

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO

ANDRÉ QUEIROZ DE ANDRADE

**A LINGUAGEM MÉDICA UTILIZADA EM PRONTUÁRIOS E SUAS
REPRESENTAÇÕES EM SISTEMAS DE INFORMAÇÃO: AS
ONTOLOGIAS E OS MODELOS DE INFORMAÇÃO**

Belo Horizonte

2013

ANDRÉ QUEIROZ DE ANDRADE

**A LINGUAGEM MÉDICA UTILIZADA EM PRONTUÁRIOS E SUAS
REPRESENTAÇÕES EM SISTEMAS DE INFORMAÇÃO: AS
ONTOLOGIAS E OS MODELOS DE INFORMAÇÃO**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Informação da Escola de Ciência da Informação da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do título de doutor em Ciência da Informação.

Linha de Pesquisa: OUI

Orientador: Maurício Barcellos Almeida

BELO HORIZONTE

2013

Andrade, André Queiroz de.

A5531 A linguagem médica utilizada em prontuários e suas representações em sistemas de informação [manuscrito] : as ontologias e os modelos de informação / André Queiroz de Andrade. – 2013.

210 f. : il., enc.

Orientador: Maurício Barcellos Almeida.

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

Bibliografia: f. 172-185

Apêndices: f. 186-210

1. Ciência da informação – Teses. 2. Representação do conhecimento (Teoria da informação) – Teses. 3. Ontologias (Recuperação da informação) – Teses. 4. Sistemas de informação gerencial – Teses. I. Título. II. Almeida, Maurício Barcellos. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 025.4:61



UFMG

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

FOLHA DE APROVAÇÃO

"A LINGUAGEM MÉDICA UTILIZADA EM PRONTUÁRIOS E SUAS REPRESENTAÇÕES EM SISTEMAS DE INFORMAÇÃO: AS ONTOLOGIAS E OS MODELOS DE INFORMAÇÃO"

André Queiroz de Andrade

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 - OUI**".

Tese aprovada em: 08 de novembro de 2013.

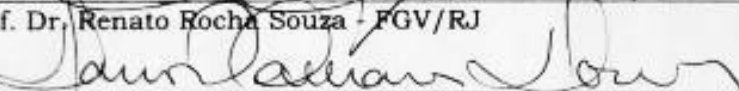
Por:



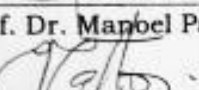
Prof. Dr. Maurício Barcellos Almeida - ECI/UFMG (Orientador)




Prof. Dr. Renato Rocha Souza - FGV/RJ



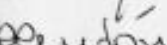
Prof. Dr. Manoel Palhares Moreira - PUC/MG



Prof. Dr. Eduardo de Mattos Pinto Coelho - PBH - Secretaria Municipal de Finanças

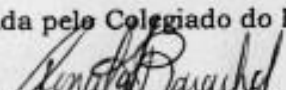


Profa. Dra. Renata Maria Abrantes Baracho Porto - ECI/UFMG



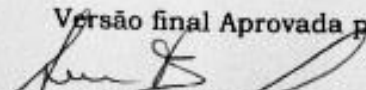
Profa. Dra. Beatriz Valadares Cendón - ECI/UFMG

Aprovada pelo Colegiado do PPGCI



Profa. Renata Maria Abrantes Baracho Porto
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 **ANDRÉ QUEIROZ DE ANDRADE**, matrícula: 2010654689

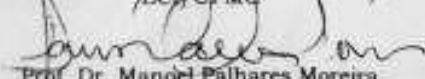
As 14:00 horas do dia 08 de novembro de 2013, reuniu-se na Escola de Ciência da Informação da UFMG a Comissão Examinadora aprovada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação em 11/10/2013, para julgar, em exame final, o trabalho intitulado ***A linguagem médica utilizada em prontuários e suas representações em sistemas de informação: as ontologias e os modelos de informação***, 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:

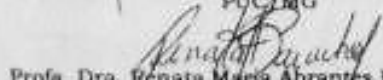
Prof. Dr. Maurício Barcellos Almeida - Orientador	APROVADO
Prof. Dr. Renato Rocha Souza	APROVADO
Prof. Dr. Manoel Palhares Moreira	APROVADO
Prof. Dr. Eduardo de Mattos Pinto Coelho	APROVADO
Profa. Dra. Renata Maria Abrantes Baracho Porto	APROVADO
Profa. Dra. Beatriz Valadares Cendón	APROVADO

Pelas indicações, o candidato foi considerado APROVADO.

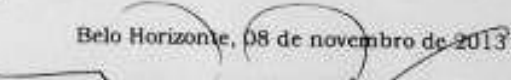
O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a sessão, da qual foi lavrada a presente ATA que será assinada por todos os membros participantes da Comissão Examinadora.

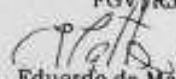

Prof. Dr. Maurício Barcellos Almeida
ECI/UFMG

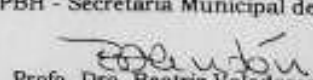

Prof. Dr. Manoel Palhares Moreira
PUC/MG



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

Belo Horizonte, 08 de novembro de 2013


Prof. Dr. Renato Rocha Souza
FGV/RJ


Prof. Dr. Eduardo de Mattos Pinto Coelho
PBH - Secretaria Municipal de Finanças


Profa. Dra. Beatriz Valadares Cendón
ECI/UFMG


Profa. Dra. Renata Maria Abrantes Baracho Porto
Coordenadora do Programa de Pós-Graduação
em Ciência da Informação

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

DEDICATÓRIA

Dedico esta tese à minha família, em especial meus pais e irmãos, sempre presentes, à Alline, minha companheira de todas as horas, e ao Gabriel, que vem por aí.

AGRADECIMENTOS

Agradeço a todos que participaram de alguma forma do desenvolvimento da tese.

Aos colegas da ECI, por conversas valiosas que eventualmente encontraram seu caminho até o texto. Aos professores, pelo rigor científico e discussões filosóficas, em especial à Profa. Dra. Beatriz Cendón, que me acolheu no início do doutorado.

Ao meu orientador Prof. Dr. Maurício Almeida, por me guiar no mar de ambiguidades da pesquisa em ontologias.

Ao Hemominas e ao Blood Project, por fornecer dados para a pesquisa.

Ao Prof. Dr. Stefan Schulz, meu supervisor durante meu estágio de um ano do doutorado-sanduíche, que fez de Graz uma segunda casa, e ainda me recebeu em Português! À Profa. Dra. Andrea Berghold, pela recepção calorosa na Universidade Médica de Graz. Agradeço também aos colegas e amigos Ward, Cati e Markus, por sua contribuição definitiva na direção da tese. Agradeço ainda a oportunidade de ter participado no Semantic Health Net, que forneceu uma visão ampla das discussões atuais e determinou muito do que planejo desenvolver daqui pra frente.

RESUMO

O enorme volume de dados produzido em pesquisa e no curso da atividade médica exige o uso de soluções baseadas em sistemas informatizados. Uma implementação técnica eficaz desta solução exige que sistemas possam manipular e interpretar as mensagens trocadas entre sistemas, armazenar registros, requisitar informações contidas no outro sistema, atender a requisições de informação, dentre outros vários usos para a informação médica. O presente trabalho busca investigar a representação de dados dos pacientes através de modelos de informação e ontologias biomédicas para identificar pontos de ligação entre uma metodologia bem fundamentada, como o realismo ontológico; e uma metodologia flexível, como os modelos de informação em saúde. Por "modelos de informação médica" entendem-se propostas para definição consensual de termos, variáveis e tipos de dados, que objetivam permitir que os dados médicos disponíveis sejam compartilhados entre sistemas diversos; por "ontologias", entende-se a estrutura de categorias, rigorosamente definida, que espelha a organização pretendida para a informação. O objetivo geral da pesquisa é avaliar a representação ontológica de dados do prontuário, a partir da estruturação do conhecimento contido em modelos de informação médicos. A revisão de literatura cobriu o estado da arte sobre padrões de interoperabilidade em medicina, abordou as ontologias como artefatos de sistemas informatizados e como disciplina filosófica, além de apresentar aspectos relacionados a linguagem e a comunicação. Ao longo da realização da parte empírica foram estudados modelos de informação médica (arquétipos) e fragmentos sentenciais de prontuários, utilizando o realismo ontológico como arcabouço para criar uma representação ontológica de modelos e de afirmações sobre pacientes. A pesquisa foi metodologicamente organizada em 3 etapas: 1) Construção do arcabouço teórico; 2) Análise ontológica do modelo OpenEHR; 3) Representação ontológica da informação médica. O arcabouço, baseado na teoria de três mundos de Karl Popper, reconhece a existência de 4 tipos de entidades representacionais: as entidades de caráter ontológico, as entidades de caráter epistemológico, os registros informacionais e os processos de raciocínio. A análise do modelo de arquétipos openEHR demonstrou que existem limites para a representação exclusivamente realista do prontuário médico, mas que esta representação, se conduzida, reduz significativamente a ambiguidade. Concluiu-se que a grande maioria dos fragmentos de informação de cunho médico foi representada adequadamente na ontologia. Observou-se que foi possível representar e recuperar adequadamente a informação médica através de ontologias representadas em lógica descritiva. Concluiu-se ainda que a abordagem ontológica apresenta limitações práticas que devem ser observadas e tratadas de acordo com a tecnologia utilizada e as necessidades de uso. Nesse sentido, os problemas mais imediatos detectados se referem ao tratamento de diferentes tipos de dados por ontologias realistas (por, exemplo, interpretações clínicas e qualidade de processos). Foram encontradas limitações inerentes à tecnologia utilizada (implementações baseadas em lógica descritiva) e mesmo inerentes à maleabilidade de interpretação da linguagem natural por seres humanos, que ainda está longe de ser imitada por sistemas automatizados.

