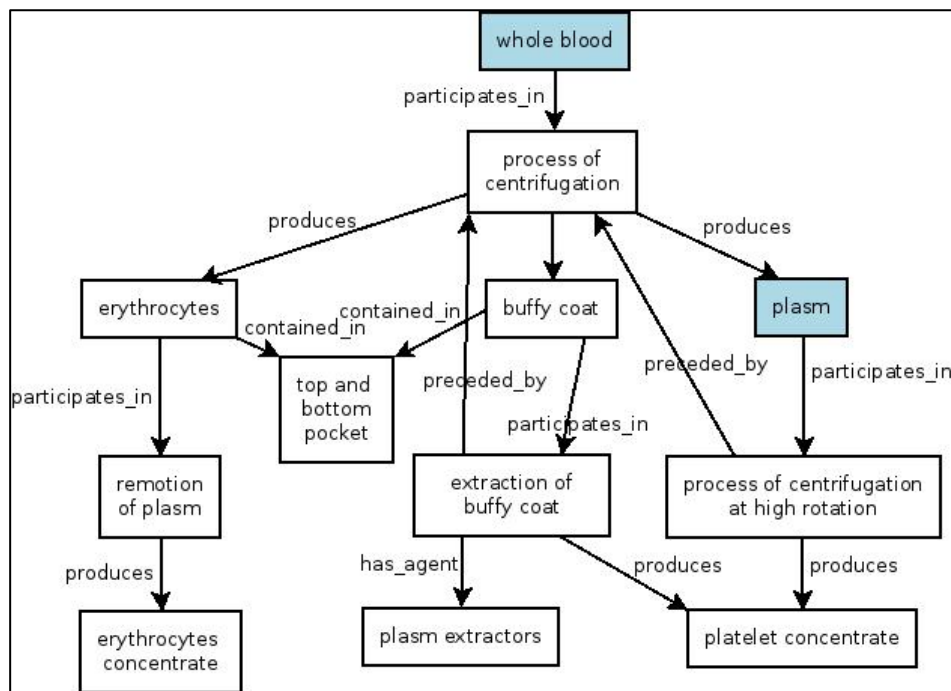
Figura 58 - Parâmetros básicos do sangue como elementos sanguíneos



Fonte: elaborado pelo autor.

Nas próximas estruturas temos a representação e descrição dos processos de obtenção de cada um dos hemocomponentes e hemoderivados do sangue, destacando as especificidades envolvidas em cada um. Os processos de obtenção do concentrado de eritrócitos (HEMONTO: *erythrocytes concentrate*) e do concentrado de plaquetas (HEMONTO: *platelet concentrate*) estão representados em uma mesma estrutura (veja Figura 59), porque possuem etapas em comum no processo de obtenção.

Figura 59 - Processos de obtenção dos hemocomponentes
Concentrado de Hemácias e Concentrado de Plaquetas



Fonte: elaborado pelo autor.

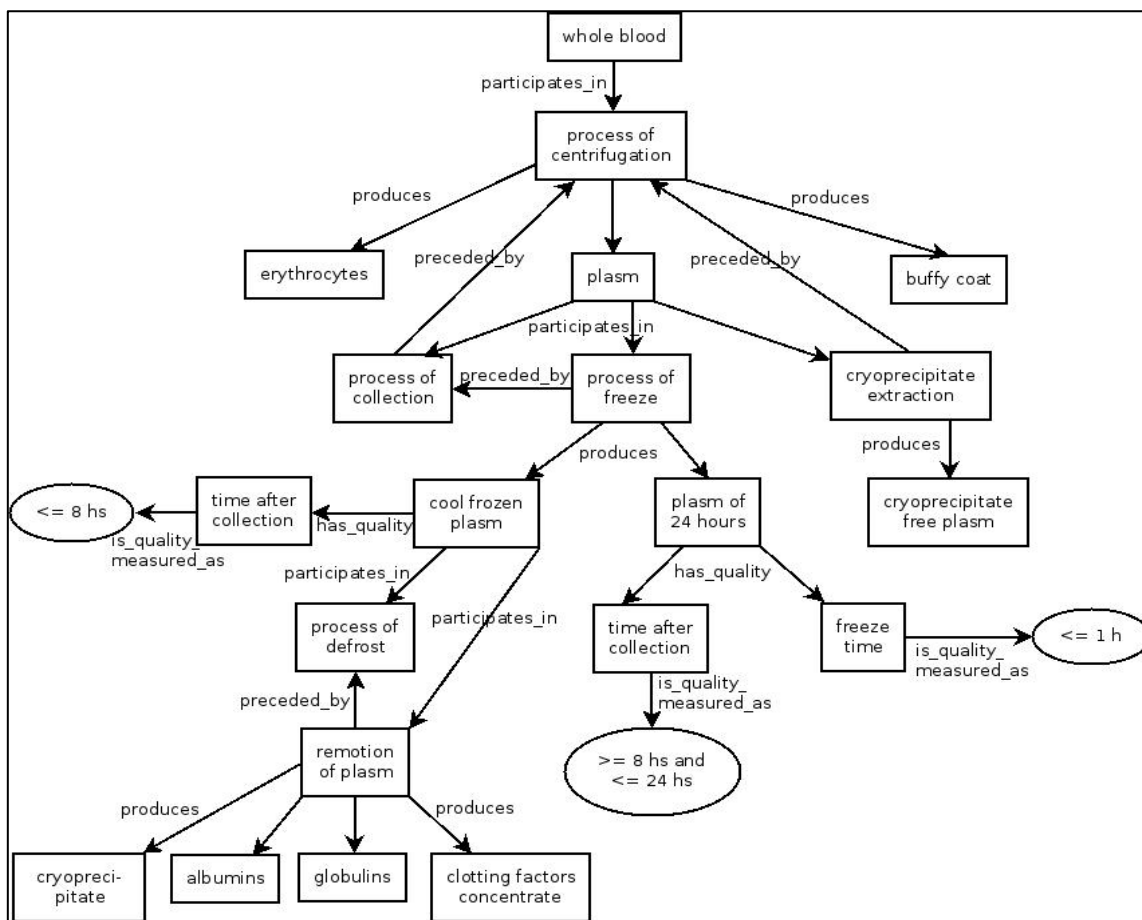
O procedimento inicial para obtenção destes dois hemocomponentes (veja Figura 59) consiste no processo de centrifugação do sangue total (HEMONTO: *process of centrifugation* do *whole blood*), que o separa nos seguintes elementos: plasma (FMA: *portion of plasm*), *buffy coat* (HEMONTO: *buffy coat*) e eritrócitos (FMA: *erythrocytes*). Dessa forma, representa-se que o sangue total participa do processo de centrifugação (*whole blood participates_in process of centrifugation*) e que esse processo produz plasma, *buffy coat* e eritrócitos (*process of centrifugation produces plasm, buffy coat and erythrocytes*). Para a obtenção do hemocomponente concentrado de eritrócitos (lado esquerdo da Figura 59), remove-se do conjunto de eritrócitos contidos em bolsas top and bottom (*erythrocytes contained_in top and bottom pockets*), o plasma que sobrou após o processo de centrifugação do sangue total. Dessa forma, temos que a remoção de plasma através de bolsas top and bottom produz o concentrado de eritrócitos (*remotion of plasm produces erythrocytes concentrate*).

Já para a obtenção do hemocomponente concentrado de plaquetas (HEMONTO: *platelet concentrate*) (lado direito da Figura 59) dois métodos diferentes podem ser usados: (i) obtenção a partir do *buffy coat*; e (ii) obtenção a partir do *plasm*. No primeiro método, o *buffy coat*, produzido após o processo de centrifugação e contido em bolsas top and bottom (*buffy coat contained_in top and bottom pockets*), é extraído por uma das saídas destas bolsas com a utilização dos extratores de plasma (*extraction of*

buffy coat has_agent plasm extractors) e após esse procedimento obtem-se o concentrado de plaquetas (*extraction of buffy coat produces platelet concentrate*). No segundo método, o *plasma* obtido após o primeiro *processo de centrifugação*, conhecida como leve, é novamente centrifugado, agora em alta rotação (*process of centrifugation at high rotation*) e, após esse processo, é produzido o concentrado de plaquetas (*process of centrifugation at high rotation produces platelet concentrate*).

O plasma é um dos mais importantes componentes do sangue humano e, a partir dele, são gerados outros quatro hemocomponentes do sangue: (i) o plasma fresco congelado (PFC) (HEMONTO: *cool frozen plasm*); (ii) o crioprecipitado livre de plasma (HEMONTO: *cryoprecipitate free plasm*); (iii) o plasma de 24 horas (P24) (HEMONTO: *plasm of 24 hour*); e (iv) o próprio crioprecipitado (HEMONTO: *cryoprecipitate*) e também três hemoderivados do sangue: (i) as albuminas (HEMONTO: *albumins*); (ii) as globulinas (HEMONTO: *globulins*) e (iii) o concentrado de fatores de coagulação (HEMONTO: *clotting factors concentrate*). O diagrama da Figura 60, a seguir, representa os processos de obtenção dos hemocomponentes e hemoderivados citados.

Figura 60 - Processos de obtenção dos hemocomponentes e hemoderivados do plasma.



Fonte: elaborado pelo autor.

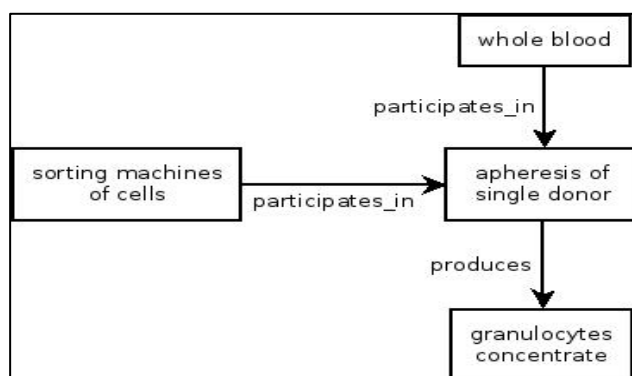
Para obtenção dos hemocomponentes *plasma fresco congelado (PFC)* e *plasma de 24 horas (P24)* (lado esquerdo da Figura 60), mais uma vez o procedimento inicial é o *processo de centrifugação do sangue total* para a separação de *eritrócitos*, *buffy coat* e *plasma*. O passo seguinte consiste no processo de coleta (HEMONTO: *process of collection*) da porção de *plasma* obtida por centrifugação (*portion of plasm participates_in process of collection*). O tempo decorrido após a coleta do plasma (HEMONTO: *time after collection*) é um parâmetro importante no processo como um todo, pois determina qual hemocomponente será gerado: (a) quando esse tempo é de no máximo 8 horas obtém-se o *plasma fresco congelado (cool frozen plasm has_quality time after collection is_quality_measured_as <= 8 hs)* e (b) quando o tempo pós-coleta está entre 8 horas e 24 horas obtém-se o *plasma de 24 horas (plasm of 24 hours has_quality time after collection is_quality_measured_as >= 8 hs and <= 24 hs)*. Para que ambos os hemocomponentes sejam gerados é necessário também que, após o processo de coleta, o plasma seja submetido ao processo de congelamento (*process of freeze preceded_by*

process of collection). No caso do *plasma de 24 horas*, o tempo de congelamento deve ser no máximo de 1 hora (*plasm of 24 hours has_quality freeze time is_quality_measured_as <= 1 h*).

O diagrama da Figura 60 também representa a obtenção dos hemocomponentes *crioprecipitado livre de plasma* e o próprio *crioprecipitado*. Apesar dos nomes de tais hemocomponentes sugerirem similaridades em seus processos de obtenção – remoção do crioprecipitado do plasma e obtenção de ambos –, na prática, esse processo é um pouco diferente. Para a obtenção do *crioprecipitado livre de plasma* (lado direito da Figura 60), a etapa inicial corresponde novamente ao *processo of centrifugação do sangue total* para a obtenção do *plasma, buffy coat e eritrócitos*. A etapa seguinte consiste no processo de extração de crioprecipitado do plasma em sistema fechado (*portion of plasma participates_in cryoprecipitate extraction*). Após essa extração obtém-se o hemocomponente *crioprecipitado livre de plasma* (*cryoprecipitate extraction produces cryoprecipitate free plasm*). Por outro lado, o hemocomponente *crioprecipitado* é obtido a partir do *plasma fresco congelado (PFC)* à temperatura de 1° C a 6° C (lado esquerdo da figura 5). Tal plasma é submetido ao processo de descongelamento (*cryoprecipitate free plasm participates_in process of defrost*) e, em seguida, o plasma sobrenadante é removido, deixando-se na bolsa coletora apenas a proteína precipitada e 10-15 ml deste plasma (*cryoprecipitate free plasm participates_in remotion of plasm*). Esses produtos formam o hemocomponente *crioprecipitado* (*remotion of plasm produces cryoprecipitate*). Nesse processo de *remoção do plasma*, o *plasma* removido também produz três hemoderivados do sangue muito importantes em procedimentos hemoterápicos: albuminas, globulinas e concentrado dos fatores de coagulação (*remotion of plasm produces albumins, globulins and clotting factors concentrate*).

Por fim, temos o processo de obtenção do hemocomponente concentrado de granulócitos (HEMONTO: *granulocytes concentrate*), representado na Figura 61, a seguir.

Figura 61 - Processo de obtenção do hemocomponente Concentrado de Granulócitos



O hemocomponente concentrado de granulócitos é também obtido a partir de uma unidade de sangue total. A diferença aqui é que o sangue total não passa pelo processo de centrifugação, mas sim por um processo de maior complexidade, conhecido como *aférese*, que corresponde à retirada do sangue do doador, seguida da separação de seus componentes por um equipamento próprio, retenção da porção de sangue que se deseja retirar na máquina e devolução dos outros componentes ao doador (BRASIL, 2008). É importante também destacar que a aférese deve ser feita de um único doador (*whole blood participates_in apheresis of single donor*) e que os equipamentos que devem ser usados nesse processo são máquinas de triagem de células (*sorting machines of cells participates_in apheresis of single donor*). Ao final desse processo é produzido o hemocomponente concentrado de granulócitos (*apheresis of single donor produces granulocytes concentrate*).

Dicionário de classes

O dicionário de classes da ontologia HEMONTO foi separado em dois dicionários distintos: um dicionário contendo apenas entidades continuantes e o outro contendo apenas entidades ocorrentes. Cada dicionário inclui as classes da ontologia e todos seus atributos, que correspondem às definições textual e formal de cada classe, suas propriedades descritivas e lógicas, e referências às classes importadas de outras ontologias. Os nomes desses atributos de cada classe da ontologia são apresentados a seguir, na Tabela 38, bem como suas descrições correspondentes no software Protégé 4.3. O dicionário completo de classes continuantes da ontologia HEMONTO é apresentado no Apêndice A deste documento, enquanto o dicionário completo de classes ocorrentes encontra-se no Apêndice B.

Tabela 38 - Correspondência entre os atributos de uma classe e seu nome no Protégé

Nome do atributo	Descrição do atributo no Protégé
ID Classe	has_ID
Classe	Label
Sinônimos	has_synonym
Ontologia de origem	imported_from has_obo_namespace has_URI
Definição textual (em propriedades)	Definition ou Elucidation
Definição formal (em propriedades)	SubClassOf
Exemplo de uso (em propriedades)	example of usage
Anotações/Comentários (em propriedades)	comments

Fonte: elaborado pelo autor.

Dicionário de relações ontológicas

O dicionário de relações da ontologia HEMONTO inclui todas relações ontológicas e os atributos necessários para uma relação: definições textual e semi-formal, propriedades descritivas e lógicas, e referências às relações importadas de outras ontologias. De maneira similar às classes da ontologia HEMONTO, os nomes dos atributos de cada relação são apresentados em uma tabela junto com suas descrições correspondentes no software Protégé 4.3 (veja Tabela 39, a seguir). O dicionário completo de relações da ontologia HEMONTO é apresentado no Apêndice C deste documento.