Palavras-chave: Ontologia, Modelos de Informação em Saúde, Registro eletrônico em saúde, Epistemologia

ABSTRACT

The huge volume of data produced in the course of research and medical practice requires the use of computerized systems. Moreover, a technical implementation of these softwares requires effective systems to "understand" the "semantic meaning" of the messages, store records, ask for specific information, meet other requests for information, besides being actually useful for a given medical informatics use case. The present study provides a thorough research on the representation of patient data through information models and biomedical ontologies, in order to identify linking points between a robust methodology, such as the ontological realism, and a flexible methodology, such as health information models. The objective of the research is to evaluate the ontological representation of clinical data, by analyzing the knowledge structure of clinical health information models. The literature review covered the state of the art of interoperability standards in medicine, ontologies as artifacts of computer systems and as a philosophical discipline, language and communication. On the empirical stage, we studied clinical models of medical data representation (archetypes) and sentential fragments of records, using the ontological realism as a framework to create a representation of ontological models and statements about patients. The methodology consisted of three steps: 1) theoretical framework creation, 2) ontological analysis of the openEHR archetype model; 3) ontological representation of medical information. The framework, based on the three worlds theory by Karl Popper, recognizes the existence of four types of representational entity: ontological entities, epistemological entities, informational records and reasoning processes. The openEHR archetype model analysis demonstrated that there are limits to the exclusive realistic representation of medical records, but also that this philosophically sound analysis significantly reduces the ambiguity in representation. Finally, the vast majority of fragments was adequately represented in the ontology. We were able to represent and retrieve information properly through medical ontologies represented in description logics. However, there are practical limitations which must be observed and treated in accordance with the technology used and usage needs. Several difficulties were encountered during the conversion of clinical data into an ontological representation. The more immediate problems relate to the treatment of different types of complex data by realist ontologies. We found limitations inherent to the technology (description logics based implementations) and even the inherent flexibility of natural language interpretation by humans, which is still far from being mimicked by automated systems.

Keywords: Ontology, Health Information Models, Electronic Health Records, Epistemology

LISTA DE FIGURAS

Figura 1 – Espectro de artefatos usualmente denominados "ontologias"	31
Figura 2 – Ontologistas descritivos e formais.	48
Figura 4 – Representação de intensão e extensão.....	74
Figura 5 – A ontologia Clinical Investigator Record (CIR) – adaptada de (BEALE; HEARD, 2007b)	86
Figura 6 – O arquétipo Sintoma.....	88
Figura 7 – Relação esquemática entre o openEHR, CEN 13606 e HL7 CDA.....	89
Figura 8 – Representação esquemática da representação ontológica dos fragmentos sentenciais	113
Figura 9 – Representação esquemática da metodologia utilizada na pesquisa	122
Figura 10 – Abordagem usada na análise	124
Figura 11 – Extrato da união entre OGMS e IAO, demonstrando subclasses da classe "information content entity"	135
Figura 12 – Hierarquia OGMS/IAO/CIR – Care information	138
Figura 13 - Hierarquia OGMS/IAO/CIR – Directive information entity	138
Figura 14 – Extrato da ontologia criada a partir dos conceitos avaliados dos arquétipos OpenEHR, baseado nas ontologias OGMS e IAO (parte 1)	142
Figura 15 - Extrato da ontologia criada a partir dos conceitos avaliados dos arquétipos OpenEHR, baseado nas ontologias OGMS e IAO (parte 2)	143
Figura 16 - Extrato da ontologia criada a partir dos conceitos avaliados dos arquétipos OpenEHR, baseado nas ontologias OGMS e IAO (parte 3)	144

LISTA DE TABELAS

Tabela 1 - Tipo de abordagem para avaliação de ontologias (ALMEIDA, 2009)	97
Tabela 2 – Seções originais dos fragmentos sentenciais selecionados	104
Tabela 3- Extrato da planilha, demonstrando o processo de criação de definições lógicas para os conceitos dos arquétipos (em inglês).....	109
Tabela 4 – Questões de competência originárias da trilha de auditoria para Trombose Venosa Profunda.....	122
Tabela 5 – Comparação de perspectivas ontológicas e não ontológicas sobre a mesma entidade	132
Tabela 6 – Correspondências ad hoc entre a CIR e IAO/OGMS	137
Tabela 7 – Arquétipos selecionados	140
Tabela 8 – Definições lógicas realistas para conceitos de arquétipos OpenEHR	141
Tabela 9 – Classificação dos conceitos segundo as categorias do arcabouço teórico.....	145
Tabela 10 – exemplo do processo de tradução dos prontuários em instâncias OWL a partir dos pares conceito-valor dos arquétipos	153
Tabela 11 – Relação entre a forma de representação das afirmativas (instâncias) e as expressões que recuperam estas instâncias	159
Tabela 12 – Escala de Borg para avaliação de dispnéia (falta de ar) pelo paciente.....	165

LISTA DE QUADROS

Quadro 1 – Definição de significado de acordo com as teorias apresentadas	76
Quadro 2 – Exemplos de fragmentos sentenciais de prontuários de-identificados.....	104
Quadro 3 – Representação dos fragmentos sentenciais em arquétipos	111
Quadro 4 - Criação das instâncias na ontologia de acordo com a definição lógica do conceito correspondente.....	111
Quadro 5 - Exemplos de um extrato de um prontuário médico	128
Quadro 6 - Informações que representam aspectos da realidade.....	128
Quadro 7 - Informações que representam construtos úteis para a prática médica.....	128
Quadro 8 - Informações que representam observações sobre a realidade.....	128
Quadro 9 - Informações que representam observações sobre a compreensão dos médicos.....	129
Quadro 10 - Comparação entre afirmações em diferentes níveis ontológicos	131
Quadro 11 – Relação entre conceitos classificados como ambíguos e definições em linguagem natural e lógica	149
Quadro 12 – Indexação da afirmação no tempo	159

LISTA DE ABREVIATURAS

SIGLA – NOME COMPLETO

BD	Banco de Dados
BFO	Basic Formal Ontology
BID	Duas vezes ao dia
CDA	Clinical Document Architecture
CEN	Comité Européen de Normalisation (Comitê Europeu de Padronização)
CEN	Centre Européen de Normalisation
CGU	Caso Genérico de Uso
CID	Classificação Internacional de Doenças
CIIS	Classificação Internacional de Intervenções em Saúde
CIR	Clinical Investigator Record
DL	Description Logics, ou Lógica Descritiva
DMIM	Domain Message Information Model
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
HL7	Health Level Seven
IA	Inteligência Artificial
IAO	Information Artifact Ontology
ISO	International Organization for Standardization
LOINC	Logical Observation Identifiers Names and Codes
MESH	Medical Subject Headings
OBI	Ontology for Biomedical Information
OBO	Open Biomedical Ontology
OGMS	Ontology for General Medical Science
OMS	Organização Mundial de Saúde
OpenEHR	Open Electronic Health Record
OWL	Ontology Web Language
RDF	Resource Description Framework
RES	Registro Eletrônico em Saúde
RIM	Reference Information Model
RMIM	Refined Message Information Model
SDO	Standards Developing Organizations
SKOS	Simple Knowledge Organization System
SNOMED-CT	Systemized Nomenclature Of Medicine Clinical Terms
SOAP	Subjetivo, Objetivo, Avaliação e Plano
TI	Tecnologia da Informação
UMLS	Unified Medical Language System
W3C	World Wide Web Consortium
WWW	World Wide Web
XML	eXtended Markup Language