Tabela 39 - Correspondência entre os atributos de uma relação e seus nomes no Protégé

Nome do atributo	Descrição do atributo no Protégé
ID Relação	has_ID
Relação	Label
Ontologia de origem	imported_from has_obo_namespace has_URI
Relação inversa	inverse_of
Definição textual (em propriedades)	Definition
Sinônimos (em propriedades)	has_synonym
Exemplo de uso (em propriedades)	example of usage
Definição semi-formal (em propriedades)	Definition
Domínio	Domains
Imagem	Ranges
Propriedades lógicas básicas	<i>Characteristics:</i> - <i>Functional</i> - <i>Inverse functional</i> - <i>Transitive</i> - <i>Symmetric</i> - <i>Asymmetric</i> - <i>Reflexive</i> - <i>Irreflexive</i>
Anotações (em propriedades)	Comments

Fonte: elaborado pelo autor.

8.6 Métricas de Avaliação

As métricas de avaliação utilizadas na ontologia HEMONTO correspondem àquelas previstas na metodologia *OntoForInfoScience*, ou seja, um conjunto de critérios avaliativos baseados nos princípios ontológicos da teoria que fundamenta a BFO (MUNN e SMITH, 2008), na caracterização dos tipos de relação e validação de suas propriedades lógicas básicas (MENDONÇA et al., 2015), na identificação de pitfalls (POVEDA-VILLALÓN, 2010) e em regras específicas das metodologias *NeOn*, *Methontology* e *método 101*. O critérios utilizados são agrupados por serem de validação ou verificação

ontológicas e também de acordo com o parâmetro de avaliação, conforme apresentado abaixo, na tabela de avaliação do conteúdo da HEMONTO – Tabela 40.

Tabela 40 - Avaliação da HEMONTO pelos critérios da metodologia *OntoForInfoScience*

Critérios de Validação: adequação ao domínio (mundo real)		
Parâmetro de avaliação	Critério avaliativo	Análise na HEMONTO
Compromisso ontológico	Fundamentação ontológica	Ontologias de fundamentação usadas: BFO e RO. BFO: todas as classes foram importadas na HEMONTO e são utilizadas. Foram também importadas relações ontológicas, que estão sendo utilizadas. RO: todas as relações foram importadas na HEMONTO e são utilizadas.
	Natureza ontológica	Todas classes e relações da HEMONTO foram consideradas genuinamente ontológicas, com uma observação para duas classes da ontologia: “blood component for transfusion” e “hemoderivative”. Tais classes podem ser questionada do ponto de vista de sua criação baseada em métodos de aprendizagem humanos.
	Não intuitividade	As classes e relações da HEMONTO foram consideradas não intuitivas, à exceção das seguintes: - Classes: “blood component for transfusion”, “buffy coat”, “hemoderivative”, - Relações: “structural_part_of”, “funcional_part_of”, “temporary_part_of”.
	Universalidade de classes	Todas classes HEMONTO foram consideradas universais.
	Universalidade de relações	Todas relações HEMONTO foram consideradas universais.
	Formalização das definições das entidades	Algumas classes da HEMONTO não possuem definições formais, devido às dificuldades de se encontrar definições textuais precisas para elas. Por exemplo: <i>protein complex</i> , <i>fibrinogen complex</i> , <i>washing</i> , <i>thawing</i> .
	Definição de propriedades	Além de algumas definições formais não encontradas, foram identificadas também poucas definições de equivalência e disjunção entre as classes da ontologia HEMONTO.
Especificação	Resposta às questões de competência	Com o conteúdo ontológico desenvolvido, conseguiu-se responder as questões de competência propostas. Entretanto, torna-se necessário a elaboração de um maior número de questões de competência.
	Grau de representatividade do domínio	O grau de representatividade da HEMONTO foi considerado significativo para o domínio de componentes do sangue, baseado na boa abrangência dos documentos de referência utilizados e no conteúdo desenvolvido até então. Falta avaliar a utilização prática da ontologia por outros usuários e especialistas do domínio.
Validação especializada	Consulta aos especialistas	Houve o envolvimento de especialistas no processo de validação da HEMONTO.
	Grau de participação	Porém, o grau de participação dos especialistas na validação ontológica foi considerado baixo, se tratando do número de passos necessários para a avaliação da ontologia. A justificativa para o grau baixo foi dada pelo tempo e escopo do desenvolvimento de uma tese.
Ex p a n d i	Capacidade de atualização	Está prevista a atualização da ontologia HEMONTO, porém não foram especificadas as formas de

		atualização, pelo fato da ontologia ainda não estar disponível para todos usuários em um ambiente web.
	Capacidade de integração	A capacidade de integração da HEMONTO é alta, devido ao grande número de classes e relações reutilizadas de ontologias de fundamentação e de domínio, especialmente ontologias integradas a OBO. Foram reutilizadas classes e relações da BFO, RO, FMA, IAO, GO, Cell Ontology, Pro Ontology, CheBI.
Critérios de Verificação: corretude ontológica		
Parâmetro de avaliação	Critério avaliativo	Análise na HEMONTO
Completeness	Única definição textual	Todas classes e relações da HEMONTO possuem uma única definição textual cada uma.
	Única definição formal	Classes e relações da HEMONTO possuem uma única definição formal cada uma e existem algumas que não possuem nenhuma definição formal.
	Definição de propriedades descritivas	A grande maioria das propriedades descritivas das classes e relações da HEMONTO estão definidas, à exceção do não preenchimento dos atributos "example of usage" e ID em algumas classes e relações.
	Caracterização das propriedades básicas da relação	A maioria das relações da HEMONTO tem definidas suas propriedades básicas, com exceção de algumas relações mais específicas do domínio, por exemplo, "coagulate_into".
Integrity	Conectividade com o domínio e contexto	A HEMONTO foi considerada uma ontologia conexa com o contexto e domínio tratado, já que todas classes, relações e propriedades analisadas possuem algum tipo de relação com o restante da ontologia.
	Definição de equivalência	Há poucas definições de equivalência entre as classes da ontologia HEMONTO.
	Definição de disjunção	Há poucas definições de classes disjuntas na ontologia HEMONTO.
Integrity and Completeness	Definição de domínio e imagem da relação	Todas relações da HEMONTO possuem domínio e imagem definidos.
	Definição de relação inversa	Todas relações específicas da HEMONTO e a maioria das relações importadas possuem uma relação inversa correspondente. Aquelas relações que não possuíam uma relação inversa e não foram reaproveitadas na HEMONTO foram excluídas.
	Definição de tipos	Os tipos de dados das classes da HEMONTO foram definidos unicamente para as entidades que representam os hemocomponentes e hemoderivados do sangue.
	Criação de cardinalidade	As cardinalidades das tuplas relação-classe da HEMONTO foram definidas unicamente para as entidades que representam os hemocomponentes e hemoderivados do sangue.
Consistency	Não polissemia	Não foram encontrados termos polissêmicos na HEMONTO.
	Ausência de ciclos na hierarquia	Não foram identificados ciclos na hierarquia principal da HEMONTO.
	Ausência de classes variadas ou genéricas	Não foram encontradas classes que representem dois ou mais conceitos diferentes sob o mesmo rótulo na HEMONTO.
	Não recursividade na definição	Há poucos casos de recursividade na definição dos termos da ontologia HEMONTO, provocados, principalmente, pela importação de definições de outras ontologias existentes. Por exemplo, o termo "blood cell" importado do FMA, cuja definição é: "general anatomical term which refers to nucleated and non-nucleated differentiated hemal cells".

		A recursividade neste exemplo é representada pelo termo “hemal cells”, que é um sinônimo de “blood cell” e, por isso, considerado recursivo.
	Consistência na definição da relação inversa	As relação inversas da ontologia definidas na ontologia HEMONTO foram consideradas consistentes.
Precisão	Definição de sinônimos nas classes	Sinônimos das classes foram apontados corretamente na HEMONTO através dos atributos <code>has_synonym</code> , <code>has_exact_synonym</code> .
	Unicidade dos rótulos (identificadores) das classes	Não foram encontrados rótulos repetidos na HEMONTO.
	Não criação de instâncias como classes	Na ontologia HEMONTO, existem algumas classes que poderiam ser consideradas instâncias, tais como: “whole portion of blood”; “albumin”, “globulin”, “coagulation factor V, X e IX”; “B e T lymphocyte”. Entretanto, como tal ontologia ainda não foi testada em um contexto de aplicação prático com especialistas do domínio, é difícil uma conclusão exata de instâncias e classes desta ontologia.
	Especificação dos diferentes tipos da relação “is_a”	Na ontologia HEMONTO há especificação dos tipos da relação “is_a”, tal que “subClassOf”, é indicado na taxonomia geral de classes construída no Protégé, “instanceOf” usada na definição de instâncias através do Protégé e “sameAs” indicado na definição de sinônimos através dos atributos <code>has_exact_synonym</code> e <code>has_synonym</code> .
	Especificação dos diferentes tipos da relação “part_of”	Na ontologia HEMONTO há especificação dos tipos da relação “part_of” na taxonomia (hierarquia) de relações (Object Properties), incluindo os tipos <code>structural_part_of</code> ; <code>located_in</code> ; <code>immaterial_part_of</code> ; <code>subquantity_of</code> ; <code>constitutes</code> ; <code>temporary_part_of</code> ; <code>member_of</code> ; <code>contained_in</code> .
Documentação	Existência de um documento formal da ontologia	O documento formal que descreve a HEMONTO é descrito em parte nesta tese e através do OWL-DOC.
	Único critério de nomeação dos termos da ontologia	Há um único critério de nomeação dos termos da HEMONTO, que envolve o uso de letras minúsculas, sem acento e separadas por hífen, no caso de palavras compostas.
	Anotação de comentários	Todos os comentários anotados para as classes e relações da HEMONTO estão anotados no atributo “comments” do Protégé.

Fonte: elaborado pelo autor.

9 Conclusões e Trabalhos Futuros

A presente pesquisa apresentou, como principal contribuição científica, uma nova metodologia para construção de ontologias – a *OntoForInfoScience* – direcionada, principalmente, a especialistas em Organização do Conhecimento, provenientes do campo da Ciência da Informação, auxiliando-os a superar problemas relativos ao jargão técnico e a questões lógicas e filosóficas no desenvolvimento de ontologias. Para atingir tal propósito junto aos desenvolvedores de ontologias não tão familiarizados com conceitos específicos da área de engenharia ontológica, a metodologia *OntoForInfoScience* realiza explicações e descrições detalhadas das atividades necessárias do processo de desenvolvimento de uma ontologia, as quais não são bem explicadas nas metodologias atualmente disponíveis para tal fim.

Iniciativas na criação de metodologias próprias para construção de ontologias, reutilizando etapas de metodologias existentes e fazendo adaptações necessárias, tem surgido em pesquisas da área e, em geral, sendo bem aceitas pela comunidade científica que investiga na área de ontologias. Ainda assim, falta em tais iniciativas abordar com maiores detalhes as atividades e processos necessários no ciclo de desenvolvimento ontológico, de forma a auxiliar os desenvolvedores de ontologias nos passos que devem ser executados em relação a questões lógicas e filosóficas e no uso de jargões técnicos.

Esses problemas atuais no processo de desenvolvimento de ontologias impactam, negativamente, na produtividade do número de ontologias desenvolvidas em áreas ligadas à Organização do Conhecimento, que inclui a Ciência da Informação, e também na qualidade das ontologias desenvolvidas. Em função disso, é fundamentalmente importante explicar melhor os conceitos técnicos, lógicos e filosóficos do processo de construção ontológica, de forma a possibilitar uma maior inserção da engenharia ontológica como uma subárea das ciências que lidam com a organização e representação do conhecimento, como é o caso da Ciência da Informação.

Nesse sentido, um dos propósitos com a elaboração da metodologia *OntoForInfoScience* foi exatamente tentar preencher essa lacuna existente entre o processo de construção de uma ontologia e a familiaridade dos profissionais de Organização do Conhecimento, provenientes da Ciência da Informação, com as atividades envolvidas nesse processo. Mas até que ponto é possível adaptar as características particulares do processo de construção de uma ontologia para uma linguagem de fácil entendimento para estes profissionais e pesquisadores?

Encontrar uma resposta exata para essa questão é extremamente complicado, entretanto, algumas conclusões retiradas da presente pesquisa conduzem a uma resposta aproximada dessa questão.

Primeiramente, concluiu-se que a metodologia *OntoForInfoScience* desenvolvida é de grande utilidade para cientistas da informação e também para outros profissionais que atuam na organização do conhecimento, considerando as possibilidades de tal metodologia para a criação de representações ontológicas formais. Além disso, é possível justificar a utilidade da *OntoForInfoScience* para a área destacando as funcionalidades incluídas em tal metodologia, dentre as quais temos: (i) alertar o desenvolvedor sobre a real necessidade ou não do uso de ontologias no problema sob estudo; (ii) explicar informações indispensáveis a uma ontologia, tais como as propriedades descritivas e lógicas das classes e relações; (iii) explicar algum método de aquisição e extração do conhecimento do domínio para uso na ontologia; (iv) tratar a colaboração como um aspecto essencial na etapa de conceitualização, através do uso de ferramentas colaborativas; (v) incentivar o uso de ontologias de fundamentação como ponto de partida do desenvolvimento ontológico; (vi) explicar formas de pesquisa por termos relacionados ao domínio para importação na ontologia desenvolvida; (vii) detalhar o processo de criação de definições textuais e formais das classes de uma ontologia; e (viii) explicar o significado de uma relação ontológica e de que forma é possível caracterizar os tipos de relação ontológica para usá-los adequadamente na representação do domínio.

Em segundo lugar, entende-se que para um maior crescimento do número de ontologias desenvolvidas por profissionais da Organização do Conhecimento e também para uma melhoria da qualidade das ontologias desenvolvidas, faz-se necessário um maior aprofundamento desses profissionais, os quais incluem os cientistas da informação, nos conceitos técnicos, lógicos e filosóficos inerentes ao processo de construção de uma ontologia. Da mesma forma, são necessárias melhorias no conteúdo das atuais metodologias para construção de ontologias, principalmente, no que diz respeito à explicação dos passos e dos termos técnicos utilizados. Acredita-se que apenas com essa contribuição bidirecional das duas áreas, teremos uma maior inserção da engenharia ontológica como uma linha de pesquisa em Organização do Conhecimento e, conseqüentemente, no campo da Ciência da Informação.

A metodologia *OntoForInfoScience* proposta é apenas uma iniciativa nesse sentido, a qual necessita de uma evolução natural e mais aplicações práticas de tal metodologia no desenvolvimento de outras ontologias. Nessa direção, nos próximos parágrafos apresentam-se algumas sugestões de trabalhos futuros para a presente pesquisa.

Inicialmente, é importante destacar que a metodologia *OntoForInfoScience* foi testada unicamente no desenvolvimento da ontologia sobre componentes do sangue humano (HEMONTA) e, ainda assim, tal ontologia não pode ser considerada completamente finalizada, devido ao seu escopo de cobertura. Sabe-se que o processo de construção de uma ontologia é um processo iterativo e sua validação em uma comunidade de especialistas, por exemplo, um corpo médico, é necessária para avaliação de sua utilidade prática. Nesse sentido, a ontologia HEMONTA precisa de validação pós-desenvolvimento entre especialistas da área e também de revisões e atualizações em seu conteúdo ontológico.