SUMÁRIO

1	PREFÁCIO.....	16
2	INTRODUÇÃO	17
3	REVISÃO DE LITERATURA.....	26
3.1	ONTOLOGIAS BIOMÉDICAS – DESENVOLVIMENTO E OBJETIVOS	26
3.1.1	O que é ontologia - definições	26
3.1.2	Objetivos das ontologias	33
3.1.3	Sistemas de classificação e sua evolução em ontologias.....	34
3.1.4	Representação simbólica de conhecimento e Lógica Descritiva.....	38
3.1.5	Ontologias e a Web Semântica	44
3.2	O PAPEL DAS ONTOLOGIAS FORMAIS – ABORDAGEM CRÍTICA DO REALISMO COMO TÉCNICA DE MODELAGEM.....	46
3.2.1	Filosofia, Ciência da Informação e Ciência da Computação	46
3.2.2	Realismo ontológico	50
3.2.3	Karl Popper e a teoria dos três mundos.....	55
3.2.4	Vantagens e limitações do realismo ontológico	59
3.3	A TROCA DE MENSAGENS EM BIOMEDICINA – SEMIÓTICA E MODELO DE INFORMAÇÃO EM SAÚDE	69
3.3.1	Semiótica e significado.....	70
3.3.2	Modelos de Informação em Medicina	80
3.3.3	Linguagem e Conhecimento.....	91
4	METODOLOGIA	95
4.1	CONSIDERAÇÕES INICIAIS	95
4.2	OBJETOS DE PESQUISA E SELEÇÃO DA AMOSTRA	101
4.3	DETALHAMENTO DAS ETAPAS DA PESQUISA.....	105
4.3.1	Construção do arcabouço teórico	106
4.3.2	Análise ontológica do Modelo OpenEHR.....	106
4.3.3	Representação ontológica da informação médica	109
4.3.4	Critérios de avaliação da modelagem ontológica.....	113
5	CONSTRUÇÃO DO ARCABOUÇO TEÓRICO	124
5.1	ABORDAGEM PARA ANÁLISE.....	124
5.2	CONCLUSÃO E IMPLICAÇÕES PARA A PESQUISA	129
6	ANÁLISE ONTOLÓGICA DO MODELO OPENEHR	133
6.1	COMPARAÇÃO DO MODELO DE ALTO-NÍVEL	133

6.2	MAPEAMENTO DOS CONCEITOS CONTIDOS NOS ARQUÉTIPOS CLÍNICOS	139
6.3	IMPLICAÇÕES PARA A PESQUISA.....	147
7	REPRESENTAÇÃO ONTOLÓGICA DA INFORMAÇÃO MÉDICA.....	153
7.1	DISCUSSÃO E QUESTÕES	154
7.1.1	Questão 1 – Como lidar com afirmativas que excedem as capacidades da lógica descritiva sob a ótica realista?.....	154
7.1.2	Questão 2 – As afirmativas devem ser representadas como classe ou como instância?	157
7.1.3	Questão 3: Como representar relações temporais?.....	159
7.1.4	Questão 4: Como representar entidades que podem ser descritas, pelo médico, de várias formas iguais para o raciocínio, mas diferentes ontologicamente	163
7.1.5	Questão 5: Como representar atributos (qualidades) de processos	164
7.1.6	Questão 6: Como representar ontologicamente entidades mentais?.....	164
7.2	CONCLUSÃO.....	168
8	CONSIDERAÇÕES FINAIS E CONTRIBUIÇÕES DA TESE	170
9	BIBLIOGRAFIA	173
10	APÊNDICE.....	187
10.1	CASO DE USO – TROMBOSE VENOSA PROFUNDA.....	187
10.2	PROCESS ATTRIBUTES IN BIOMEDICAL ONTOLOGIES – PREPRINT VERSION	192
10.2.1	Ontological Realism and their problem with process “qualities”	193
10.2.2	Preliminary evaluation	196
10.2.3	Results	199
10.2.4	Discussion.....	208
10.2.5	Final remarks.....	212

1 PREFÁCIO

Escrever sobre um assunto interdisciplinar que envolve ciência da informação e saúde requer um grande esforço. Não só requer um novo conjunto de habilidades, mas também uma nova forma de pensar para olhar os problemas a partir de perspectivas diferentes. Como consequência, esta tese é um trabalho circular, escrita dentro dos ciclos de aprofundamento da análise teórica e aplicação pragmática. Alguns trechos podem conter argumentos puramente filosóficos, sem aplicação prática clara. Outros podem conter partes com aplicações novas de tecnologias já conhecidas. No entanto, as partes são complementares para alcançar o objetivo final, que é organizar informações em saúde. Este objetivo não foi totalmente alcançado aqui, mas esperamos ter caminhado a passos concretos nesse sentido.

Dito isso, cabe destacar que esta tese é um *remix* de centenas de brilhantes pesquisas anteriores em ciência da informação, inteligência artificial, ontologia aplicada, filosofia e informática em saúde. Agradecemos aos pesquisadores que contribuíram para o desenvolvimento destes campos diversos. Mais do que uma mudança paradigmática kuhniana, procuramos encontrar um meio termo de pacífica co-existência de dois paradigmas em informática em saúde: uma compartilhada por ontologistas formais e outra compartilhada por desenvolvedores de terminologias e padrões em informática em saúde. Estamos confiantes de que o melhor desses dois mundos pode ser alcançado.

2 INTRODUÇÃO

Informações estruturadas no registro eletrônico em saúde (RES) ¹ são um componente essencial no desenvolvimento de aplicações compartilhadas entre profissionais de saúde. Tal estruturação permite a troca de informações entre diferentes sistemas automatizados e a comunicação entre profissionais, mantendo o significado original do registro. Os registros estruturados são também essenciais para o desenvolvimento de aplicações cujo objetivo é manipular dados em ambientes heterogêneos, ou seja, sem definição clara e consensual do significado destes dados em diferentes sistemas. Tais situações são comuns atualmente, em função da diversidade de formas em que a informação médica é utilizada. Para ilustrar esse fato, cabe mencionar algumas situações: registro e posterior uso em pesquisa científica; comunicação entre profissionais no atendimento ao paciente; registro e reuso da informação sobre um paciente por outros profissionais, em outros locais, em outros períodos; gerenciamento de instituições de saúde.

O enorme volume de dados produzido em pesquisa e no curso da atividade médica, exige o uso de soluções baseadas em sistemas informatizados. Além disso, uma implementação técnica eficaz desta solução exige que sistemas possam manipular e interpretar as mensagens entre sistemas, armazenem registros, requisitem informações contidas no outro sistema, atendam a requisições de informação, dentre outros vários usos mencionados acima para a informação médica. Os termos “semântica” e “significado” são passíveis de interpretações diversas e, no contexto desse trabalho, são utilizados conforme a acepção descrita em Almeida, Souza e Fonseca (2011b), descritos na seção 3.3. No presente trabalho, propõem-se investigar as possibilidades de representação de informações médicas, no contexto de heterogeneidade descrito, considerando-se ainda os diversos tipos de agentes envolvidos – sistemas criados para diferentes fins e profissionais com formações e objetivos diversos. Aborda-se a questão a partir de propostas, aqui consideradas complementares, para compartilhamento de informações: os modelos de informação médicos e as ontologias. No restante dessa seção, examinam-se brevemente essas duas propostas.

¹ Reservamos o termo RES para os sistemas informatizados de representação médica e o termo prontuário médico para as representações feitas em papel

Por “modelos de informação médica” entendem-se propostas para definição consensual de termos, variáveis e tipos de dados, com o objetivo de permitir que os dados médicos disponíveis sejam compartilhados entre sistemas diversos. Esta definição conjunta que resulta em um modelo é, na maioria das vezes, orientada por casos de uso reais e de necessidades práticas de informação. É comum nesses modelos, por exemplo, a existência de campos padronizados para itens necessários à cadastros administrativos, como “nome” e “data de nascimento”; e itens utilizados para controle por computadores, como o “id” (um tipo de identificador).

O escopo dos modelos é modificado de acordo com as necessidades de sistemas e, atualmente, as principais utilizações dos modelos dizem respeito à definição de dados do prontuário de forma detalhada. A escolha dos termos é feita através de instituições que congregam os agentes interessados no compartilhamento das informações, usualmente representantes do governo, da academia, de empresas e instituições de saúde. A principal vantagem dos modelos é exatamente a criação do consenso orientada à prática, uma vez que o modelo, após definido, é publicado e disseminado para uso. Dentre os diversos modelos existentes, cabe citar os três principais, assim caracterizados por sua abrangência de adoção: o HL7 v3 RIM (*Reference Information Model*) e o OpenEHR (*Open Electronic Health Record*) e a ISO/CEN 13606 *standard*.

Apesar da utilidade dos modelos de informação médica, diversos problemas surgem na obtenção de consenso necessário para a definição de termos e variáveis. Smith e Ceusters (2006), por exemplo, apontam algumas questões no âmbito do HL7 v3 RIM, como problemas de implementação e uso, usabilidade restrita à domínios de especialistas, limitação de escopo, deficiência na documentação, dificuldade na aprendizagem e influência de *marketing*. Outros trabalhos apontam a grande complexidade e custo de implementação do HL7 v3 (NEHTA, 2006), a inerente ambigüidade do padrão – resultante do processo democrático de criação (SCHADOW *et al.*, 2006) – e problemas de modelagem que surgem ao se conectarem fluxo de trabalho e conhecimento de domínio médico em um mesmo modelo de informação (DE ROOS, 2007).