Quanto à aplicabilidade prática da metodologia *OntoForInfoScience*, torna-se necessário testá-la em outros domínios do conhecimento, utilizando-a na construção de ontologias dos mais diferentes tipos de domínios. Além disso, é preciso que outros cientistas da informação façam uso de tal metodologia, testando seus passos e relatando suas dificuldades na aplicação da metodologia para a construção de suas ontologias. Essa é uma medida imprescindível para avaliar a eficácia e real aplicabilidade prática da metodologia proposta, uma vez que apenas a partir da visão de “terceiros” e de cientistas da informação, com formação básica na área, é possível se chegar a uma conclusão mais geral de tal metodologia.

Para contribuir com que mais desenvolvedores de ontologia da área de Ciência da Informação e também de outras áreas, que tratam a Organização do Conhecimento, possam utilizar a metodologia *OntoForInfoScience*, sugere-se outros três trabalhos futuros principais: (i) a elaboração de um handbook detalhado sobre a metodologia desenvolvida; (ii) a especificação, a ser feita dentro do próprio handbook, de quais passos e atividades da metodologia são obrigatórios ou opcionais para um desenvolvimento correto do ponto de vista ontológico; (iii) o desenvolvimento de um software que guie o processo de construção de ontologias a partir da metodologia *OntoForInfoScience*.

A elaboração de um handbook para a metodologia *OntoForInfoScience* vai de encontro à definição do termo como um trabalho de referência, que contenha uma coleção de instruções e informações concisas de forma que possa ser facilmente consultado e dar respostas rápidas para o propósito em que ele se aplica, neste caso para a construção de uma ontologia em qualquer domínio do conhecimento. Esse tipo de trabalho tem grande importância para disseminação da metodologia proposta entre todos os desenvolvedores de ontologias, principalmente, aqueles menos familiarizados com o processo de construção, tornando mais simples o acesso as atividades previstas na metodologia e possibilitando uma melhor compreensão de seu conteúdo. A elaboração do handbook

também pode contribuir bastante para o processo de melhoria da *OntoForInfoScience*, gerando feedbacks dos desenvolvedores que utilizam a metodologia e possíveis novas versões da metodologia contemplando melhorias em seu conteúdo.

Outro trabalho futuro sugerido refere-se à elicitación dos passos e atividades da metodologia *OntoForInfoScience* que são obrigatórios e que são opcionais no processo de construção de uma ontologia, considerando sua corretude do ponto de vista ontológico-formal. Dependendo do propósito de desenvolvimento de uma dada ontologia, possivelmente, alguns passos se tornam opcionais nesse processo. Por exemplo, quando o objetivo é o desenvolvimento de uma ontologia leve (*lightweight ontology*) para um domínio específico menos complexo comparado à área biomédica, o uso de uma ontologia de fundamentação no início do projeto ontológico pode se tornar opcional, sendo necessário apenas importar algumas entidades de uma determinada ontologia de fundamentação para desenvolver a ontologia. Neste exemplo, a referência à abordagem filosófica adotada no desenvolvimento da ontologia também não se faz mais necessário, porque as justificativas sobre as decisões ontológicas do desenvolvedor são mais simples. Essa especificação dos elementos obrigatórios e opcionais da metodologia desenvolvida deve estar contida tanto na documentação da *OntoForInfoScience* quanto no seu handbook a ser elaborado.

Por fim, sugere-se como outro trabalho futuro o desenvolvimento de um software que guie o processo de construção de ontologias a partir da metodologia *OntoForInfoScience*. O objetivo principal com este software é formalizar as regras e passos da metodologia proposta em um sistema de informação, de tal forma que o desenvolvedor possa ser orientado sobre as atividades que deve realizar e as restrições que deve seguir no processo de construção de sua ontologia. O software deve direcionar melhor as ações do desenvolvedor na construção da ontologia ao fazer ele seguir as regras e fundamentos básicos da metodologia *OntoForInfoScience*. Isso poderá contribuir para o resultado final da ontologia desenvolvida. Trabalhos científicos similares a essa sugestão já foram desenvolvidos com outras metodologias de construção, tais como o software da metodologia NeOn, o software OntoKEM, a OntoLingua, entre outros projetos deste tipo.

Os trabalhos futuros sugeridos servem para divulgar melhor a metodologia *OntoForInfoScience* e permitir que mais desenvolvedores possam utilizá-la na construção de ontologias dos mais diferentes domínios do conhecimento.

Por fim, como conclusão geral, pode-se afirmar que a investigação aqui conduzida representa uma importante contribuição para as áreas que tratam da Organização e Representação do Conhecimento e, de alguma forma, auxilia em uma maior aproximação entre a Engenharia Ontológica e a Ciência da Informação, cujo futuro

esperado é de que ontologias e Ciência da Informação possam caminhar juntas com o propósito comum de permitir a organização e uso da informação nas mais diversas áreas do conhecimento humano.

Referências

- ABRIAL, J. R. Data semantics. In: KLIMBIE, J.W. e KOFFEMAN, K. L. (Eds.), **Proceedings of the IFIP Working Conference Data Base Management** (pp. 1–60). Amsterdam: North-Holland, 1974.
- ALMEIDA, M. B.; SOUZA, R. R.; COELHO, K. C. Ontologias aplicadas a modelos de sistemas de informação: perspectivas de pesquisa e atuação em Ciência da Informação. In: IV Encontro Ibérico EDIBCIC, 2009, Coimbra. **Anais do IV Encontro Ibérico EDIBCIC**, 2009.
- ALMEIDA, M. B.; TEIXEIRA, L. M. D.; COELHO, K. C.; SOUZA, R. R. Relações semânticas em ontologias: estudo de caso do *Blood Project*. **Liinc em Revista**, v.6, n.2, setembro, 2010, Rio de Janeiro, p. 384- 410.
- ALMEIDA, M. B. Revisiting Ontologies: a necessary clarification. **Journal of the American Society of Information Science and Technology**. v. 64, n. 8. p. 1682–1693, 2013.
- AMERICAN ASSOCIATION of BLOOD BANKS (AABB) 17th edition Technical Manual. Bethesda, Maryland: AABB 2011.
- ANSI/NISO Z39.19:1993. **American National Standard Guidelines for Thesaurus Structure, Construction, and Use**. New York, 1993.
- ANSI/NISO Z39.19:2005. **Guidelines for the construction, format, and management of monolingual controlled vocabularies**. Bethesda: NISO Press, 2005.
- ARISTÓTELES. **Metafísica**. Edição trilingue – grego, latim e espanhol de Valentín García Yebra. Madrid: Gredos, 1998.
- AUGER, A.; BARRIÈRE. Probing semantic relations. **Probing Semantic Relations: Exploration and Identification in Specialized Texts**, vol. 23, p. 1, 2010.
- BACHELARD, G. **O novo espírito científico**, 1934. Disponível em: <<http://pt.scribd.com/doc/65380809/Gaston-Bachelard-O-Novo-Espirito-Cientifico>>. Acesso em: 23 de junho de 2012.
- BARD, J.; RHEE, S. Y.; ASBURNER, M. An ontology for cell types. **Genome Biology** 6:R21, 2005.
- BARRET, K. E.; BARMAN, S.M. ; BOITANO, S.; BROOKS, H. **Ganong's Review of Medical Physiology**, 23rd edition. New York: McGraw-Hill Medical, 2009.
- BEISSWANGER, E.; SCHULZ, S.; STENZHORN, H.; HAHN, U. BIOTOP: An upper domain ontology for the Life Sciences. **Applied Ontology**, p. 1–3, 2007.
- BERNARAS, A.; LARESGOITI, I.; CORERA, J. Building and reusing ontologies for electrical network applications. In: THE EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, ECAI, 1996. **Proceedings...**1996. p. 298-302.
- BERZELL, M. **Electronic Healthcare Ontologies: philosophy, the real world and IT structures**. PhD thesis, Linköping University, Faculty of Arts and Sciences, Division of

Health and Society, Department of Medical and Health Science. Linköping, Sweden, 2010. 163 p.

BHASKAR, R. **A Realist Theory of Science**. Sussex: Harvester Press; 1978.

BIOTOP Ontology, 2012. Disponível em: <<http://www.imbi.uni-freiburg.de/ontology/biotop/>>. Acesso em: 16 de Agosto de 2012.

BIREME – Biblioteca Regional de Medicina, 2012. Disponível em <<http://new.paho.org/bireme/>>. Acesso em: 30 de Abril de 2012.

BITTNER, T.; DONNELLY, M. Logical properties of foundational relations in bio-ontologies. **Artificial Intelligence in Medicine**, v. 39, n. 3, p. 197–216, 2007.

BLACKBURN, S. **The Oxford dictionary of philosophy**. Oxford; New York: Oxford University Press, 1996.

BODENREIDER, O. **The Unified Medical Language System (UMLS): integrating biomedical terminology**. Oxford University. 32(1) Suppl.1: D267- D270, 2004.

BODENREIDER, O. Biomedical Ontologies in Action: Role in Knowledge Management, Data Integration and Decision Support. In: GEISSBUHLER, A.; KULIKOWSKI, C (editors). IMIA Yearbook of Medical Informatics, 2008. **Methods Inf. Med**; 47 Suppl 1:67-79, 2008.

BOOCH, G. **Object-oriented analysis and design with applications** (2nd ed.). Redwood City, CA: Benjamin Cummings, 1993.

BORST, W.N. **Construction of engineering ontologies**. 1997. Tese - University of Twente. Twente, NL, 1997. Disponível em: <<http://www.ub.utwente.nl/webdocs/inf/1/t0000004.pdf>>. Acesso em: 20 de agosto de 2012.

BOSAK, R., RICHARD, F. CLIPPINGER, R.F., DOBBS, C., GOLDFINGER, R., JASPER, R.B., KEATING, W., KENDRICK, G.; e SAMMET, J.E. An information algebra: Phase 1 report – language structure group of the CODASYL development committee. **Communications of the ACM**, 5(4), p. 190–204, 1962.

BRASIL. Ministério da Saúde. **Guia para o uso de hemocomponentes**. Brasília, DF, 2008.

BREITMAN, K. **Web Semântica: a Internet do futuro**. Rio de Janeiro: LTC, 2005.

BUNGE, M. **Teoria e Realidade**. São Paulo: Editora Perspectiva, 1974.

BUSH, V. As we may may think. **Atlantic Monthly**, v.176, n.1, p. 101-108, 1945. Disponível em: <http://www.w3.org/History/1945/vbush/vbush-all.shtml>. Acesso em: 07 de junho de 2012.

CAMPBELL, D. T. Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes, **Psychological Review**, 67(6): 380–400, 1960.

CAMPBELL, K.E.; OLIVER, D.E.; e SHORTLIFFE, E.H. The Unified Medical Language System. Toward a collaborative approach for solving terminologic problems. **J. Am. Med. Inform. Assoc.** 5, p. 12–16, 1998.

CAMPOS, M. L. A. **Linguagem Documentária: Teorias que fundamentam sua elaboração**. Niterói-RJ: EUFF, 2001.

CAMPOS, M. L. A.; BRASIL, M. I.; COELHO, B. A. S.; BASTOS, D. R.. **Vocabulário Sistematizado: a experiência da Fundação Casa de Rui Barbosa**. In: INTEGRAR, 2002, São Paulo. INTEGRAR. São Paulo, 2002.

CAMPOS, M. L. A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 22-32, 2004.

CAPRA, F. **O tao da física**. São Paulo: Cultrix, p. 160, 1983.

CARDOSO, J. The semantic web vision: where are we? **IEEE Intelligent Systems**, p. 22-26, Sept./Oct. 2007.

CARNAP, R. **Foundations of Logic and Mathematics**. In: NEURATH, O.; MORRIS, C.; e CARNAP, R. (eds.), *International Encyclopaedia of Unified Science*. Vol. 1. Chicago: University of Chicago Press, p. 139-213, 1938.

CARTWRIGHT, N. **How the Laws of Physics Lie**. Oxford: Oxford University Press, 1983.

CEUSTERS, W.; SMITH, B.; FLANAGAN, J. Ontology and medical terminology: why description logics are not enough. **Towards an Electronic Patient Record (TEPR 2003)**, Boston, MA: Medical Records Institute, 2003.

CEUSTERS, W., SMITH, B., KUMAR, A. e DHAEN, C. Mistakes in medical ontologies: where do they come from and how can they be detected? **Stud. Health Technol. Inform.** **102**, p. 145–164, 2004.

CEUSTERS, W.; SMITH, B.; GOLDBERG, L. **A terminological and ontological analysis of the NCI thesaurus**. In: *Methods of Information in Medicine 2005*. p. 498–507, 2005.

CEUSTERS, W.; ELKIN, P.; e SMITH, B. Negative Findings in Electronic Health Records and Biomedical Ontologies: A Realist Approach. **International Journal of Medical Informatics** 2007; 76: 326-333.

CHURCHLAND, P. M.; HOOKER, C. A. **Images of Science: Essays on Realism and Empiricism with a Reply from Bas C. Van Fraassen**, 1985. Disponível em: <http://books.google.com.br/books/about/Images_of_Science.html?id=Suc98NDaPMQC&redir_esc=y>. Acesso em: 01 de junho de 2012.

CIA - Central Intelligence Agency, 2012. Disponível em: <<https://www.cia.gov/library/publications/the-world-factbook/>>. Acesso em: 5 de setembro de 2012.

CINTRA, A. M. M.; KOBASHI, N. Y. ; LARA, M. L. G.; TÁLAMO, M. F. M. G. **Para entender as linguagens documentárias**. 2a ed. São Paulo: Editora Pólis, 2002.

CODD, E.F. Extending the database relational model to capture more meaning. **ACM Transactions on Database Systems**, 4(4), p. 397–434, 1979.

CRESWELL, J.W. **Research design: qualitative, quantitative, and mixed methods approaches**. 2. ed. Thousand Oaks: Sage Publications, p. 245, 2003.

CYC Project, 2012. **What's in Cyc?** Disponível em: <http://www.cyc.com/cyc/technology/whatisincyc_dir/whatsincyc>. Acesso em: 20 de Março 2012.

DAHLBERG, I. Classification theory, yesterday and today. **International classification**, v. 3, n. 2, p. 85-90, 1976.

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

DE NICOLA, A.; MISSIKOFF, M.; e NAVIGLI, R. A software engineering approach to ontology building. **Information Systems** 34, p. 258–275, 2009.

DEGEN, W.; HELLER, B.; HERRE, H.; SMITH, B. **GOL: Toward an axiomatized upper-level ontology**. In: Proceedings of the 2nd Int. Conf. on Formal Ontology in Information Systems. New York, USA: ACM, 2001. p.34–46. Ogunquit, USA, 2001.

DEGTYARENKO, K.; MATOS, P.; ENNIS, M.; HASTINGS, J.; ZBINDEN, M.; McNAUGHT, A.; ALCÁNTARA, R.; DARSOW, M.; GUEDJ, M.; e ASHBURNER, M. ChEBI: a database and ontology for chemical entities of biological interest. **Nucleic Acids Research**, Vol. 36, D344–D350, 2008.

DING, Y.; ENGELS, R. IR and AI: Using Co-occurrence Theory to Generate Lightweight Ontologies. **DEXA Workshop**, p. 961-965, 2001.

DODEBEI, Vera Lucia Doyle. **Tesouro: linguagem de representação da memória documentária**. Niterói, Rio de Janeiro: Intertexto; Ed. Interciência, 2002.

ECO, Humberto. **A estrutura ausente**. 7a ed. São Paulo, Perspectiva, 1997.

EDGAR, S. Hermann Cohen. **The Stanford Encyclopedia of Philosophy** (Winter 2011 Edition), Edward N. Zalta (ed.), Disponível em: <<http://plato.stanford.edu/archives/win2011/entries/cohen/>>. Acesso em: 01 de junho de 2011.