No escopo do presente trabalho, a limitação mais importante é a ambigüidade inerente às definições das variáveis, pela dificuldade que ela traz para comunicação no

âmbito das práticas médicas. Por exemplo, a documentação do RIM (HL7, 2010) descreve a classe “*Acts*” como “uma coleção de classes relacionadas a ações e eventos que constituem serviços de atenção à saúde”², e a classe “*Act*” (pertencente a “*Acts*”) como um “registro de algo que está sendo feito, foi feito, pode ser feito, ou é planejado ou solicitado a ser feito”³. Nesse exemplo, percebem-se pelo menos duas fontes importantes de ambigüidade: 1) a relação entre “*Acts*” e “*Act*” não é explicitamente definida, e não é intuitivo *a priori* se todos os “*Acts*” são um “*Act*”; 2) a definição de “*Act*” é ambígua, pois se refere ao registro e não ao ato em si (SMITH; CEUSTERS, 2006). Outro modelo de informação utilizado no domínio da hematologia, o ISBT-128 *Standard*, proporciona outro exemplo de ambigüidade (ALMEIDA *et al.*, 2011a): membros da classe “*blood components*” podem ter modificadores, como “*deglycerolized*” e “*deglycerolized rejuvenate*”. O modificador “*deglycerolized*” é definido como “a remoção de glycerol por lavagem”, e o modificador “*deglycerolized rejuvenate*” é definido como “um produto em que as células são rejuvenecidas, glycerol é adicionado e então congelado, e subseqüentemente descongelado e deglicerolizado” (ICCBBA, 2010). Pode-se notar que o mesmo modificador executa duas funções distintas, representando em um caso um processo (“*the removal*”) e em outro um objeto (“*the product*”).

A segunda proposta aqui analisada aborda as ontologias. Ontologia diz respeito ao estudo do que existe, das estruturas dos objetos, propriedades, eventos, processos e relações na realidade (SMITH; WELTY, 2001). Entretanto, tal termo tem sido usado na Ciência da Informação para se referir a a estrutura de categorias, rigorosamente definida, que espelha a organização pretendida para a informação. Trata-se, de fato, de um assunto multidisciplinar que envolve Filosofia da Linguagem, Linguística e Computação. Sua primeira menção na comunidade de inteligência artificial foi feita por *John McCarthy* em 1980. O termo se popularizou nos anos 1990 como uma forma de representação formal de conhecimento, em uma tentativa de criar um sistema de categorias independente de domínio específico (VICKERY, 1997). Para tal, foi reconhecida a importância de utilizar abordagens fundamentadas no estudo filosófico milenar. Como Guarino e Poli esclarecem, “uma base de conhecimento pode adquirir valor *per se* ao se tornar transparente, no sentido que seus compromissos em relação ao mundo (...) são explícitos, e o conteúdo do

² A collection of classes including the Act class and its specializations. These relate to the actions and events that constitute health care services.

³ A record of something that is being done, has been done, can be done, or is intended or requested to be done.

conhecimento pode ser mapeado nas estruturas de tal mundo externo”⁴ (GUARINO; POLI, 1995).

Além de seu uso tradicional em Filosofia, as ontologias começaram a ser usadas exclusivamente na representação de conhecimento, com fins de inferência automática⁵. Em seguida, foram também utilizadas por pesquisadores de bancos de dados, de linguagens orientadas a objetos como uma abordagem de modelagem para sistemas de informação (WAND; WEBER, 1990), bem como no contexto da Web Semântica (BERNERS-LEE, 1998). Atualmente, ontologias têm sido utilizadas em várias aplicações e em vários contextos, inclusive como referência para organização da informação. Podem ser criadas em diversos graus de formalidade, sendo que o mais rigoroso tem como objetivo formular axiomas básicos que descrevem: 1) as propriedades básicas que todas as entidades de um determinado tipo têm em comum; 2) os critérios necessários que devem ser atendidos para que uma entidade “x_1” seja instancia de um tipo “X”. Um exemplo para o caso 1) é: todas as células têm uma membrana, ou que todas as moléculas de glicose contêm carbono; um exemplo para o caso 2) é que todas as instâncias de "Hepatite" são instâncias de "doença inflamatória" localizadas em um fígado (e vice versa).

A disseminação de ontologias favorece a comunicação (entre agentes, sejam humanos ou máquinas) e seu uso pode incluir busca semântica, especificação da semântica dos dados em bancos de dados, integração de dados em múltiplas fontes, criação de linguagens de comunicação para agentes, vocabulários controlados para registro de diagnósticos médicos, dentre outros (GRUNINGER *et al.*, 2008). O grande número de aplicações possíveis gerou um número também grande de metodologias e abordagens para criação de ontologias, com graus variados de robustez e consistência para fins de modelagem e de representação. Para os objetivos supramencionados – compartilhamento de dados do prontuário – é necessário contar com uma abordagem que defina explicitamente o significado dos termos, e que permita realizar as inferências necessárias durante a solicitação de informações clínicas. Por exemplo, quem busca casos de doenças inflamatórias em partes do sistema digestivo pode receber como retorno casos de hepatite.

⁴ a knowledge base can in principle acquire value per se to the extent that it is transparent, in the sense that its commitments to the world beyond the knowledge base are explicit, and the knowledge content can be actually mapped onto the structure of such an external world

⁵ Considera-se aqui "raciocínio automatizado" e "inferência automatizada" como sinônimos.

Uma vez que co-existem atualmente diversas propostas de ontologias de referência como a BFO (GRENON *et al.*, 2004), DOLCE (GANGEMI *et al.*, 2002) e outras, cujas diferenças em geral envolvem aspectos filosóficos ainda em aberto, é preciso adotar uma referência. Isso ocorre pela necessidade de conectar ontologias com modelos de informação, o que exige que os compromissos ontológicos – a determinação precisa do significado de um termo e sua relação com certa conceitualização (GÓMEZ-PÉREZ *et al.*, 2004) – sejam no mínimo similares. No presente trabalho, optou-se pela metodologia que tem recebido o nome do “realismo ontológico”, conforme descrito em (SMITH, 2006; SMITH; CEUSTERS, 2010).

O assim denominado realismo ontológico propõe uma metodologia formal para garantir que os desenvolvedores de ontologias mantenham consistência na modelagem, ao adotarem o compromisso ontológico com a realidade. No escopo do presente trabalho, associa-se o termo “realismo ontológico” a uma metodologia para desenvolvimento de ontologias que se baseia nos princípios filosóficos do realismo para modelagem dos artefatos (ver seção 3.2.2). A escolha foi motivada pela presença de uma teoria detalhada e descrita em inúmeros artigos (SMITH, 2003a; GRENON *et al.*, 2004; SMITH, 2004; SMITH *et al.*, 2005; SMITH, 2006), e por seu uso consagrado por uma comunidade ampla e ativa na pesquisa biomédica, que já conta com um abrangente repositório de ontologias⁶ (SMITH *et al.*, 2007).

Por outro lado, a adoção de ontologias pela comunidade médica é significativamente inferior quando comparada à comunidade de pesquisadores em biologia. Cimino (2006) argumenta que o realismo é insuficiente para descrever as entidades que os clínicos usam para comunicação. Esse fato ocorre pois tal abordagem parece ter sido idealizada para a descrição do conhecimento científico, como a descrição de genes e funções biológicas das proteínas (HILL *et al.*, 2008). Ao ser utilizada para comunicação, a abordagem realista parece sofrer de limitações na representação de aspectos da linguagem natural que não têm referente direto no mundo. Denominam-se estes aspectos, genericamente, de epistemológicos e contextuais, seguindo a nomenclatura proposta por Bodenreider e outros (2004). Por exemplo, um clínico que usa o termo “hepatite” nem sempre se refere a uma hepatite real, mas pode referir-se a uma suspeita (de um paciente ter essa doença), assim

⁶ Ver <http://www.obofoundry.org/>

como a uma ação preventiva (como uma vacinação). Nesse contexto, não parece direta a correspondência com a realidade das ciências naturais e a possibilidade de compromisso ontológico pelo menos similar.

Além disso, Schulz e outros (2009b) argumentam que a tentativa de codificar conhecimento que os próprios autores denominam *default* ou *probabilístico* utilizando ontologias, da forma como são aplicadas atualmente na abordagem realista, tem grandes chances de produzir modelos incorretos. De fato, a abordagem ontológica realista não parece capaz de representar adequadamente afirmativas que não sejam universalmente verdadeiras, como “suspeita de febre”, “história de febre” ou “sem febre”. Cabe mencionar que, enquanto os modelos de informação buscam definir o conteúdo de mensagens, ou seja, enfatizam a comunicação na prática clínica, o realismo ontológico refuta questões da linguagem e referencia universais ou tipos existentes na realidade independentemente do nome atribuído.

Visto a importância e a abrangência de uso das duas propostas – modelos de informação médicos e ontologias realistas – acredita-se na necessidade de identificar formas para diminuir o hiato entre as duas abordagens, reduzindo as limitações e propondo complementaridade. Nessa direção, Stenzhorn e outros (2008) argumentam que o uso de ontologias para registro e comunicação clínica exige o uso concomitante de modelos de informação, como os já mencionados HL7 e OpenEHR. Entretanto, tal tarefa não é simples e encontra-se em aberto, uma vez que existe sobreposição entre a informação expressa por ontologias e por modelos de informação, além de dificuldade na interpretação das mensagens (RECTOR, 2010). Como exemplo, vale citar o projeto TERMINFO⁷, uma tentativa de identificar a superposição de conteúdo entre ontologias clínicas e modelos de informação que sugeriu heurísticas de evitar incompatibilidades; e o projeto Semantic HealthNet⁸, cujo objetivo é desenvolver uma organização pan-européia escalável e sustentável para a interoperabilidade semântica de conhecimento clínico e biomédico. O autor do presente trabalho teve a oportunidade de participar deste último projeto durante o estágio de doutorado no exterior na Universidade Médica de Graz, sob a supervisão do Prof. Stefan Schulz, e vários achados descritos neste trabalho são frutos direto desta colaboração.

⁷ <http://gforge.hl7.org/gf/project/terminfo/>

⁸ <http://www.semantichealthnet.eu/>

O presente trabalho busca uma investigação aprofundada das questões descritas até aqui, ou seja, a representação de dados dos pacientes através de modelos de informação e ontologias biomédicas. Dessa forma, espera-se identificar pontos de ligação entre uma metodologia bem fundamentada, como o realismo ontológico, e uma metodologia flexível, como os modelos de informação em saúde.

O objeto desta tese é amplo – estudar a representação da linguagem médica utilizada em um prontuário médico em sistemas de informação, utilizando ontologias e modelos de informação. O objetivo geral da pesquisa é *avaliar a representação ontológica de dados do prontuário, a partir da estruturação do conhecimento contido em modelos de informação médicos*. Durante a pesquisa, pretende-se explorar a natureza ontológica de dados do prontuário representados por modelos de informação, avaliar a capacidade de representar estes dados por ontologias realistas e propor mecanismos que explorem a complementaridade entre modelos de informação e ontologias. Tal complementaridade permitirá, por exemplo, identificar dados em uma abordagem que não podem ser adequadamente mapeados diretamente na outra. Por exemplo, dados de finalidade administrativa presentes nos modelos de informação, ou mesmo necessários à prática médica, não encontram contrapartida nas ontologias de referências da abordagem realista.

Como objetivos específicos, definem-se:

- Propor uma arcabouço teórico para classificação ontológica das informações contidas em prontuários médicos;
- Identificar a sobreposição entre modelos de informação e ontologias de referência orientadas pelo realismo ontológico, por exemplo, a Basic Formal Ontology (BFO);
- Propor metodologia para representação consistente de informações clínicas através de extensões do realismo ontológico
- Testar a metodologia proposta a partir de uma série de fragmentos retirados de prontuários médicos reais.

Para atingir estes objetivos, a pesquisa foi dividida em três etapas:

1. Construção do arcabouço teórico
2. Análise ontológica do modelo OpenEHR
3. Representação ontológica da informação médica

Afim de fornecer base empírica para embasar a transformação entre modelos de informação e ontologias, foram extraídos fragmentos sentenciais de prontuários previamente anonimizados, coletados durante a execução do Blood Project. O Blood Project é um projeto colaborativo entre o Ontology Research Group da State University of New York at Buffalo, a Fundação Hemominas e a Escola de Ciência da Informação da Universidade Federal de Minas Gerais.

Quatro hipóteses orientaram a análise dos resultados:

- H1 - Ontologias podem expressar o mesmo conteúdo de modelos de informação
- H2 - A representação ontológica reduz ambiguidade conceitual e permite inferências robustas
- H3 - É possível recuperar informação de saúde representada ontologicamente através de operações em lógica descritiva
- H4 - O processo de representação da linguagem médica por ontologias é custo-efetivo no desenvolvimento de sistemas de informação em saúde

A presente pesquisa demonstrou que o alinhamento entre modelos de informação e ontologia é possível para a maior parte das informações registradas no prontuário. Os modelos analisados mostraram-se heterogêneos e podem ser melhorados com regras simples oriundas do realismo. Por outro lado, a representação ontológica não captura todas as nuances da comunicação médica e requer complementações para ser fiel ao prontuário médico.

O restante do trabalho está organizado como segue. A seção 3 apresenta a revisão da literatura sobre ontologias biomédicas e modelos de informação em saúde, com ênfase nas ontologias formais oriundas do realismo científico e nos modelos de informação em dois níveis. A seção 4 apresenta os métodos de pesquisa utilizados na pesquisa. Em seguida, cada um dos sub-itens do detalhamento dos métodos de pesquisa é descrito em uma seção independente. A seção 5 apresenta os resultados da construção de um arcabouço teórico que reúne informação e realidade, baseado na teoria dos três mundos de Karl Popper. A seção 6 apresenta os resultados do alinhamento entre o modelo openEHR e ontologias construídas sob o paradigma do realismo ontológico. A seção 7 apresenta uma representação prática de asserções do prontuário médico, demonstrando os benefícios e

limitações da representação ontológica. Finalmente, a seção 8 sumariza as contribuições da pesquisa.

3 REVISÃO DE LITERATURA

Nesta seção, a revisão de literatura é apresentada. O tema "ontologias biomédicas" é explorado na seção 3.1, apresentando um breve histórico de sistemas de classificação em medicina e sua evolução em sistemas terminológicos complexos. As ontologias formais são apresentadas na seção 3.2, com ênfase na descrição e fundamentação filosófica da metodologia do realismo ontológico. A seção 3.2 apresenta os conceitos de semiótica e significado utilizados neste trabalho, que servem de partida para a descrição e discussão dos modelos de informação em medicina.

3.1 ONTOLOGIAS BIOMÉDICAS – DESENVOLVIMENTO E OBJETIVOS

Ontologias biomédicas têm se mostrado muito importantes no desenvolvimento de sistemas de troca de conhecimento em biologia e medicina, essenciais para realização de estudos multicêntricos e organização do conhecimento crescente em genética, imunologia, dentre outros campos de estudo biológico. Nesta seção, explora-se o tema, incluindo definições, objetivos, exemplos e tecnologias para manipulação e uso de ontologias.

3.1.1 O que é ontologia - definições

A palavra ontologia, que significa o estudo da essência do ser ("ontos": ser, existência; "logos": estudo de), pode ser encontrada desde 1613, tendo sido criada por dois filósofos, Rudolf Gockel (Goclenius), no seu *Lexicon philosophicum* e Jacob Lorhard (Lorhardus), em seu *Theatrum philosophicum*. No entanto, a busca filosófica por uma teoria do ser é bem mais antiga, indo até os primeiros filósofos gregos. De acordo com Smith and

Welty (2001, p.1), a “ontologia filosófica é a ciência do que é, dos tipos e estruturas dos objetos, propriedades, eventos, processos e relações em todas as áreas da realidade⁹”. De forma mais pragmática, Guarino (1998, p.4) referencia ontologias no sentido filosófico como “sistemas particulares de categorias representando uma determinada visão do mundo”¹⁰. Fonseca (2007) distingue dois tipos de ontologias: ontologias de sistemas de informação e ontologias para sistemas de informação. O primeiro tipo utiliza os métodos e teorias ontológicas para encontrar os construtos básicos de sistemas de informação, com o objetivo de criar métodos consistentes de modelagem que assegurem compartilhamento (ALMEIDA *et al.*, 2009). O segundo tipo descreve o vocabulário de um domínio para criação de esquemas conceituais, para ser usado como um artefato de sistemas de informação, representando o conhecimento sobre um assunto.