EVANS, V, Green, M. 2006. Cognitive Linguistics: an introduction. Edinburgh University Press.

FERNÁNDEZ, M. et al. Building a chemical ontology using methontology and the ontology design environment. **Intelligent Systems**, v. 14, n. 1, p. 37-46, Jan./Feb.1999.

FERNÁNDEZ, M.; CORCHO, O. Methodologies and methods for building ontologies. In: FERNÁNDEZ, M.; CORCHO, O. **Ontological engineering**. London: Springer, 2004. p. 107-153.

FEYERABEND, P. **Against Method: Outline of an Anarchistic Theory of Knowledge**, 1975. Disponível em: <<http://plato.stanford.edu/entries/feyerabend/#2.13>>. Acesso em: 01 de junho de 2012.

FARQUHAR, A.; FIKES, R.; RICE, J. The Ontolingua Server: A Tool for Collaborative Ontology Construction. **Knowledge Systems Laboratory**, KSL-96-26, September 1996.

FAUCONNIER, G.; TURNER, M. Conceptual Integration Networks. **Cognitive Science**, vol. 22 (2), pp. 133-187, 1998.

FERNÁNDEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. METHONTOLOGY: From ontological art towards ontological engineering. AAI Technical Report SS-97-06, 1997.

FMA - Foundational Model of Anatomy, 2012. University of Washington School of Medicine, Spokane, Washington, USA, 2012. Disponível em: <<http://sig.biostr.washington.edu/projects/fm/AboutFM.html>>. Acesso em: 27 de Abril de 2012.

FONSECA, F. The Double Role of Ontologies in Information Science Research. **Journal of the American Society for Information Science and Technology**, 58 (6), pp. 786-793, 2007.

FONSECA, F.; MARTIN, J. Learning the Differences Between Ontologies and Conceptual Schemas Through Ontology-Driven Information Systems. **JAIS - Journal of the Association for Information Systems - Special Issue on Ontologies in the Context of IS**, 8 (2), pp. 129–142, 2007.

FREITAS, F.; SCHULZ, S. Pesquisa de terminologias e ontologias atuais em biologia e medicina. **RECIIS – R. Eletr. de Com. Inf. Inov. Saúde**. Rio de Janeiro, v.3, n.1, p.8-20, mar., 2009.

GANDON, F. **Distributed artificial intelligence and knowledge management: ontologies and multi-agent systems for a corporate semantic web**. Tese - INRIA and University of Nice, Nice, France, 2002.

GENESERETH, M.; FIKES, R. **Knowledge Interchange Format: version 3.0 reference manual report logic-92-1**, 1992. Disponível em: <<ftp://www-ksl.stanford.edu:/pub/knowledge-sharing/papers/kif.ps>>. Acesso em: 21 de julho de 2005.

GERSENOVIC, M. The ICD family of classifications. **Methods Inf. Med.** 1995. Mar; 34(1-2):172-5, 1995.

GERSTL, P.; PRIBBENOW, S. Midwinters, end games, and body parts: a classification of part-whole relations. **International Journal of Human-Computer Studies**, 43, 865-889, 1995.

GIERE, R. **Explaining Science: A Cognitive Approach**. Chicago: University of Chicago Press, 1988.

GIL, A.C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, p. 231, 1991.

GO - Gene Ontology, 2012. Disponível em: <<http://amigo.geneontology.org/cgi-bin/amigo/go.cgi>>. Acesso em: 10 de maio de 2012.

GOLBECK, J.; FRAGOSO, G.; HARTEL, F.; HENDLER, J.; OBERTHALER, J.; PARSIA, B. The National Cancer Institute's thesaurus and ontology. **Journal of Web Semantics**, v. 1, n. 1, p. 75–80, 2003.

GÓMEZ-PÉREZ, A.; FERNÁNDEZ, M.; VICENTE, A. J. **Towards a method to conceptualize domain ontologies**. In: ECAI WORKSHOP ON ONTOLOGICAL ENGINEERING, 1996, Budapest. Disponível em: <<http://citeseer.ist.psu.edu/483876.html>>. Acesso em: 13 de março de 2014.

GÓMEZ-PÉREZ, A. Some Ideas and Examples to Evaluate Ontologies. **Technical Report KSL- 94-65. Knowledge System Lab. Stanford Univ.** Also in Proceedings of CAIA94. 1994.

GÓMEZ-PÉREZ, A.; FERNÁNDEZ-LÓPEZ, M.; CORCHO, O. Ontological Engineering.. Springer Verlag. **Advanced Information and Knowledge Processing series**. ISBN 1-85233-551- 3, November 2003

GÓMEZ-PÉREZ, A. **Ontology Evaluation. Handbook on Ontologies**. S. Staab and R. Studer Editors. Springer. International Handbooks on Information Systems. Pp: 251-274. 2004.

GÓMEZ-PÉREZ, A.; LOZANO-TELO, A. Applying ONTOMETRIC Method to Measure the Suitability of Ontologies. **Business Systems Analysis with Ontologies**. Eds: Green, P.; Rosemann, M. Idea Group Publishing. 2005. PP: 249-269. ISBN: 1-59140-339-1.

GONÇALVES, B. **An Ontological Theory of the electrocardiogram with applications**. 2009. Tese - Universidade Federal do Espírito Santo (UFES), Vitória, ES, Brasil, 2009.

GRENON, P.; SMITH, B.; GOLDBERG, L. **Biodynamic Ontology: Applying BFO in the Biomedical Domain**. In: PISANELLI (ed.), *Ontologies in Medicine*, Amsterdam: IOS Press, p. 20–38, 2004.

GRENON, P.; SMITH, B. SNAP and SPAN: Towards Dynamic Spatial. **Spatial Cognition & Computation**, v.4, n.1, p. 69-104, 2004. Disponível em: <http://ontology.buffalo.edu/smith/articles/SNAP_SPAN.pdf>. Acesso: 12 de Abril de 2012.

GREER, J. P.; FOERSTER, J.; RODGERS, G. M.; PARASKEVAS, F. ; GLADER, B.; ARBER; D. A.; MEANS, R.T. **Wintrobe's Clinical Hematology 12th Edition**. Philadelphia: Lippincott Williams & Wilkins, 2009.

GRÜBER, T. R. **Towards Principles for the Design of Ontologies Used for Knowledge Sharing**. Stanford Knowledge Systems Laboratory, Palo Alto, CA, 1993.

GRUNINGER, M.; FOX, M. S. “**Methodology for the Design and Evaluation of Ontologies**”. In: WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWLEDGE SHARING (IJCAI). [S.l.; s.n.], 1995.

GUARINO, N.; GIARETTA, P. **Ontologies and Knowledge Bases, towards a terminological clarification**. In: MARS, N. (Ed.). *Towards a Very Large Knowledge Bases; Knowledge Building and Knowledge Sharing*. [S.l.]: IOS Press, p. 25-32, 1995.

GUARINO, N. Formal Ontology in Information Systems. **Proceedings of FOIS'98**, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15, 1998.

GUARINO, N.; WELTY, C. Evaluating ontological decisions with ONTOCLEAN. **Communications of the ACM**. February 2002, vol. 45, no. 2.

GUARINO, N.; WELTY, C. An Overview of OntoClean. In: **Handbook on Ontologies. International Handbooks on Information Systems 2004**. Springer Berlin Heidelberg: Springer 2004, pp 151-171.

GUIZZARDI, G.; WAGNER, G.; GUARINO, N.; van SINDEREN, M. **An ontologically well-founded profile for UML conceptual models**. In: Advanced Information Systems Engineering. Berlin / Heidelberg: Springer, 2004. (LNCS, Volume 3084/2004), p. 112–126.

GUIZZARDI, G.; FALBO, R.; GUIZZARDI, R. S. S. Grounding software domain ontologies in the Unified Foundational Ontology (UFO): The case of the ODE Software Process Ontology. In: **Proc. of the Iberoamerican Workshop on Requirements Engineering and Software Environments 2008**. [S.l.: s.n.], p. 127–140. Recife, Brazil, 2008.

GUIZZARDI, G.; WAGNER, G. **Using the Unified Foundational Ontology (UFO) as a foundation for general conceptual modeling languages**. Berlin: Springer-Verlag, 2009.

HAAV, H.M.; LUBI, T.L. **A survey of concept-based information retrieval tools on the web**, 2001. Disponível em: <<http://www.science.mii.lt/ADBIS/local2/haav.pdf>>. Acesso em: 25 de março de 2012.

HEMPEL, C. G., **Aspects of Scientific Explanation and other Essays in the Philosophy of Science**. New York: Free Press, 1965.

HENDERSON, B.; DORSEY, J. **Medical Terminology For Dummies**. Indianapolis: Wiley Publishing, 2009.

HERRE, H. General Formal Ontology (GFO): A Foundational ontology for conceptual modelling. In: **Theory and Applications of Ontology: Computer Applications**. Springer Netherlands, 2010. p. 297-345.

HERRE, H.; HELLER, B.; BUREK, P.; HOEHNDORF, R.; LOEBE, F.; MICHALEK, H. General Formal Ontology (GFO): A Foundational Ontology Integrating Objects and Processes. Part I: Basic Principles. **Leipzig: Institute of Medical Informatics, Statistics and Epidemiology (IMISE)**, p. 16 - 18, 2006.

HERSH, W. R.; DETMER, W. M.; FRISSE, M. E. **Information Retrieval Systems**. In: SHORTLIFFE, E. H.; PERREAULT, L. E.; WIEDERHOLD, G.; FAGAN, L. M. **Medical Informatics: Computer Applications in Health Care and Biomedicine**. 2nd ed. New York: Springer, 2001.

HERSH, W. R. **Information retrieval: a health and biomedical perspective**. 2nd ed. New York : Springer, p.149-60, 2003.

HULL, D. **Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science**. Chicago: The University of Chicago Press, 1988.

HUSSERL, E. **Formal and Transcendental Logic**, 1969 [1929]. CAIRNS, D. (ed.). The Hague: Nijhoff, 1969.

IAO - Information Architect Ontology, 2012. Disponível em: <<http://code.google.com/p/information-artifact-ontology/>>. Acesso em: 16 de setembro de 2012.

ICCBBA – ISBT 128 Standard. **Standard Terminology for Blood, Cellular Therapy and Tissue Product Descriptions**, v 3.33, January 2010. Disponível em: www.iccbba.org. Acesso em: 8 de janeiro de 2014.

IEEE - INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. **Standard 1074 Standard for developing software life cycle processes**, 1995. Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?isnumber=10452&arnumber=490501&count=2&index=0>. Acesso em: 20 de julho de 2012.

IFOMIS-BFO - Institute for Ontology and Medical Information Science - Basic Formal Ontology. Disponível em: <<http://www.ifomis.org/bfo>>. Acesso em: 30 de agosto de 2012.

IHTSDO-SNOMED – International Health Terminology Standards Development Organisation Systematized Nomenclature of Medicine. Disponível em: <<http://www.ihtsdo.org/snomed-ct/>> 2012. Acesso em: 05 de maio de 2012.

JACOBSON, I., CHRISTERSON, M., JONSSON, P., e OVERGAARD, G. **Object oriented software engineering: A use case driven approach**. Boston: Addison-Wesley, 1992.

JARDINE, D. A. The ANSI/SPARC DBMS model. **Proceedings of the Second SHARE Working Conference on Database Management Systems**. Amsterdam: North Holland, 1976.

JONES, D.; BENCH-CAPON, T.; VISSER, P. **Methodologies for ontology development**. 1998. Disponível em: <<http://cweb.inria.fr/Resources/ONTOLOGIES/methodo-for-ontodev.pdf>>. Acesso em: 13 de março de 2014.

KIM, H. M.; FOX, M. S.; GRUNINGER, M. **An ontology for quality management; enabling quality problem identification and tracing**, 1999. Disponível em: <<http://citeseer.ist.psu.edu/context/1728409/0>>. Acesso em: 12 de abril de 2012.

KETT, C.; ARTALE, A. Representing and reasoning over a taxonomy of part-whole relations. **Applied Ontology 3** (1-2), p. 91-110, 2008.

KUHN, T. S. **The Structure of Scientific Revolutions**. Chicago: University of Chicago Press, 3 ed., 1962.

KUMAR, A.; SMITH, B. **The Unified Medical Language System and the Gene Ontology: Some critical reflections**. Berlin / Heidelberg: Springer, vol. 2821/2003, p. 135–148, 2003.

KUMAR, A.; SMITH, B. **Artificial intelligence in medicine**. In: [S.I.]: Springer Berlin Heidelberg, Lecture Notes in Computer Science, Volume 3581/2005. cap. Oncology ontology in the NCI thesaurus, p. 213–220, 2005.

LAKATOS, I. **Criticism and the Growth of Knowledge**. Cambridge: Cambridge University Press, 1970.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos de metodologia científica**. 3ed. São Paulo: Atlas, 1991.

LANCASTER, F.W. **Indexação e resumos: teoria e prática**. Brasília: Briquet de Lemos, 1993.

- LANCASTER, F.W.; WARNER, A. J. Information Retrieval today. **Information Resources Press**, 1993.
- LE MOIGNE, J. **A teoria do sistema geral: teoria da modelização**. Lisboa: Instituto Piaget, 1977.
- LIM, S. C. J.; LIU, Y.; e LEE, W. B. A methodology for building a semantically annotated multi-faceted ontology for product family modelling. **Advanced Engineering Informatics** 25 (2), pp. 147-161, 2011.
- LOWE, E. **A Survey of Metaphysics**. Oxford/New York: Oxford University Press, 2002.
- MASOLO, C.; BORGIO, S.; GANGEMI, A.; GUARINO, N.; OLTRAMARI, A. **Ontology Library: WonderWeb Deliverable D18**, v 1.0, p. 31-12, Trento, Italy, 2003. Disponível em: <<http://www.loa-cnr.it/Papers/D18.pdf>>. Acesso em: 05 de agosto de 2012.
- MASOLO, C.; VIEU, L.; BOTTAZZI, E.; CATENACCI, C.; FERRARIO, R.; GANGEMI, A.; GUARINO, N. Social Roles and their Descriptions. **9th International Conference on Principles of Knowledge Representation and Reasoning (KR2004)**. Banff, Canada, 2004.
- MCGUINNESS, D. L.; e VAN HARMELEN, F. **OWL Web Ontology Language Overview**. W3C Recommendation 10 February 2004. Disponível em: <http://www.w3.org/TR/owl-features/>. Acesso em: 17 de março de 2014.
- MENDONÇA, F. M.; ALMEIDA, M. B. Modelos e teorias para representação: uma teoria ontológica do sangue humano. In: XIII Encontro Nacional de Pesquisa em Ciência da Informação - XIII ENANCIB, 2012, Rio de Janeiro. **Anais do XIII Encontro Nacional de Pesquisa em Ciência da Informação - ENANCIB**, 2012.
- MENDONÇA, F. M.; SOARES, A. L.; ALMEIDA, M. B.; SOUSA, C. P. From a consensual conceptual level to a formal ontological level: a case study in healthcare organizations. **The Ninth International Conference on Advances in Semantic Processing 2015 (SEMABRO 2015)** (em publicação), Nice, France.
- MENDES, R. D. Inteligência artificial; sistemas especialistas no gerenciamento da informação. **Ciência da Informação**, v. 26, n. 1, p. 39-45, 1998. Disponível na Internet: <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=458&layout=abstract>>. Acesso: 10 de agosto de 2012.
- MILLER, G. WordNet: a lexical database for English. **Communications of the ACM**. 1995.
- MILTON, S. **An ontological comparison and evaluation of data modelling frameworks**. Tese, University of Tasmania, Hobart, AU, School of Information Systems, 2000.
- MUNN, K; SMITH, B. **Applied Ontology: An Introduction**. Heusenstamm, Germany: Ontos Verlag, 2008.
- MYLOPOULOS, J. Information modeling in the time of revolution. **Information Systems**, 23(3), p. 127–155, 1998.

MYLOPOULOS, J. Conceptual modeling and telos. In: LOUCOPOULOS P., ZICARI, R. (Eds.), **Conceptual modeling, databases and case: an integrated view of information systems development**. New York: John Wiley e Sons, 1992.

NATALE, D. A.; ARIGHI, C. N.; BARKER, W. C.; BLAKE, J. A.; BULT, C. J.; CAUDY, M.; DRABKIN, H. J.; D'EUSTACHIO, P.; EVSIKOV, A. V.; HUANG, H.; NCHOUTMBOUBE, J.; ROBERTS, N. V.; SMITH, B.; ZHANG, J; WU, C. H. The Protein Ontology: a structured representation of protein forms and complexes. **Nucleic Acids Research** **2011**, vol. 39, Database issue D539–D545, 2011.

NCI Thesaurus – National Center Institute's Thesaurus. Disponível em: <<http://nciterns.nci.nih.gov/NCIBrowser/Dictionary.do>>. Acesso em: 02 de agosto de 2012.

NELSON, S. J.; POWELL, T.; HUMPHREYS, L. B. The Unified Medical Language System (UMLS) of the National Library of Medicine. **Journal of American Medical Record Association**, 61: 40-42, 2006.

NELSON, S.J.; SCHULMAN, J. A. Multilingual Vocabulary Project - Managing the Maintenance Environment. MeSH Section, **National Library of Medicine**, Bethesda, Maryland, 2007.

NEUHAUS, F.; e SMITH, B. Modelling Principles and Methodologies. Relations in anatomical ontologies. In: BURGER, A.; DAVIDSON, D.; e BALDOCK, R (eds.): **Anatomy Ontologies for Bioinformatics: Principles and Practice**, 2007.

NEWTON-SMITH, W. H. **Berkeley's Philosophy of Science**. In: Essays on Berkeley (ed. Foster, J. & Robinson, H.). Oxford: Clarendon Press, 1985.

NIST - National Institute of Standards and Technology. **Announcing the Standard for Countries, Dependencies, Areas of Special Sovereignty and their principal Administrative Divisions**, 1995. Disponível em: <<http://www.itl.nist.gov/fipspubs/fip10-4.htm>>. Acesso em: 20 de julho de 2012.

NLM-UMLS – National Library of Medicine - Unified Medical Language System. Disponível em: <<http://www.nlm.nih.gov/research/umls/>>. Acesso em: 10 de abril de 2012.

NOY, N. F.; McGUINNESS, D. L. **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, March 2001.

OBO - Open Biomedical Ontologies. Disponível em: <<http://www.obofoundry.org>>. Acesso em 04 de Maio de 2012.

OMS – Organização Mundial de Saúde. Disponível em: <<http://new.paho.org/bra/>>. Acesso em: 13 de Setembro de 2012.

PELLIZZON, Rosely de Fátima. Pesquisa na área da saúde: 1. Base de dados: Descritores em Ciências da Saúde (DeCS). **Acta Cir. Bras.**, mar./abr., vol.19, no.2, p.153-163, 2004.

PEREIRA, C.; SOARES, A. Ontology development in collaborative networks as a process of social construction of meaning. **OTM 2008 Workshops, LNCS 5333**, pp. 605–614, Berlin Heidelberg: Springer-Verlag, 2008.

PEREIRA, C.; SOUSA, C.; SOARES, A. A socio-semantic approach to support conceptualisation processes: a case study in an R&D project. **International Journal of Computer Integrated Manufacturing** 2012 (July 2): 1–21. doi:10.1080/0951192X.2012.684714

PESSOA JR., O. **Filosofia e Sociologia da Ciência**, 1993. Disponível em: <<http://www.cfh.ufsc.br/~wfil/sociociencia.htm>>. Acesso em: 29 de maio de 2012.

PETTIGREW, K. E.; MCKECHIE, L. E. F. The use of theory in Information Science research. **Journal of the American Society for Information Science and Technology**, 52(1):62–73, 2001.

POPPER, K. R. **Conjectures and Refutations**, London: Routledge and Keagan Paul, 1963, pp. 33-39. In: Theodore Schick, ed., *Readings in the Philosophy of Science*, Mountain View, CA: Mayfield Publishing Company, 2000, pp. 9-13.

POPPER, K. R. **Três concepções acerca do conhecimento humano**. São Paulo: Abril Cultural, 1980.

POPPER, K. R. **Conjecturas e refutações: o progresso do conhecimento científico**. 5.ed. Brasília: UNB, 2008.

POVEDA-VILLALÓN, M.; SUÁREZ-FIGUEROA, M.C.; GÓMEZ-PÉREZ, A. A Double Classification of Common Pitfalls in Ontologies. **OntoQual 2010 - Workshop on Ontology Quality at the 17th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2010)**. Proceedings of the Workshop on Ontology Quality - OntoQual 2010. ISBN: ISSN 1613-0073. CEUR Workshop Proceedings. Pages: 1-12. 15 October 2010. Lisbon, Portugal.

POVEDA-VILLALÓN, M.; SUÁREZ-FIGUEROA, M.C.; GÓMEZ-PÉREZ, A. Validating ontologies with OOPS!. **18th International Conference on Knowledge Engineering and Knowledge Management**. (EKAW2012) Galway, Ireland, 8 - 12 October 2012.

PUTNAM, H. **Razão, Verdade e História**. Lisboa: Dom Quixote, 1992.

RANGANATHAN, S. R. **Prolegomena to library classification**. Bombay: Asia Publishing House. p. 640, 1967.

RANGANATHAN, S. R. **Facet analysis: fundamental categories**. In: CHAN, L. M.; RICHMOND, P.; SVENONIUS, E. (org.) *Theory of subject analysis: a sourcebook*. Littleton, Co.: Libraries Unlimited, p. 86-93, 1985.

RECTOR, A. L.; ROGERS, J. E.; POLE, P. The GALEN high level ontology. **Medical Informatics in Europe (MIE) 1996**, Copenhagen, Dinamarca, 1996.

RECTOR, A. L. Clinical terminology: why is it so hard? **Methods of Information in Medicine** 38(4):239-252, 2000.

RECTOR, A.; ROGERS, J. E.; ZANSTRA, P. E.; HARING, E. OpenGALEN: Open Source Medical Terminology and Tools. **AMIA Annual Symposium Proceedings 2003**; 982, 2003.

REED, S.L.; LENAT, D.B. **Mapping Ontologies into Cyc**, 2002. Disponível em:

<http://www.cyc.com/doc/white_papers/mapping-ontologies-into-cyc_v31.pdf>. Acesso em: 25 de abril de 2002.

REICHENBACH, H. Experience and prediction: an analysis of the foundations and the structure of knowledge. **Chicago: University of Chicago Press, 1938.**

REYNOLS, P.D. **A primer in theory construction.** New York: Pearson, p. 191, 2007.

RIOS, E. R. G.; FRANCHI, K. M. B.; SILVA, R. M.; AMORIM, R. F.; COSTA, N. C. Senso comum, ciência e filosofia - elo dos saberes necessários à promoção da saúde. **Ciência & Saúde Coletiva**, 12(2):501-509, 2007.

ROBREDO, J. **Documentação de hoje e amanhã.** 2ed. rev. e amp. Brasília: Ed. Autor, p.201-263, 1986.

ROSSE, C.; MEJINO, J. L. V. A reference ontology for biomedical informatics: The Foundational Model of Anatomy. **Journal of Biomedical Informatics**, v. 36, p. 478–500, 2003.

RUMBAUGH, J., BLAHA, M., PREMERLANI, W., EDDY, F., e LORENSEN, W. **Object-oriented modeling and design.** NewYork: Prentice Hall, 1991.

RUTTENBERG, A. Introduction to Ontology. In: **From Basic Formal Ontology to the Information Artifact Ontology - International Conference on Biomedical Ontology (ICBO)**, 2009. Venue: Norton 112, UB North Campus, July 22-23, 2009.

SANTOS, M. T. **Estudo do processo de apropriação da ontologia pela Ciência da Informação no Brasil.** Dissertação (Mestrado), Universidade Federal de Pernambuco, Recife, PE, Brasil, 2014.

SARACEVIC, T. Ciência da Informação: origem, evolução, relações. **Perspectivas em Ciência da informação**, Belo Horizonte, v.1, n.1, p.41-62, 1996.

SATIJA, M. P. Library classification: an essay in terminology. **Knowledge Organization.** Vol. 27, n.4, p. 221-229, 2000.

SAYÃO, L.F. Modelos teóricos em ciência da informação; abstração e método científico. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 82-91, 2001.

SCHIESSL, M.; BRÄSCHER, M. Do texto às ontologias: uma perspectiva para a ciência da informação. **Ciência da Informação**, Brasília, v. 40, n. 2, maio/ago. 2011.

SCHULZ, S.; BEISSWANGER, E.; HAHN, U.; WERMTER, J.; KUMAR, A.; STENZHORN, H. From GENIA to BIOTOP - Towards a top-level ontology for biology. In: BENNETT, B.; FELLBAUM, C. (Ed.). **Proc. of the 4th Int. Conf. of Formal Ontology in Information Systems (FOIS 2006).** Amsterdam: IOS Press, 2006. (Frontiers in Artificial Intelligence and Applications, v. 150), p. 103–114, 2006.

SCHULZ, S.; KUMAR, A.; BITTNER, T. Biomedical ontologies: What part-of is and isn't. **Journal of Biomedical Informatics** 39, p. 350–361, 2006.

SCHULZ, S.; e JANSEN, L. Molecular Interactions: On the Ambiguity of Ordinary Statements in Biomedical Literature. **Applied Ontology**, 2008.

SCHULZ, S.; STENZHORN, H.; BOEKER, M.; e SMITH, B. Strengths and limitations of formal ontologies in the biomedical domain. **RECIIS – Electronic Journal Communication Information & Innovation in Health**. Rio de Janeiro, v.3, n.1, 31-45, Mar., 2009.

SHANNON, C. E. A Mathematical Theory of Communication. **Bell System Technical Journal**, vol. 27, p. 379–423, p. 623–656, 1948. Disponível em: <<http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html>>. Acesso em: 15 de agosto de 2012.

SHANNON, C. E., WEAVER, W. **The Mathematical Theory of Communication**. Springfield: University of Illinois Press, 1949.

SILVA, M. Realismo e Anti-realismo na ciência: aspectos introdutórios de uma discussão sobre a natureza das teorias. **Revista Ciência e Educação**, 5(1), 7–13, 1998.

SILVA, D. L. **Uma proposta metodológica para construção de ontologias: uma perspectiva interdisciplinar entre as Ciências da Informação e da Computação**. Dissertação - Escola de Ciência da Informação, UFMG, Belo Horizonte, MG, 2008.

SILVA, D. L.; SOUZA, R. R.; e ALMEIDA, M. B. Ontologias e vocabulários controlados: comparação de metodologias para construção. **Ciência da Informação**, Brasília, v. 37, n. 3, p. 60-75, set./dez. 2008.

SIMONS, P. **Parts: a Study in Ontology**. Oxford: Clarendon Press, 1987.

SMITH, B.; WELTY, C. Ontology: Towards a new synthesis. In: SMITH B., e WELTY, C. (Eds.). **Proceedings of the International Conference on Formal Ontology in Information Systems** (pp. 3–9). New York: ACM Press, 2001.

SMITH, B.; WILLIAMS, J.; SCHULZE-KREMER, S. The ontology of the Gene Ontology. In: **Proc. of the AMIA Symposium 2003**. [S.l.: s.n.], p. 609–13, 2003.

SMITH, B. **Ontology and Information Systems**, 2004. Disponível em: <<http://www.ontology.buffalo.edu/ontology>> Acesso em: 16 de março de 2012.

SMITH, B.; KUMAR, A.; BITTNER, T. Basic Formal Ontology for Bioinformatics. 2004. Disponível em: <<http://blog.51xuewen.com/upload/Blog/ArticleFile/bb07869e-1899-4187-aca9-b6a50e6910de.pdf>>. Acesso em: 13 de maio de 2012.

SMITH, B.; CEUSTERS, W.; KLAGGES, B.; KÖHLER, J.; KUMAR, A.; LOMAX, J.; MUNGALL, C.; NEUHAUS, F.; RECTOR, A. L.; ROSSE, C. Relations in biomedical ontologies. **Genome Biology** 6, R46, 2005.

SMITH, B.; KUMAR, A.; BITTNER, T. Basic formal ontology for bioinformatics. **Journal of Information Systems**, p. 1-16, 2005.

SMITH, B. Against Fantology. In: M. Reicher & J. Marek (Eds.), **Experience and Analysis** (pp. 153-170). Vienna: HPT&ÖPV, 2005.

SMITH, B. et al. The OBO foundry: Coordinated evolution of ontologies to support biomedical dataintegration. **Nature Biotechnology**, v. 25, n. 11, p. 1251–1255, 2007.

SMITH, B.; CEUSTERS, W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. **Applied Ontology** 5 p. 139–188. IO Press, 2010.

SMITH, B. et al. **Basic Formal Ontology 2.0 – Draft specification and user’s guide**, 2012. Disponível em: <<http://bfo.googlecode.com/svn/trunk/docs/bfo2-reference/BFO2-Reference.docx>>. Acesso em: 18 de agosto de 2012.

SOERGEL, D. **Functions of a Thesaurus / Classification / Ontological Knowledge Base**. College of Library and Information Services, University of Maryland, 1997.

SOUSA, C.; PEREIRA, C.; SOARES, A. Collaborative Elicitation of Conceptual Representations: A Corpus-Based Approach. In: **Advances in Information Systems and Technologies**, vol. 206, Á. Rocha, A. M. Correia, T. Wilson, and K. A. Stroetmann, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 111–124.

SOWA J. F. **Building, sharing and merging ontologies**, 1999. Disponível em: <<http://users.bestweb.net/~sowa/ontology/ontoshar.htm>>. Acesso em: 13 de maio de 2012.

SOWA, J. F. **Ontology**, 2000. Disponível em: <<http://www.jfsowa.com/ontology/>>. Acesso em: 13 de maio de 2012.

SPEAR, A.D. **Ontology for the twenty first century: an introduction with recommendations**. Saarbrücken, Germany, 2006.

SUÁREZ-FIGUEROA, M. C.; AGUADO de CEA, G.; BUIL C.; DELLSCHFT, K.; FERNANDEZ-LOPEZ, M.; GARCIA, A; GOMEZ-PEREZ, A.; HERRERO, G.; MONTIEL-PONSODA, E.; SABOU, M.; VILLAZON-TERRAZAS, B.; YUFEI, Z. **NeOn D5.4.1. NeOn Methodology for Building Contextualized Ontology Networks**. NeOn project. <http://www.neon-project.org>. February 2008.

SUÁREZ-FIGUEROA, M. C.; BLOMQUIST, E.; D’AQUIN, M.; ESPINOZA, M.; GOMEZ-PEREZ, A.; LEWEN, H.; MOZETIC, I.; PALMA, R.; POVEDA, M.; SINI, M.; VILLAZON-TERRAZAS M.; ZABLITH, F.; DZABOR, M. **NeOn D5.4.2. Revision and Extension of the NeOn Methodology for Building Contextualized Ontology Networks**. NeOn project. <http://www.neon-project.org>. February 28, 2009.

SUÁREZ-FIGUEROA, M. C. **NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse**. Tese (Doutorado), Madrid: Facultad de Informatica da Universidad Politécnica de Madrid, 2010.