O tema das ontologias (no sentido de um artefato usado em sistemas de informação) adquiriu fama nos anos 90, ganhando popularidade nos primeiros anos do século 21. A palavra “ontologia” foi introduzida na ciência da computação por McCarthy (1980), ao descrever um arcabouço necessário para inferências com senso comum. Mesmo após 30 anos, a definição precisa do termo ainda é objeto de conferências e argumentos (GRUNINGER *et al.*, 2008). Definidas como uma “especificação de uma conceitualização” (GRUBER, 1995), ontologias podem ser entendidas como uma descrição formal das coisas que existem em um determinado domínio. Guarino refina esta distinção ao fazer explícita a diferença entre uma conceitualização e uma ontologia:

“Uma ontologia é uma teoria lógica representando o significado intensional de um vocabulário formal, ou seja, seu comprometimento ontológico a uma conceitualização particular do mundo. Os modelos intensionais de uma linguagem lógica usando tal vocabulário são restritos por este comprometimento ontológico. Uma ontologia indiretamente reflete este comprometimento (e a conceitualização subjacente) ao aproximar estes modelos¹¹” (GUARINO, 1998, p.7)

⁹ “(p)hilosophical ontology is the science of what is, of the kinds and structures of objects, properties, events, processes and relations in every area of reality

¹⁰ “A particular system of categories accounting for a certain vision of the world”

¹¹ “An ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualization of the world. The intended models of a logical language using such a vocabulary are constrained by its ontological commitment.

De acordo com Fonseca, a definição mais comum de ontologia é uma “teoria que explica um domínio¹²” (FONSECA, 2007, p.791). Chandrasekaran e outros também descrevem ontologia como um vocabulário descrevendo uma conceitualização e o conhecimento que descreve algum domínio. Nesta visão, ontologias são “teorias de conteúdo sobre os tipos de objetos, propriedades dos objetos e relações entre objetos que são possíveis em um domínio especificado de conhecimento”¹³. (CHANDRASEKARAN *et al.*, 1999, p.20). Smith também define ontologias como “uma teoria formal na qual não apenas definições, mas também um arcabouço de axiomas que dê suporte estão incluídos (talvez os axiomas forneçam definições implícitas dos termos envolvidos)”¹⁴. (SMITH, 2003a, p.158). Ainda outra definição por Stenzhorn e outros formula que “ontologias formais são teorias que buscam fornecer representações precisas dos tipos de entidades na realidade, suas propriedades e as relações entre eles, usando axiomas e definições capazes de processamento algorítmico”¹⁵ (STENZHORN *et al.*, 2008, p.3768).

O desenvolvimento errático da engenharia de conhecimento e a diversidade de objetivos dos pesquisadores que utilizam lógica, semântica e conceitos de Inteligência Artificial (IA) criaram um grande espectro de artefatos denominados ontologia. Este cenário tornou-se ainda mais complexo com o surgimento da comunidade de Web Semântica, que viu nas ontologias um artefato perfeito para definir significado para compartilhamento de informação. O termo perdeu, portanto, sua ligação com a realidade para se aproximar de vocabulários e terminologias.

Ao invés de aumentar a precisão da definição, a comunidade que trabalha com ontologias procurou criar perspectivas para descrever cada artefato. Durante o Ontology Summit 2007, foram identificadas seis dimensões para melhor caracterizar ontologias (BODENREIDER; OLKEN, 2007; GRUNINGER *et al.*, 2008):

An ontology indirectly reflects this commitment (and the underlying conceptualization) by approximating these intended models.”

¹²“ theory that explains a domain”

¹³ “content theories about the sorts of objects, properties of objects, and relations between objects that are possible in a specified domain of knowledge”

¹⁴ “a formal theory within which not only definitions but also a supporting framework of axioms is included (perhaps the axioms themselves provide implicit definitions of the terms involved)”

¹⁵ “formal ontologies are theories that attempt to give precise representations of the types of entities in reality, of their properties and of the relations between them, using axioms and definitions that support algorithmic reasoning”

1. **Expressividade** uma propriedade da linguagem de representação de conhecimento que descreve a amplitude e a facilidade com que a linguagem pode descrever cada vez mais semântica complexa, ex. lógica proposicional, lógica descritiva, a lógica de primeira ordem, lógica de classificação, lógica modal, e assim por diante. Ontologias também podem ser comparadas de acordo com a expressividade mínima necessária para definir seus vocabulários. Por exemplo, apesar de ser possível especificar uma taxonomia em linguagens de lógica complexa, esta especificação exige apenas linguagens semiformais mais simples, como XML, de forma a se tornar uma alternativa prática.
2. **A estrutura** é uma propriedade da ontologia, que registra quão elaborada (ou bem organizada) está a semântica codificada pelo ontologia. Ela corresponde à formalidade de definições para os termos do vocabulário. Outra forma de visualizar a estrutura é caracterizar uma ontologia de acordo com a medida em que a interpretação pretendida do vocabulário é definida logicamente. Por exemplo, definições matemáticas formais são geralmente muito estruturadas, com seus conceitos, definidos em declarações lógicas, enquanto que conceitos como "nome" ou "título" são geralmente informais, definida apenas nos rótulos por linguagem natural e restrita por tipo de dados (texto, booleana, etc.).
3. A dimensão **granularidade** diz respeito ao nível de detalhe em que a ontologia é especificada. Uma ontologia com muitas ligações entre os termos, propriedades que detalham conceitos e relações possui baixa granularidade. Uma ontologia com poucas classes primitivas possui alta granularidade. Granularidade é independente da linguagem de representação e estrutura. Por exemplo, ontologias formais de alto-nível, como o BFO, geralmente possuem granularidade mais alta, ou seja, aquela que proporciona organização em um nível genérico. É importante não confundir granularidade como apresentado neste contexto com a definição apresentada por (RECTOR *et al.*, 2006, p.334), que afirma que granularidade é o “grau de coletivização, por exemplo, considerando a água que enche um lago, a

relação “encher” considera água como, entre outras coisas, uma coleção de moléculas de água, não as moléculas individuais”¹⁶.

4. **Uso pretendido** é uma dimensão pragmática, que registra o propósito original da ontologia.
5. **Inferência automatizada** é uma dimensão que registra o quanto espera-se que uma ontologia possa ser usada por *software* de inferência automatizada, por exemplo, sistemas de resposta a questões. Ontologias podem ser comparadas de acordo com os tipos de inferências possíveis, variando desde as capazes de provar teoremas até os motores de inferência *ad hoc*, capazes de interpretar o conteúdo da ontologia. Três tipos de inferências automatizadas foram identificados: Simples (capaz de fazer inferências de relações de subclasse); Especial (classificação das classes em lógica descritiva) e Geral (envolve o uso de regras dedutivas que combinam informações de toda a ontologia). É possível atribuir diversos nomes a esquemas terminológicos de acordo com sua capacidade de raciocínio automatizado. Na Figura 1, mostra-se os diferentes nomes dados a esses artefatos, classificados de acordo com a complexidade da inferência automatizada empregada.
6. **Prescritiva vs. Descritiva** dizem respeito a uma dimensão que caracteriza a fonte das interpretações pretendidas do vocabulário da ontologia. Ontologias descritivas caracterizam termos conforme seu uso linguístico comum, sem a exigência de existência destes objetos representados no mundo. Exemplos de tais ontologias descritivas incluem *folksonomias* e a maioria das ontologias linguísticas. Por outro lado, uma ontologia pode ser concebida como um documento normativo prescritivo cuja veracidade é uma preocupação considerável. Por exemplo, uma baleia é um mamífero e não um peixe. Outras ontologias prescritivas incluem terminologias médicas de diagnóstico, ontologias jurídicas ou regulatórias, as ontologias de contabilidade, matemática ou ontologias da engenharia, etc.

¹⁶ “degree of collectivisation, e.g., with respect to water filling a lake, the relation ‘filling’ is to the water as, amongst other things, a collective of water molecules, not to the individual molecules themselves”.

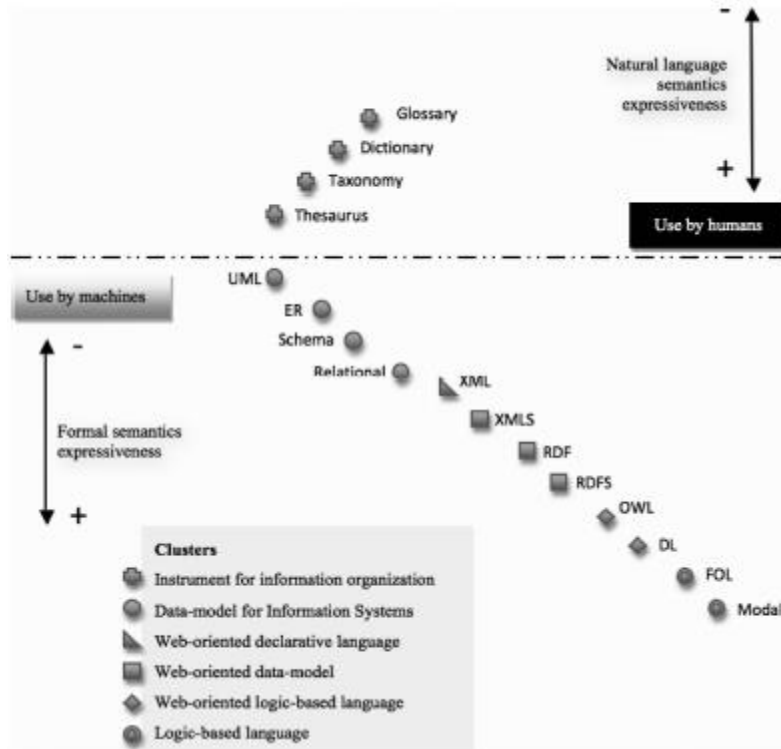


Figura 1 – Espectro de artefatos usualmente denominados "ontologias"

Fonte: Almeida, Souza e Fonseca (2011b)

Considerando estas seis dimensões, um termo melhor para esta definição ampla de ontologia seria “serviço terminológico” (FREITAS *et al.*, 2009). Para maior clareza, ao longo do presente trabalho utilizam-se os termos e respectivas definições descritas abaixo, baseadas em (SMITH *et al.*, 2006) e (STENZHORN *et al.*, 2008).