SUPPE, F. (org.), **The Structure of Scientific Theories**. 2a ed. Urbana: University of Illinois Press, pp. 1-241, 1977.

SUPPES, P. **Representation and Invariance of Scientific Structures**. Stanford: CSLI Publications, 2002.

SURE, Y.; STAAB, S.; e STUBER, R. On-To-Knowledge Methodology (OTKM). 2003. Disponível em: http://www.sfu.ca/~mhatala/iat881/papers/2003_ontohandbook_otkmethodology.pdf. Acesso em: 17 de março de 2014.

SWARTOUT, B. et al. **Toward distributed use of large-scale ontologies**. 1996. Disponível em: <http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html>. Acesso em: 13 de março de 2014.

TUDORACHE T.; VENDETTI, J.; NOY N. Web-Protégé: A lightweight OWL ontology editor for the Web. In: C. Dolbear, A. Ruttenberg, and U. Sattler (Eds.), **Proceedings of the Fifth OWLED Workshop on OWL: Experiences and Directions**, volume 432 of CEUR Workshop Proceedings (CEUR-WS), 2008.

UNESCO. **Guidelines for the establishment and development of monolingual thesauri**. Paris, p. 37, 1973.

USCHOLD, M.; KING, M. **Towards a methodology for building ontologies**. 1995. Disponível em: <<http://citeseer.ist.psu.edu/uschold95toward.html>> Acesso em: 11 mar. 2014.

USCHOLD, M.; GRUNINGER, M. Ontologies: Principles, methods and applications. **Knowledge Engineering Review**, v. 11, n. 2, p. 93–136, 1996.

VAN FRAASSEN, B. C. **The Scientific Image**. Oxford: Oxford University Press, 1980.

VAN DE GRAAFF, K.M.; RHEES, R.D.; PALMER, S. L. **Human Anatomy and Physiology**, 3 rd Edition. New York: McGraw-Hill, 2010.

VARZI, A. Parts, wholes, and part-whole relations: the prospects of mereotopology. **Data and Knowledge Engineering**, v. 20, p. 259-286, 1996.

VICKERY, B. C. Ontologies. **Journal of Information Science**, v. 23, n. 4, p. 277-286, 1997.

WAND, Y., WEBER, R. Reflection: Ontology in Information Systems. **Journal of Database Management**, 15 (2), iii-vi, 2004.

WINSTON, M.; CHAFFIN R., HERRMANN, D. A taxonomy of part-whole relations. **Cognitive Science**, 11(4), 417-444, 1987.

WÜESTER, E. **Introduction to the General Theory of Terminology and Terminological Lexicography**. Viena: Springer, 1979.

YOUNG, J.W., e KENT, H.K. Abstract formulation of data processing problems. **Journal of Industrial Engineering**, 9(6), p. 471–479, 1958.

APÊNDICE A – Dicionário de Classes Continuantes

Tabela 41 - Dicionário de Classes Continuantes da HEMONTO

ID	Classe	Sinônimos	Ontologia de origem	Propriedades
001	portion of body substance	body substance	FMA: portion of body substance (FMAID: 9669)	<p>Definição textual: Material anatomical entity in a gaseous, liquid, semisolid or solid state.</p> <p>Definição formal: is_a 'object aggregate' and ('has quality at all times' some 'physical state')</p> <p>Comentários: with or without admixture of cells and biological macromolecules. A portion of body substance is produced by anatomical structures or derived from inhaled or ingested substances that have been modified by anatomical structures.</p> <p>Exemplos de uso: portion of saliva, portion of semen, portion of cerebrospinal fluid, portion of respiratory air, portion of urine, feces, portion of blood, portion of plasma, portion of lymph.</p>
002	physical state	state of matter	id: FMA:85812	<p>Definição textual: One of the principal conditions in which matter exists. Matter is traditionally divided into three states: solid, liquid, and gas.</p> <p>Comentários: Ice, liquid water, and steam, for example, are three states of matter of the same substance. The electrically neutral condition known as plasma is often considered a fourth state of matter.</p>
003	portion of blood	blood	FMA: portion of blood (FMAID: 9670)	<p>Definição textual: Portion of body substance which has as its parts plasma and blood cells.</p> <p>Definição formal: is_a 'portion of body substance' and ('has continuant part at all times that part exists' some 'blood cell') and ('has continuant part at all times that part exists' exactly 1 'portion of plasma')</p>
004	portion of plasma	plasma; blood plasma	FMA: portion of plasma (FMAID: 62970)	<p>Definição textual: Portion of body substance corresponding to the liquid component of blood in which blood cells are immersed.</p> <p>Definição formal: is_a 'portion of body substance' and ('has quality at all times' only liquid) and (contains some 'blood cell')</p> <p>Comentários: Unless otherwise specified the product has been obtained from Whole Blood and frozen.</p>
005	portion of arterial	arterial blood;	FMA: portion of	Definição textual:

	blood	blood in artery	arterial blood (FMAID: 83066)	Blood that is oxygenated in the lungs, is found in the left chambers of the heart and in the arteries, and is relatively bright red.
006	portion of blood in capillary	—	FMA: portion of blood in capillary (FMAID: 263901)	Definição textual: Portion of blood localized in one of the minute blood vessels between the terminations of the arteries and the beginnings of the veins.
007	portion of venous blood	blood in vein; venous blood	FMA: portion of venous blood (FMAID: 83067)	Definição textual: Blood that has passed through the capillaries of various tissues other than the lungs, is found in the veins, in the right chambers of the heart, and in pulmonary arteries, and is usually dark red as a result of a lower content of oxygen.
008	whole portion of blood	entire portion of blood; maximum portion of blood; portion of blood in whole body	FMA: whole portion of blood (FMAID: 263001)	Definição textual: Portion of blood collected into an anticoagulant and not further processed. Definição formal: is_a 'portion of blood' collected_in some anticoagulant and not (separated) Comentários: A unit of blood collected into an anticoagulant and not further processed unless otherwise specified.
009	anticoagulant BFO function	anticoagulante, anticoagulants	id: CHEBI:50249	Definição textual: An agent that prevents blood clotting. Definição formal: is_a realizable_entity and prevent_of some 'blood coagulation'
010	coagulation factor	—	HEMONTO	Definição textual: Enzymes or glycoproteins that performs the blood coagulation. Definição formal: is_a realizable_entity and agent_in some 'blood coagulation'
011	anatomical structure	biological structure	id: FMA:67135	Definição textual: Material anatomical entity which is generated by coordinated expression of the organism's own genes that guide its morphogenesis. Comentários: Anatomical structure has inherent 3D shape; its parts are connected and spatially related to one another in patterns determined by coordinated gene expression. Exemplos de uso: heart, right ventricle, mitral valve, myocardium, endothelium, lymphocyte, fibroblast, thorax, cardiovascular system, hemoglobin, T cell receptor.
012	human body	—	FMA: human body (FMAID: 20394)	Definição textual: Anatomical structure which is the aggregate material substance of an

				individual member of the Homo sapiens species.
013	cell	normal cell	FMA: cell (FMAID: 68646)	<p>Definição textual: Anatomical structure which has as its boundary the external surface of a maximally connected plasma membrane.</p> <p>Definição formal: is_a 'anatomical structure' and has_boundary exactly 'external surface' 'proper part of continuant at all times' 'plasma membrane' max connected has_boundary exactly 1 'external surface' and 'proper part of continuant at all times' only 'plasma membrane'</p> <p>Exemplos de uso: lymphocyte, fibroblast, erythrocyte, neuron.</p>
014	cell component	—	id: FMA:86454	<p>Definição textual: Cardinal cell part which is demarcated from other cell parts by predominantly bona fide anatomical surfaces.</p> <p>Exemplos de uso: golgi complex, endosome, myofilament.</p>
015	plasma membrane	cell membrane, cellular membrane, cytoplasmic membrane, plasmalemma	id: FMA:63841	<p>Definição textual: Cell component which has as its parts a maximal phospholipids bilayer and two or more types of protein embedded in the bilayer.</p> <p>Exemplos de uso: plasma membrane of hepatocyte, sarcolemma, plasma membrane of erythrocyte.</p>
016	blood cell	hemal cell	FMA: blood cell (FMAID: 62844)	<p>Definição textual: General anatomical term which refers to nucleated and non-nucleated differentiated hemal cells.</p>
017	erythrocyte	anucleate erythrocyte; red blood corpuscle; red blood cell(RBC)	FMA: erythrocyte (FMAID: 62845)	<p>Definição textual: A red-colored blood cell that carries oxygen from the lungs to other parts of the body and are responsible for the red color of vertebrate blood.</p>
018	platelet	thrombocyte; blood platelet	FMA: platelet (FMAID: 62851)	<p>Definição textual: A small colorless disk-shaped cell fragment without a nucleus, found in large numbers in blood and involved in clotting.</p>
019	leukocyte	leucocyte; white blood cell (WBC); white cell	FMA: leukocyte (FMAID: 62852)	<p>Definição textual: a colorless blood corpuscle capable of ameboid movement, whose chief function is to protect the body against microorganisms causing disease and which may be classified in two main groups: <i>granular</i> and <i>nongranular</i>.</p>
020	buffy coat	—	HEMONTA	<p>Definição textual: Blood layer composed of leukocytes and platelets, obtained after centrifugation of portion of whole blood.</p>
021	reticulocyte	polychromatophilic erythrocyte	FMA: reticulocyte	<p>Definição textual: An immature red blood cell containing a</p>

			(FMAID: 66785)	network of basophilic filaments and occurring during the blood regeneration.
022	echinocyte	burr cell	FMA: echinocyte (FMAID: 81099)	Definição textual: An RBC with 30+ crenations, bumps or spurs, which reflect damage to the <i>normal</i> cell membrane by various lytics—eg, saponin, bile salts, ionic detergents, lecithin; slow drying; aged blood; rarely, echinocytes reflect disease—eg, uremia or pyruvate kinase deficiency.
023	granular leukocyte	granular leucocyte granulocyte	FMA: granular leukocyte (FMAID: 62854)	Definição textual: A leukocyte with abundant granules in the cytoplasm.
024	nongranular leukocyte	mononuclear leukocyte; agranular leukocyte	FMA: nongranular leukocyte (FMAID: 62855)	Definição textual: A leukocyte that lacks granules.
025	neutrophil	neutrophil leucocyte; neutrophile; neutrophil leukocyte; heterophile leucocyte; polymorphonuclear leukocyte; polymorphonuclear leucocyte	CELL: neutrophil (CL_ID: 775)	Definição textual: Any of the immature or mature forms of a granular leukocyte that in its mature form has a nucleus with three to five lobes connected by slender threads of chromatin, and cytoplasm containing fine inconspicuous granules and stainable by neutral dyes.
026	eosinophil	eosinophil leucocyte; eosinophil leukocyte;	CELL: eosinophil (CL_ID: 771)	Definição textual: Any of the immature or mature forms of a granular leukocyte with a nucleus that usually has two lobes connected by one or more slender threads of chromatin, and cytoplasm containing coarse, round granules that are uniform in size and which can be stained by the dye eosin. Eosinophils are CD9-positive, CD191-positive, and CD193-positive.
027	basophil	basophil leukocyte	CELL: basophil (CL_ID: 767)	Definição textual: Any of the immature or mature forms of a granular leukocyte that in its mature form has an irregularly shaped, pale-staining nucleus that is partially constricted into two lobes, and with cytoplasm that contains coarse, bluish-black granules of variable size. Basophils contain vasoactive amines such as histamine and serotonin, which are released on appropriate stimulation. A basophil is CD123-positive, CD193-positive, CD203c-positive, and FcεR1a-positive.
028	lymphocyte	—	CELL: lymphocyte (CL_ID: 542)	Definição textual: A lymphocyte is a leukocyte commonly found in the blood and lymph that has the characteristics of a large nucleus, a neutral staining cytoplasm, and prominent heterochromatin.
029	lymphocyte of B lineage	B-lymphocyte	CELL: lymphocyte of B lineage (CL_ID: 945)	Definição textual: A lymphocyte of B lineage is a lymphocyte that expresses CD19 on the cell surface. An additional defining

				characteristic is the commitment to express an immunoglobulin complex.
030	T lymphocyte	T cell, T-cell, T-lymphocyte	CELL: T lymphocyte (CL_ID: 84)	<p>Definição textual: A type of lymphocyte whose defining characteristic is the expression of a T cell receptor complex.</p> <p>Comentários: Types of B lineage lymphocytes include B cells and antibody secreting cells (plasmablasts and plasma cells).</p>
031	plasma cell	plasmacyte	CELL: plasma cell (CL_ID: 786)	<p>Definição textual: A terminally differentiated, post-mitotic, antibody secreting cell of the B cell lineage with the phenotype CD138-positive, surface immunoglobulin-negative, and MHC Class II-negative. Plasma cells are oval or round with extensive rough endoplasmic reticulum, a well-developed Golgi apparatus, and a round nucleus having a characteristic cartwheel heterochromatin pattern and are devoted to producing large amounts of immunoglobulin</p>
032	thymocyte	thymic lymphocyte	CELL: thymocyte (CL_ID: 893)	<p>Definição textual: An immature T cell located in the thymus.</p>
033	monocyte	—	CELL: monocyte (CL_ID: 576)	<p>Definição textual: Myeloid mononuclear recirculating leukocyte that can act as a precursor of tissue macrophages, osteoclasts and some populations of tissue dendritic cells.</p> <p>Comentários: Morphology: Mononuclear cell, diameter, 14 to 20 μm, N/C ratio 2:1-1:1. Nucleus may appear in variety of shapes: round, kidney, lobulated, or convoluted. Fine azurophilic granules present; markers: CD11b (shared with other myeloid cells), human: CD14, mouse: F4/80-mid, GR1-low; location: Blood, but can be recruited into tissues; role or process: immune & tissue remodelling; lineage: hematopoietic, myeloid.</p>
034	lymphoblast	blast cell	FMA: lymphoblast (FMAID: 83030)	<p>Definição textual: An immature lymphocyte that has enlarged in response to antigenic stimulation.</p> <p>Comentários: Unlike other usages of the suffix -blast a lymphoblast is a further differentiation of a lymphocyte, T- or B-, occasioned by an antigenic stimulus. The lymphoblast usually develops by enlargement of a lymphocyte, active re-entry to the S phase of the cell cycle, mitogenesis and production of much m-RNA and ribosomes.</p>
035	cellular component	cellular_component	GO: cellular component (GO_ID: 5575)	<p>Definição textual: The part of a cell or its extracellular environment in which a gene product is</p>

				located. A gene product may be located in one or more parts of a cell and its location may be as specific as a particular macromolecular complex, that is, a stable, persistent association of macromolecules that function together. Comentários: Note that, in addition to forming the root of the cellular component ontology, this term is recommended for use for the annotation of gene products whose cellular component is unknown. Note that when this term is used for annotation, it indicates that no information was available about the cellular component of the gene product annotated as of the date the annotation was made; the evidence code ND, no data, is used to indicate this.
036	blood component for transfusion	—	HEMONTO	Definição textual: Part of a unit of blood that is separated of the human body for treatment, healthcare and therapeutic purposes.
037	blood component	hemocomponent, hemo-component	HEMONTO	Definição textual: Part of a unit of blood which is separated by physical or mechanical means and administered based on specific indications. Definição formal: Part_of whole portion of blood and (separated_by physical process or mechanical process).
038	hemoderivado	blood product	HEMONTO	Definição textual: Blood product obtained in industrial measure, from subdivision of the plasma by physical-chemical processes.
039	erythrocytes concentrate	packed red blood cells; RBCs	HEMONTO	Definição textual: Blood from which most of the plasma has been removed and it is composed by RBCs and reduced plasma (75% of hematocrits); WBCs and platelets.
040	platelets concentrate	—	HEMONTO	Definição textual: A product that contains platelets as the major cellular component.
041	granulocytes concentrate	—	HEMONTO	Definição textual: A product in which the major cellular component is granulocytes; preparation includes a sedimenting agent.
042	plasma of 24 hours	—	HEMONTO	Definição textual: Blood component that is obtained of the one whole portion of blood by centrifugation between 8 and 24 hours after the collection and frozen fully at most 1 hour.
043	cryoprecipitate free plasma	—	HEMONTO	Definição textual: Plasma which was removed, by “closed system”, the cryoprecipitate.
044	fresh frozen	FFP; thawed	HEMONTO	Definição textual:

	plasma	plasma		Plasma that has been frozen by a process and to a temperature that will maintain the activity of labile protein fractions.
045	pooled plasma	—	HEMONTO	Definição textual: A product prepared by combining two or more single units of plasma into one container.
046	pooled fresh frozen plasma	—	HEMONTO	Definição textual: Pooled plasma that has been frozen by a process and to a temperature that will maintain the activity of labile protein fractions.
047	platelet rich plasma	—	HEMONTO	Definição textual: Plasma containing platelets removed from whole blood by a process designed to obtain maximum platelet recovery.
048	cryoprecipitate	cryoprecipitated antihemophilic factor	HEMONTO	Definição textual: A product containing the major portion of Factor VIII and fibrinogen prepared from a unit of Fresh Frozen Plasma.
049	pooled cryoprecipitate	—	HEMONTO	Definição textual: A product prepared by combining two or more single units of Cryoprecipitate into one container.
050	platelet rich buffy coat	—	HEMONTO	Definição textual: Buffy-coat prepared by initial hard centrifugation of whole blood for later recovery of the platelets in a second, gentle centrifugation step.
051	pooled platelet rich buffy coat	—	HEMONTO	Definição textual: A product prepared by combining two or more single units of Platelet-Rich Buffy-Coat into one container.
052	pooled platelet	—	HEMONTO	Definição textual: A product prepared by combining two or more single units of Platelets into one container.
053	granulocyte platelet	—	HEMONTO	Definição textual: A product in which the major cellular components are granulocytes and platelets.
054	pooled granulocyte	—	HEMONTO	Definição textual: A product prepared by combining two or more single units of Cryoprecipitate into one container.
055	serum	—	HEMONTO	Definição textual: The liquid portion of blood following the completion of the clotting process.
056	pooled serum	—	HEMONTO	Definição textual: A product prepared by combining two or more single units of Serum into one container.

057	RBCs adeninesaline added	red blood cells adeninesaline added	HEMONTO	Definição textual: A product obtained of a unit of whole blood that is composed by RBCs and 100 ml of additive solution (60% of hematocrits); WBCs, platelets, little plasma.
058	RBCs leukocytes reduced	red blood cells leukocytes reduced	HEMONTO	Definição textual: A product obtained of a unit of whole blood for filtration, that contain 85% original volume of red blood cells and less than 5×10^6 white blood cells.
059	RBCs washed	red blood cells washed	HEMONTO	Definição textual: A product obtained of a unit of whole blood that is composed by RBCs (approx. 70% of hematocrits); reduced WBCs and no plasma. Its volume is 225 ml.
060	RBCs frozen	red blood cells frozen	HEMONTO	Definição textual: A product obtained of a unit of whole blood that is composed only of RBCs (75% of hematocrits).
061	RBCs deglycerolized	red blood cells deglycerolized	HEMONTO	Definição textual: A product obtained of a unit of whole blood that is composed by less than 5×10^6 white blood cells, no platelets and no plasma.
062	granulocytes pheresis	—	HEMONTO	Definição textual: A product obtained of a unit of whole blood that is composed by Granulocytes ($>1.0 \times 10^{10}$ polymorphonuclear cells/unit); lymphocytes; platelets ($>2.0 \times 10^{11}$ /unit); some RBCs.
063	platelets pheresis	—	HEMONTO	Definição textual: A product obtained of a unit of whole blood that is composed by Platelets ($>3 \times 10^{11}$ /unit); RBCs; WBCs; plasma
064	platelets leukocytes reduced	—	HEMONTO	Definição textual: A product obtained of a unit of whole blood that is composed by Platelets (as above); $<5 \times 10^6$ WBCs/final dose of pooled or pheresis platelets.
065	electrolyte composition	—	PATO: electrolyte composition (PATOID: 27)	Definição textual: Set of electrolytes, which are substances that ionizes when dissolved in suitable ionizing solvents such as water. This includes most soluble salts, acids, and bases.
066	water	Oxidane; H2O	CHEBI: water (CHEBI_ID: 15377)	Definição textual: An inorganic hydroxy compound that has formula H2O.
067	protein	—	PRO: protein (PROID: 001)	Definição textual: An amino acid chain that is produced de novo by ribosome-mediated translation of a genetically-encoded mRNA. Comentários: Proteins descended from a common ancestor can be classified into families and superfamilies composed of products

				of evolutionarily-related genes. The domain architecture of a protein is described by the order of its constituent domains. Proteins with the same domains in the same order are defined as homeomorphic.
068	albumin	ALB, Alb-1, Alb1	PRO: albumin (PROID: 3918)	Definição textual: A protein that is a translation product of the human ALB gene or a 1:1 ortholog thereof.
069	globulin	globulins	NDF-RT: globulins (NDF-RTID: 7615)	Definição textual: Globulin proteins also called globular proteins, or spheroproteins are one of the main protein classes. These have a globelike protein structure, that is more or less soluble in aqueous solutions. The two major sub classes are alpha and beta globulins.
070	nutrient	nutrients	CHEBI: nutrient (CHEBI_ID: 33284)	Definição textual: A nutrient is a chemical element needed by all life forms.
071	hormone	hormones, endocrine	CHEBI: hormone (CHEBI_ID: 24621)	Definição textual: An endogenous compound that is formed in specialized organ or group of cells and carried to another organ or group of cells in the same organism. Comentários: The hormone has a specific regulatory function. The term is now commonly used to include non-endogenous, semi-synthetic and fully synthetic analogues of such compounds.
072	top and bottom pocket	—	HEMONTO	Definição textual: Recipient used in the processes for obtaining of blood components to separate the plasma of other elements of the portion of blood.
073	blood container	—	HEMONTO	Definição textual: Recipient used in the processes of blood collection that hold a blood product prepared from one collection.
074	plasma extractor	—	HEMONTO	Definição textual: Product used to express blood components from collection containers.
075	coagulation factors concentrate	—	HEMONTO	Coagulation factors contained in plasma that operate together calcium to produce insoluble fibrin clot.
076	fibrinogen complex (Factor I)	fibrinogen; fibrinogen alpha chain; fibrinogen beta chain; fibrinogen gamma chain	GO: fibrinogen complex (GOID: 5577)	Definição textual: A highly soluble, elongated protein complex found in blood plasma and involved in clot formation. Comentários: The fibrinogen is converted into fibrin monomer by the action of thrombin. Containing two sets of nonidentical chains (alpha, beta, and gamma) linked together by disulfide bonds.
077	prothrombin (Factor II)	F2, coagulation factor II	Protein Ontology	Definição textual: A protein that is a translation product of

			PR:000007299	the human F2 gene or a 1:1 ortholog thereof.
078	coagulation factor V	activated protein C cofactor, F5, proaccelerin, labile factor	Protein Ontology PR:000007300	Definição textual: A protein that is a translation product of the human F5 gene or a 1:1 ortholog thereof.
079	coagulation factor VII	serum prothrombin conversion accelerator, proconvertin, F7, eptacog alfa, SPCA	Protein Ontology PR:000007301	Definição textual: A protein that is a translation product of the human F7 gene or a 1:1 ortholog thereof.
080	coagulation factor VIII	procoagulant component, antihemophilic factor, F8, AHF	Protein Ontology PR:000007302	Definição textual: A protein that is a translation product of the human F8 gene or a 1:1 ortholog thereof.
081	Von Willebrand factor	factor VIII: vWF	Protein Ontology PR_000017364	Definição textual: A protein that is a translation product of the human VWF gene or a 1:1 ortholog thereof.
082	coagulation factor IX	F9, plasma thromboplastin component, christmas factor	Protein Ontology PR:000007303	Definição textual: A protein that is a translation product of the human F9 gene or a 1:1 ortholog thereof.
083	coagulation factor X	stuart-Prower factor, F10, stuart factor	Protein Ontology PR:000007294	Definição textual: A protein that is a translation product of the human F10 gene or a 1:1 ortholog thereof.
084	coagulation factor XI	PTA, FXI, F11, plasma thromboplastin antecedent	Protein Ontology PR:000007295	Definição textual: A protein that is a translation product of the human F11 gene or a 1:1 ortholog thereof.
085	coagulation factor XII	HAF, F12, hageman factor	Protein Ontology PR:000007296	Definição textual: A protein that is a translation product of the human F12 gene or a 1:1 ortholog thereof.
086	coagulation factor XIII A chain	F13A1, transglutaminase A chain, protein-glutamine gamma-glutamyltransferase A chain, coagulation factor XIIIa	Protein Ontology PR:000007297	Definição textual: A protein that is a translation product of the human F13A1 gene or a 1:1 ortholog thereof.
087	coagulation factor XIII B chain	transglutaminase B chain, protein-glutamine gamma-glutamyltransferase B chain, fibrin-stabilizing factor B subunit, F13B	Protein Ontology PR:000007298	Definição textual: A protein that is a translation product of the human F13B gene or a 1:1 ortholog thereof.
088	antithrombin-III	serpin C1, SERPINC1, ATIII	Protein Ontology PR:000003252	Definição textual: A serpin protein that is a translation product of the human SERPINC1 gene or a 1:1 ortholog thereof.

Fonte: elaborado pelo autor.

APÊNDICE B – Dicionário de Classes Ocorrentes

Tabela 42 - Dicionário de Classes Ocorrentes da HEMONTO

ID	Classe	Sinônimos	Ontologia de origem	Propriedades
089	blood draw method;	blood collection process; blood collection; collection process	Measurement method ontology (MMO_ID: 090)	Definição textual: Method for withdrawing blood from an organism.
090	human blood collection	human blood collection process;	HEMONTO	Definição textual: Process of extraction of blood of the one person, in which the blood is used for transfusions and/or made into biopharmaceutical medications by a process called fractionation (separation of whole-blood components). Comentários: Collection may be of whole-blood or of specific components directly (the latter called apheresis).
091	venipuncture	—	Measurement method ontology (MMO_ID: 090)	Definição textual: Extraction of blood from a vein using a needle. Comentários: In human, the basilic, cephalic and median cubital veins are most commonly used as well as the femoral vein.
092	blood component separation	—	HEMONTO	Definição textual: Separation of blood components through some method or process for use of the component in therapeutic purposes.
093	centrifugation	—	HEMONTO	Definição textual: Process of separating molecules by size or density using centrifugal forces generated by a spinning rotor.
094	centrifugation at high rotation	ultracentrifugation	OBI (OBI_ID: 302886)	Definição textual: Process of separating molecules by centrifugal G-forces approximately several hundred thousand times gravity.
095	blood transfusion	—	HEMONTO	Definição textual: Blood transfusion is generally the process of receiving blood products into one's circulation intravenously.
096	blood coagulation	blood clotting	GO: blood coagulation (GOID: 7596)	Definição textual: The sequential process in which the multiple coagulation factors of the blood interact, ultimately resulting in the formation of an insoluble fibrin clot. Comentários: Blood coagulation may be divided into three stages: stage 1, the formation of intrinsic and extrinsic prothrombin converting principle; stage 2, the formation of thrombin; stage 3, the formation of stable fibrin polymers.

097	removal of plasma	—	HEMONTO	Definição textual: Removal of plasma of some portion of human body fluid, separating its of the whole portion.
098	removal of cryoprecipitate	—	HEMONTO	Definição textual: Removal of cryoprecipitate of some portion of human blood, separating its of the portion of plasma.
099	extraction of buffy coat		HEMONTO	Definição textual: Removal of leukocytes and platelets of some portion of human blood, separating these cells of the whole portion and creating the layer of leukocytes and platelets, named buffy coat.
100	apheresis	—	HEMONTO	Definição textual: A blood collection process in which some part of the donation is returned to the donor.
101	deglycerolize	—	HEMONTO	Definição textual: The removal of glycerol by washing.
102	freeze	—	HEMONTO	Definição textual: The treatment of a cellular product freezing it and keeping in temperature appropriate.
103	thawing	—	HEMONTO	Definição textual: Defrosting of a blood product that is currently in the frozen state.
104	washing	—	HEMONTO	Definição textual: The treatment of a cellular product using a compatible solution to remove most of the plasma proteins.
105	rejuvenate	—	HEMONTO	Definição textual: The treatment of Red Blood Cells by a method known to restore 2,3 DPG and ATP to normal levels or above.
106	liquid storage	—	HEMONTO	Definição textual: A product that has been stored in the liquid state and has not been previously frozen.
107	deglycerolized apheresis	—	HEMONTO	Definição textual: The removal of glycerol by washing from an apheresis product.
108	frozen apheresis	—	HEMONTO	Definição textual: An apheresis product maintained in the frozen state after preparation.
109	liquid apheresis	—	HEMONTO	Definição textual: An apheresis product that has been stored in the liquid state and has not been previously frozen.
110	rejuvenated apheresis	—	HEMONTO	Definição textual: The treatment of apheresis Red Blood Cells by a method known to restore 2,3 DPG and ATP to normal levels or above.
111	washed apheresis	—	HEMONTO	Definição textual:

				The treatment of an apheresis cellular product using a compatible solution to remove most of the plasma proteins.
112	leukocyte reduction	Leukocyte depletion; leukodepletion	HEMONTO	Definição textual: Any technique which WBCs in transfused blood products, for example, RBCs, platelets; virtually eliminates nonhemolytic (immunologic) transfusion reactions.
113	basic blood analysis	—	Measurement method ontology (MMO_ID: 030)	Definição textual: Determination of the component parts or elements of blood and their quantity or quality.
114	hematocrit analysis	—	Measurement method ontology (MMO_ID: 214)	Definição textual: A method for measuring the percentage of whole blood volume occupied by red blood cells.
115	erythrocytes analysis	—	HEMONTO	Definição textual: Morphological analysis of red blood cells for identification of pathologic erythrocytes.
116	leukocytes analysis	—	HEMONTO	Definição textual: Morphological analysis of white blood cells for identification of pathologic leukocytes.
117	platelets analysis	—	HEMONTO	Definição textual: Morphological analysis of blood cells for identification of pathologic platelets.
118	morphological analysis of blood cells	—	HEMONTO	Definição textual: Morphological analysis of blood cells for identification of pathologic blood cells.
119	platelets count	—	HEMONTO	Definição textual: Counting of the number of platelets in some portion of human body fluid.
120	red blood cells count	RBCs count	HEMONTO	Definição textual: Counting of the number of red blood cells in some portion of human body fluid.
121	white blood cells count	WBCs count	HEMONTO	Definição textual: Counting of the number of white blood cells in some portion of human body fluid.

Fonte: elaborado pelo autor.