Terminologia é um artefato de representação constituído por unidades de representação, as quais são os termos gerais de alguma linguagem natural usado para se referir a entidades em algum domínio específico. Terminologias são coleções de termos, são construídas para reunir termos e relacionam os diversos significados das entidades linguísticas entre si. Terminologias biomédicas não usam descrições formais, mas sim definem os termos através de expressões em linguagem natural, estabelecendo as associações entre os termos por relações de linguagem informal (por exemplo, "*Tratamento tem Sentido Mais Abrangente Que Cirurgia*"). Os termos são normalmente organizados em

hierarquias que se relacionam com o seu significado em termos de sinonímia (mesmo significado), hiperonímia (sentido mais amplo), hiponímia (sentido mais estrito), etc. Termos sinônimos são comumente referidos como conceitos.

Taxonomia é um artefato representacional apresentado via grafos com formato de árvore e contendo nós que representam universais ou classes. Além disso, contém arestas que representam relações hierárquicas (do tipo *é-um*) ou relações de subconjunto. Pode ser orientada à termos ou orientada à objetos da realidade. Será tratada aqui como um tipo de terminologia.

Ontologia é um artefato de representação, o qual contém uma taxonomia como estrutura central. As unidades representacionais designam uma combinação de universais¹⁷, classes definidas e relações definidas entre os universais e classes. Ontologias devem ter uma relação direta com a realidade sob alguma perspectiva definida e compartilhada. Isso exclui afirmações probabilísticas, contextuais e epistemológicas. Ontologias são expressas em formalismos baseados em lógica, de modo que motores de inferência e algoritmos baseados em lógica possam manipular as definições de acordo com a estrutura da ontologia. Portanto, algum tipo de inferência automatizada é necessário em nossa definição de ontologias.

Ontologia realista é uma ontologia com referência direta a realidade, utilizando como componentes os “universais” (eventualmente denominado de categorias ou tipos). Os universais representam o conhecimento expresso em teorias científicas formais, refletindo a melhor aproximação da verdade que dispomos em um determinado momento sobre determinado assunto. Assim, ontologias realistas não representam termos ou conceitos e são independentes da linguagem. Uma vez que não é possível distinguir as ontologias realistas de outras ontologias sem uma auditoria detalhada, o termo “ontologia realista” será reservado para ontologias desenvolvidas de acordo com a metodologia do realismo ontológico.

Sistemas terminológicos incluem ontologias e terminologias. Às vezes, será difícil determinar se um sistema terminológico mencionado na literatura é ou não uma ontologia,

¹⁷ Universais representam aquilo que as entidades individuais correspondentes (ex. todas as baleias) possuem em comum, sendo invariantes na realidade. Sua existência não é consensual. Este assunto é abordado com mais detalhes na seção 3.2.

devido à heterogeneidade das definições e metodologias. Uma vez que não se propõe auditoria de sistemas terminológicos neste trabalho, faz-se uso da descrição oferecida pelos autores do artefato sob análise.

3.1.2 Objetivos das ontologias

Após os primeiros usos como estrutura para organização da informação, as ontologias foram adotadas por várias comunidades de diferentes disciplinas. Atualmente, os objetivos mais comuns do uso de ontologias em sistemas de informação são (GRUNINGER *et al.*, 2008):

- Compartilhar bases de conhecimento;
- Permitir comunicação entre agentes;
- Integrar conjuntos de dados heterogêneos;
- Apoiar suporte à decisão;
- Proporcionar arcabouços semânticos para sistemas de informação organizacionais;
- Representar vocabulários em linguagem natural;
- Representar a semântica de serviços e aplicações de *softwares*;
- Apoiar a busca por conhecimento, proporcionando aprimoramento e melhorias.

No domínio da saúde, todos estes objetivos são importantes, mesmo que exigidos em diferentes graus. Rector aponta quatro funções essenciais para o uso de ontologias em sistemas de informação para a saúde (RECTOR, 2010):

- Especificar as entidades e as relações sobre as quais há informação a ser transmitida;
- Organizar as entidades e as descrições em uma rede de relações é-um (taxonômicas);
- Representar as características universais destas entidades e relações;
- Especificar o que pode e não pode ser dito sobre essas entidades e suas relações

Na próxima seção discutem-se três aspectos importantes das ontologias na Medicina: a evolução natural das terminologias em ontologias, o uso intensivo de representações de conhecimento médico em linguagens lógicas para inferência automática e a Web Semântica, bem como seus desdobramentos para a prática em saúde.

3.1.3 Sistemas de classificação e sua evolução em ontologias

Conforme mencionado, ontologias são tipos especiais (e mais complexos) de sistemas terminológicos. Os profissionais de saúde têm uma longa tradição de criar sistemas de classificação. Encontrados, por exemplo, em registros históricos de John Graunt¹⁸ e em registros de óbito de Londres, sistemas de classificação foram utilizados pela primeira vez como uma diretriz para registrar informações relevantes para a saúde pública, para fins epidemiológicos e de planejamento de longo prazo. Segundo a Organização Mundial de Saúde (OMS)¹⁹, as classificações médicas capturam retratos instantâneos da saúde da população através de parâmetros como mortalidade, doenças, funcionalidade para atividades diárias, incapacidade, saúde e intervenções em saúde. Eles também fornecem *insights* sobre tendências que orientam o planejamento e tomada de decisão pelas autoridades de saúde público ou privado. O número de objetivos para os sistemas de classificação é amplo, e conseqüentemente, existem diversos sistemas disponíveis para uso.

Um dos primeiros sistemas de classificação na Medicina foi a Lista Internacional de Causas de Morte, mais tarde renomeada Classificação Internacional de Doenças (CID). O CID tem como objetivo classificar as doenças registradas em documentos governamentais e hospitalares, inclusive certidões de óbito e prontuários médicos, e é obrigatória em alguns países (incluindo Brasil). Atualmente, é amplamente adotada por hospitais ao redor do

¹⁸ John Graunt (1620-1674), inglês, foi dos primeiros epidemiologistas a realizar análises sobre registros de óbito (*Bills of Mortality*) e é o autor de um trabalho seminal na saúde pública, o *Natural and Political Observations Made upon the Bills of Mortality*.

¹⁹ <http://www.who.int/topics/classification/en/>

mundo em sistemas de reembolso (operadoras de seguro-saúde), registro e exploração de dados do prontuário, com fins epidemiológicos e para avaliação da qualidade dos serviços. Devido à sua utilização em sistemas de reembolso, o CID foi modificado em muitos países para incluir, além das doenças, várias intervenções clínicas e cirúrgicas (JETTE *et al.*, 2010).

CID é apenas um dos muitos sistemas de classificação oficial publicados e mantidos pela OMS, que também publica duas outras classificações de referência. A Classificação Internacional de Funcionalidade, Incapacidade e Saúde (CIF) fornece uma lista de funções e estruturas corporais, além de uma lista de domínios de atividade e participação, tanto no nível individual quanto populacional. Uma vez que as deficiências de saúde de um indivíduo ocorrem em contextos pré-definidos, o CIF também inclui uma lista de fatores ambientais. A Classificação Internacional de Intervenções em Saúde (CIIS) visa a representar uma lista abrangente de ações com fins curativos e preventivos executados por serviços médicos, cirúrgicos e outros relacionados a serviços de saúde.