APÊNDICE C – Dicionário de Relações Ontológicas

Tabela 43 - Dicionário de Relações da ontologia HEMONTO

ID Relação	Relação	Relação inversa	Ontologia de origem	Propriedades
01	adjacent_to	-	RO (RO_ID: 2220)	<p>Definição: x adjacent to y if and only if x and y share a boundary</p> <p>Domínio: independent_continuant</p> <p>Imagem: independent_continuant</p> <p>Propriedades básicas: Assimétrica, não transitiva, reflexiva</p>
02	agent_in (actively_participates_in)	has_agent	RO (RO_ID: 2217)	<p>Definição: x actively participates in y if and only if x participates in y and x realizes some active role</p> <p>Domínio: realizable_entity</p> <p>Imagem: process</p>
03	collected_into	has_collected	HEMONTTO	<p>Definição: C collected_into P Relação ontológica identificada entre um continuante C e um ocorrente P, que é uma entidade realizável, tal que o continuante C é extraído ou coletado através de algum meio ou processo, que corresponde à entidade realizável P.</p> <p>Domínio: material_entity</p> <p>Imagem: realizable_entity</p> <p>Exemplo de uso: whole portion of blood is_a 'portion of blood' collected_in some anticoagulant and not (separated)</p>
04	constitutes	constitute_of	HEMONTTO	<p>Definição textual: Tipo de relação part-of verificada entre dois continuantes na qual a parte é uma quantidade de matéria que constitui o todo, o qual é um objeto físico.</p> <p>Definição formal: $\forall x,y(\text{constitutes}_{it}(x,y) \equiv \text{constituted_of}_{it}(y,x) \equiv \text{mpart_of}(x,y) \wedge \text{POB}(y) \wedge M(x))$, onde M é an amount of matter e POB representa physical objects.</p> <p>Domínio:</p>

				<p>material_entity</p> <p>Imagem: material_entity</p> <p>Exemplo de uso: nutrients constitutes portion of blood</p> <p>water constitutes portion of blood</p>
05	contained_in	-	RO (RO_ID: 1018)	<p>Definição textual: Tipo de relação part-of verificada entre dois continuantes (universal-universal ou particular-particular) na qual a parte ocupa uma região espacial 2D que está inserida (dentro) na região espacial 3D ocupada pelo todo.</p> <p>Definição formal: $\forall x, y(\text{contained_in}(x, y) \equiv \text{part_of}(x, y) \wedge R(x) \wedge R(y) \wedge \exists z, w(\text{has_3D}(z, x) \wedge \text{has_3D}(w, y) \wedge \text{ED}(z) \wedge \text{ED}(w)))$, onde ED é um endurente ou continuante e R é a região especial que ele ocupa. 2D e 3D são usadas para diferenciar o tipo de região especial ocupada.</p> <p>Domínio: material_entity</p> <p>Imagem: independent_continuant</p> <p>Exemplo de uso: blood in coronary artery contained_in heart</p> <p>blood cells contained_in portion of plasma</p>
06	derives_from	derived_into	RO (RO_ID: 1000)	<p>Definição: A relation between two distinct material entities, the new entity and the old entity, in which the new entity begins to exist when the old entity ceases to exist, and the new entity inherits the significant portion of the matter of the old entity</p> <p>Domínio: material_entity</p> <p>Imagem: material_entity</p> <p>Exemplo de uso: this cell derives from this parent cell (cell division); this nucleus derives from this parent nucleus (nuclear division).</p> <p>Comentários: This is a very general relation. More specific relations are preferred when applicable, such as 'directly</p>

				develops from'
07	has_boundary	boundary_of	RO (RO_ID: 2002)	<p>Definição: The relationship between a material entity and the boundary (2D immaterial entity) that delimits it.</p> <p>Termo alternativo: has_2D_boundary</p> <p>Domínio: material_entity</p> <p>Imagem: immaterial_entity</p> <p>Exemplo de uso: my body has 2D boundary the surface of my skin</p>
08	immaterial_part_of	has_immaterial_part	HEMONTO	<p>Definição: Tipo de relação part-of verificada entre dois continuantes, tal que a parte é um objeto imaterial e está conectada ao todo, o qual representa a entidade material.</p> <p>Domínio: immaterial_entity</p> <p>Imagem: material_entity</p> <p>Exemplo de uso: lumen of coronary artery immaterial_part_of heart cavity of ventricle immaterial_part_of heart</p>
09	instance_of	has_instance	RO	<p>Definição: A relation between an instance and a class. For components: a primitive relation between a component instance and a class which it instantiates at a specific time. For processes: a primitive relation, between a process instance and a class which it instantiates, holding independently of time" [PMID:15892874]</p> <p>Domínio: entity</p> <p>Imagem: entity</p> <p>Exemplo de uso: fresh frozen plasma instance_of blood component albumin instance_of protein</p> <p>Comentários: The instance_of relationship is considered axiomatic by the obo file format specification; ie it is taken for</p>

				granted. The is_a relation is still included in this ontology for completeness.
10	involved_in	involves	RO (RO_ID: 2431)	<p>Definição textual: c involved in or regulates p if and only if either (i) c is involved in p or (ii) c is involved in regulation of p. Tipo de relação part-of verificada entre dois ocorrentes (universal-universal ou particular-particular) na qual a parte representa uma etapa do todo (processo inteiro).</p> <p>Definição formal: $\forall x, y(\text{involved_in}(x, y) \equiv \text{part_of}(x, y) \wedge \text{PD}(x) \wedge \text{PD}(y))$ onde PD é um perdurante ou ocorrente.</p> <p>Termo alternativo: involved in regulation of</p> <p>Domínio: occurrent</p> <p>Imagem: occurrent</p> <p>Exemplo de uso: centrifugation involved_in process for obtaining erythrocytes concentrate</p>
11	located_in	has_located	BFO	<p>Definição textual: Tipo de relação part-of verificada entre dois continuantes (universal-universal ou particular-particular) na qual a parte ocupa uma porção (ou limite menor) do espaço ocupado pelo todo (porção inteira ou todo limite). b located_in c at t = Def. b and c are independent continuants, and the region at which b is located at t is a (proper or improper) continuant_part_of the region at which c is located at t.</p> <p>Definição formal: $\forall x, y(\text{located_in}(x, y) \equiv \text{part_of}(x, y) \wedge R(x) \wedge R(y) \wedge \exists z, w(\text{has_2D}(z, x) \wedge \text{has_2D}(w, y) \wedge \text{ED}(z) \wedge \text{ED}(w)))$, onde ED é um endurente ou continuante e R é a região espacial que ele ocupa. 2D e 3D são usadas para diferenciar o tipo de região espacial ocupada.</p> <p>Sinônimo: located_in at some time</p> <p>Domínio: independent_continuant</p> <p>Imagem: independent_continuant</p> <p>Exemplo de uso: portion of blood in capillary</p>

				<p>located_in portion of blood of human body</p>
12	member_part_of	has_member	BFO	<p>Definição textual: Tipo de relação part-of verificada entre dois continuantes tal que a parte é um objeto físico que compõem o todo, o qual é um objeto social não-físico. b member_part_of c at t =Def. b is an object & there is at t a mutually exhaustive and pairwise disjoint partition of c into objects x1, ..., xn (for some n > 1) with b = xi for some 1 ≤ i ≤ n.</p> <p>Definição formal: $\forall x,y(\text{member_of}_n(x,y) \equiv \text{mpart_of}(x,y) \wedge (\text{POB}(x) \vee \text{SOB}(x)) \wedge \text{SOB}(y))$, onde POB representa physical objects e SOB representa non-physical social objects.</p> <p>Sinônimo: member_part_of at some time</p> <p>Domínio: independent_continuant</p> <p>Imagem: independent_continuant</p> <p>Exemplo de uso: erythrocytes member_part_of whole portion of blood</p>
13	occupies_spatial_region at some time	has_spatial_occupant at some time	BFO	<p>Definição: b occupies_spatial_region r at t means that r is a spatial region in which independent continuant b is exactly located.</p> <p>Domínio: independent_continuant</p> <p>Imagem: spatial_region</p>
14	occupies_spatiotemporal_region	has_spatiotemporal_occupant	BFO	<p>Definição: p occupies_spatiotemporal_region s. This is a primitive relation between an occurrent p and the spatiotemporal region s which is its spatiotemporal extent.</p> <p>Domínio: occurrent</p> <p>Imagem: spatiotemporal_region</p>
15	occurs_in	contains	BFO	<p>Definição: b occurs_in c =def b is a process and c is a material entity or immaterial entity & there exists a spatiotemporal region r and b occupies_spatiotemporal_region r. & forall(t) if b exists_at t then c exists_at t & there exist spatial regions s and s' where & b spatially projects_onto s at</p>

				<p>t & c is occupies_spatial_region s' at t & s is a proper_continuant_part_of s' at t</p> <p>Domínio: process or process_boundary</p> <p>Imagem: site or material_entity</p> <p>Propriedades básicas: Funcional, Assimétrica e Reflexiva</p>
16	participates_in	has_participant	RO (RO_ID: 0056)	<p>Definição textual: A relation between a continuant and a process, in which the continuant is somehow involved in the process.</p> <p>Definição formal: $\forall x,y(\text{participates_in}_{it}(x,y) \equiv \text{mpart_of}(x,y)^{\text{ED}}(x)^{\text{PD}}(y))$ onde ED é um endurante ou continuante e PD é um perdurante ou ocorrente.</p> <p>Domínio: continuant</p> <p>Imagem: occurrent</p> <p>Exemplo de uso: plasma-extractors participates_in process of separation of the plasma this blood clot participates_in this blood coagulation</p>
17	part_of_continuant at all times	has_continuant_part at all times	BFO	<p>Definição: C part of continuant at all times C1 b continuant_part_of c at t =Def. b is a part of c at t & t is a time & b and c are continuants. (axiom label in BFO2 Reference: [002-001])</p> <p>Domínio: continuant</p> <p>Imagem: continuant</p> <p>Propriedades básicas: Reflexiva, Transitiva, Assimétrica</p> <p>Exemplo de uso: - portion of blood 'has continuant part at all times' some 'blood cell' and 'has continuant part at all times' exactly 1 'portion of plasma'.</p>
18	part_of_continuant at some time	has_continuant_part at some time	BFO	<p>Definição textual: Tipo de relação part-of verificada entre continuantes ou entre ocorrentes a ser usada quando a parte em algum instante de tempo está localizada como parte do todo, mas não durante toda sua existência.</p>

				<p>Domínio: continuant</p> <p>Imagem: continuant</p> <p>Propriedades básicas: Reflexiva, Transitiva, Assimétrica</p> <p>Sinônimo: temporary_part_of</p> <p>Exemplo de uso: blood component for transfusion part_of_continuant at some time portion of body substance</p> <p>portion of blood collected by venipuncture part_of_continuant at some time portion of body substance</p>
19	part_of_occurent	has_occurent_part	BFO	<p>Definição textual: b occurent_part_of c =Def. b is a part of c & b and c are occurents. (axiom label in BFO2 Reference: [003-002]).</p> <p>Domínio: occurent</p> <p>Imagem: occurent</p> <p>Propriedades básicas: Reflexiva, Transitiva, Assimétrica</p> <p>Exemplo de uso: centrifugation at high rotation part_of_occurent process for obtaining platelets concentrate</p>
20	preceded_by	precedes	BFO (BFO_ID:0062)	<p>Definição: P preceded_by P1 $\forall p$, se p instance_of P então há algum p1 tal que p1 instance_of P1 and p preceded_by p1, onde p preceded_by p1 = $\forall t, t1$, se p occurring_at t and p1 occurring_at t1, então t1 earlier t, onde t earlier t1 é uma relação primitiva entre dois tempos tal que t ocorre antes de t1 e p occurring_at t = for some c, p has_participant c at t.</p> <p>Domínio: process</p> <p>Imagem: process</p> <p>Propriedades básicas: Reflexiva, Transitiva</p> <p>Exemplos de uso: - Process of centrifugation at high rotation preceded_by Process of</p>

				centrifugation. - Process of collection preceded_by Process of centrifugation.
21	prepared_by	prepares	HEMONTO	<p>Definição: C prepared_by P Relação ontológica identificada entre um continuante C e um ocorrente P, tal que o continuante C é obtido ou gerado a partir da realização do processo (ocorrente) P.</p> <p>Domínio: material_entity</p> <p>Imagem: process</p> <p>Exemplo de uso: pooled platelet is_a 'blood component for transfusion' prepared by combining some platelets into one container.</p>
22	prevents	prevent_of	HEMONTO	<p>Definição: P prevents P1 Relação ontológica identificada entre dois ocorrentes, P e P1, sendo que o ocorrente P é uma entidade realizável que evita ou previne a realização do processo (ocorrente) P.</p> <p>Domínio: realizable_entity</p> <p>Imagem: process</p> <p>Exemplos de uso: - anticoagulant is_a realizable_entity and prevents some 'blood coagulation'</p>
23	produces	produced_by	RO (RO_ID: 3000)	<p>Definição: P produces C $\forall p$, se p instance_of P então há algum c e algum t; tal que se c instance_of C1 at t and p produces c at t, onde p produces c at t é uma relação entre o processo p, o continuante c e um tempo t, no qual p produces c se algum processo que occurs_in p has_output c. A produces b if some process that occurs_in a has_output b, where a and b are material entities. Examples: hybridoma cell line produces monoclonal antibody reagent</p> <p>Domínio: process (material_entity)</p> <p>Imagem: material_entity</p> <p>Exemplos de uso: - Process of centrifugation produces Buffy coat.</p>

				- Cryoprecipitate extraction produces Cryoprecipitate free plasm.
24	realizes	realized_in	BFO	<p>Definição: To say that b realizes c at t is to assert that there is some material entity d & b is a process which has participant d at t & c is a disposition or role of which d is bearer_of at t& the type instantiated by b is correlated with the type instantiated by c.</p> <p>Domínio: realizable_entity</p> <p>Imagem: process</p> <p>Exemplos de uso: - water realizes cellular washing.</p>
25	same_as	has_synonym	GO	<p>Definição: Sinonímia, igualdade ou equivalência: Tipo de relação is_a verificada entre duas entidades que são idênticas, podendo ocorrer entre particulares e universais, continuantes e ocorrentes, tal quando a, A são continuantes b, B também devem ser; e se a, A são ocorrentes b, B também devem ser.</p> <p>Domínio: continuant or process</p> <p>Imagem: continuant or process</p> <p>Exemplos de uso: leukocyte same_as white blood cell portion of venous blood same_as blood in vein</p>
26	structural_part_of	has_structural_part	HEMONTO	<p>Definição textual: Tipo de relação part-of verificada entre dois continuantes (universal-universal ou particular-particular) na qual a parte compõem a estrutura do todo, funcionalmente ou estruturalmente.</p> <p>Definição formal: $\forall x, y (s_part_of(x, y) \equiv part_of(x, y) \wedge ED(x) \wedge ED(y))$, onde ED é um endurante ou continuante.</p> <p>Domínio: material_entity</p> <p>Imagem: material_entity</p> <p>Exemplos de uso: circulatory system structural_part_of human body</p>
27	subclass_of	has_subclass	RO	<p>Definição: Especialização ou especificação</p>

				<p>(tipo-subtipo): Tipo de relação is_a verificada entre dois universais (A e B), tal que uma entidade é um tipo da outra. Quando A é um continuante B também deve ser, se A é um ocorrente B deve também ser.</p> <p>Sinônimo: is_a</p> <p>Domínio: continuant or process</p> <p>Imagem: continuant or process</p> <p>Exemplos de uso: - cryoprecipitate subclass-of blood component for transfusio - process of centrifugation at high rotation subclass-of Process of centrifugation.</p>
28	subquantity_of	has_subquantity	HEMONTO	<p>Definição textual: Tipo de relação part-of verificada entre dois continuantes, tal que ambos são quantidades de matéria e a parte é uma porção menor de toda a matéria.</p> <p>Definição formal: $\forall x, y(\text{sub_quantity_of}_n(x, y) \equiv \text{mpart_of}(x, y) \wedge M(x) \wedge M(y))$, onde M é an amount of matter</p> <p>Domínio: material_entity (amount of matter)</p> <p>Imagem: material_entity (amount of matter)</p> <p>Exemplo de uso: platelet subquantity_of platelets concentrate</p>

Fonte: elaborado pelo autor.