A proliferação do uso de terminologias médicas em sistemas de informação tem resultado em mudanças qualitativas nos requisitos para o desenvolvimento de sistemas de classificação. Embora ainda sejam importantes para criação de relatórios epidemiológicos, as terminologias foram incorporadas ao núcleo de sistemas clínicos, sendo utilizadas para a relatórios de cobrança, sistemas de prescrição de medicação e exames complementares, inteligência empresarial (*business intelligence*) do hospital e até mesmo em sistemas de apoio à decisão. Para incorporar essas terminologias, os desenvolvedores de sistemas hospitalares frequentemente desenvolvem terminologias locais para uso interno – na verdade, as terminologias locais ainda são quase onipresentes nos sistemas de saúde hoje em dia – de acordo com os requisitos específicos de cada realidade hospitalar. Apesar de prática, esta abordagem cria grandes dificuldades para a interoperabilidade entre sistemas, uma vez que é a regra a co-existência de vários sistemas de informação nas instituições de saúde. Surgiu, portanto, uma necessidade de padronização da tecnologia da informação (TI) na saúde, incluindo a criação vocabulários compartilhados para a troca de informações

Esforços distribuídos para criação de terminologias foram feitos por várias organizações, como o *Systematized Nomenclature of Medicine* - SNOMED (desenvolvida no *American College of Pathologists*), o *Medical Subject Headings* - MESH (na *National Library of Medicine*) e *Logical Observation Identifiers Names and Codes* - LOINC (no Instituto *Regestrief*). Diferenças significativas entre as várias terminologias levaram ao desenvolvimento de alguns critérios para a avaliação das mesmas. Campbell e outros, por exemplo, citam completude e abrangência, clareza, mapeamento a outras terminologias

administrativas, o caráter de composição, a incorporação de sinônimos, atributos de termos, a atribuição de incerteza para as declarações, as hierarquias e herança, identificadores independentes de contexto, identificadores únicos, definições, independência de linguagem, sintaxe e gramática (CAMPBELL *et al.*, 1997).

Para resolver questões dessa natureza, Cimino (1998) publicou um *Desiderata* para vocabulários médicos, listando características desejáveis que possam orientar o desenvolvimento de tais artefatos. Essa lista contém:

- Conteúdo do vocabulário – a terminologia deve crescer constantemente para atender a demanda do usuário, conforme o uso e o conhecimento da área médica se expande;
- Orientação a conceito – significa que os termos devem corresponder a pelo menos um significado, não mais de um significado e que significados não podem corresponder a mais de um termo. Essa questão era tida como consensual na época, mas tornou-se o foco de discussão significativa (abordada adiante).
- Persistência do conceito – o significado de um conceito deve ser mantido constante durante as novas revisões subsequentes da terminologia;
- Identificadores conceituais não-semânticos – deve haver identificadores para cada conceito, para facilitar o processamento por computador e a organização da terminologia após as revisões;
- Poli-hierarquia – cujo objetivo é tornar a busca mais fácil para o usuário, uma vez que a presença de múltiplas hierarquias liga conceitos-pai e conceitos-filho através de relações definidas informalmente;
- Definições formais – normalmente estas definições são expressas como uma coleção de relações entre conceitos na terminologia.
- Rejeição de termos "não especificados" - todo sistema deve representar as informações da mesma forma, na medida do possível. Na busca por essa uniformidade, dois termos não podem se sobrepôr, ou seja, o mesmo evento não pode ser descrito de duas maneiras diferentes. Assim, o uso de modificadores como "outros" (presente, por exemplo, na CID) ou "não classificado" não são desejáveis pois não há como assegurar a interpretação correta. Por exemplo, o termo do CID10 D72.8 – *Outros transtornos especificados dos glóbulos brancos* – só pode ser interpretado uma vez que se está informado sobre todos os transtornos dos glóbulos brancos listadas

em todo o CID10. O termo exige, além disso, a interpretação da palavra “transtorno”, o que pode induzir à erros no registro do diagnóstico. Outro exemplo são os itens D75 – *Outras doenças do sangue e dos órgãos hematopoéticos* – e D76 – *Algumas doenças que envolvem o tecido linforreticular e o sistema reticulohistiocítico* – cuja diferenciação não pode ser feita pela simples descrição dos termos;

- Granularidade múltipla – vocabulários médicos devem ser capazes de lidar de maneira consistente com conceitos de granularidade baixa, como "molécula de insulina" assim como conceitos genéricos, de granularidade alta, como "resistência insulínica";
- Múltiplas visualizações consistentes – propósitos diferentes podem exigir apresentação diferente da terminologia. Pode ser necessário buscar um conceito através de uma representação gráfica em árvore, ou através da busca literal do termo;
- Representação do contexto – vocabulários devem conter a representação de contexto através de informação explícita e formal sobre como os conceitos são usados;
- Evolução “suave” – a evolução constante da terminologia deve ser mantida usando as boas práticas no desenvolvimento de terminologias, privilegiando motivos razoáveis para realização de alterações (adição simples, refinamento, etc.) e evitando outros (redundância, a reutilização de código, etc);
- Reconhecimento da redundância – a possibilidade de declarar a mesma informação de duas maneiras diferentes deve ser evitada;

Apesar das diretrizes, a maioria dos sistemas de classificação em saúde não segue os preceitos da desiderata. Tais sistemas são estruturas genéricas e sua capacidade de combinar categorias ou expressar significados complexos é limitada. O CID, por exemplo, adota uma visão etiológica e fisiopatológica de Medicina e, portanto, classifica as doenças de acordo com o sistema humano envolvido. Os termos não são formalmente descritos, os relacionamentos não são explícitos e o conteúdo é limitado. Os sistemas enumerativos não são escaláveis, e a explosão combinatória que resulta da criação de novos termos é insustentável. Rector (2010) aponta que o número de códigos para “acidente de bicicleta” na terminologia *Read CID* passou de oito (em 1980) a oitenta (no início de 1990), e para 587 no CID modificado pela Austrália (de 1999).

As novas exigências impostas aos sistemas terminológicos fizeram de sua modelagem e avaliação tarefas muito difíceis. Um requisito importante para terminologias é que o conhecimento representado deve ser de boa qualidade, mesmo que conhecimento de “qualidade” seja uma definição complexa. Consideram, entretanto, especialmente a consistência interna e fidelidade à realidade. Consistência interna significa garantir que: termos não sejam hierarquicamente superiores a eles mesmos; termos sejam redundantes; e múltiplas formas de visualização da ontologia representem adequadamente as hierarquias. Uma representação formal é capaz de fornecer semântica explícita para o conhecimento representado, portanto, possibilitando a determinação da consistência interna e fidelidade aos fatos. Uma das vantagens dos sistemas terminológicos baseados em lógica descritiva - uma família de lógicas amplamente utilizada desde os anos 1990 - é a possibilidade de auditoria ao se buscar por definições duplicadas, omissão de propriedades e por sub-especificação.

Em resumo, os sistemas de classificação têm sido postos a prova em tarefas para as quais não foram projetados originalmente. Para fazer frente às novas necessidades, devem ser reescritos de forma clara, concisa, abrangente e formal, suportando algum nível de inferência automatizada. Além disso, mais importante, devem ser capazes de definir claramente o significado dos termos. Ontologias são artefatos terminológicos particularmente bem adaptados para atender a essas demandas, devido à sua ênfase em tornar explícitos os compromissos ontológicos e a capacidade de fixar significado através de relacionamentos entre termos e referência de universais do mundo real.

3.1.4 Representação simbólica de conhecimento e Lógica Descritiva

Apesar de amplamente discutidas e teorizadas em filosofia, as ontologias têm se tornado muito mais populares nas últimas décadas devido ao seu uso como um artefato de ciência da computação na IA, e na pesquisa e prática de sistemas de informação. Como Smith e Welty apontam, a IA logicista se focou desde o início em sistemas que pudessem manipular ou simular conhecimento para criação de mecanismos de inferência automatizada.

Enquanto alguns problemas interessantes foram solucionados por algoritmos de linguagem procedural²⁰, problemas de processamento verbal simples continuavam longe de uma solução eficaz. Dentre os princípios que John McCarthy considera essenciais para a criação de um sistema com inteligência de ordem humana, cabe citar sua afirmação de que "todos os comportamentos devem ser representáveis no sistema" (MCCARTHY, 1959), e a proposição de um programa que representasse conhecimento de senso comum através de sentenças em uma linguagem lógica apropriada.

Embora a ênfase claramente tenha recaído sobre os algoritmos de inferência, as bases de conhecimento também foram objeto de estudo intenso nos anos subsequentes. Iniciativas para melhorias no rigor no desenvolvimento destas bases levou à criação da área de engenharia de conhecimento. Já em 1979, havia evidências da relevância do processamento simbólico de apoio à decisão médica (SHORTLIFFE *et al.*, 1979), como forma de fazer os computadores "compreenderem" problemas e utilizarem conhecimento de senso comum para resolvê-los. Conforme sistemas e aplicações se disseminaram, os desenvolvedores de sistemas baseados em conhecimento tiveram de lidar com bancos de dados heterogêneos, linguagens de programação variadas e protocolos de rede não padronizados. Este desenvolvimento exigiu a criação de convenções em três níveis: formato da linguagem de representação, o protocolo de comunicação entre agentes e formato de especificação (GRUBER, 1995).

Independentemente da comunidade de representação do conhecimento, duas outras correntes da ciência da computação também fizeram incursões significativas no desafio de modelagem. No projeto de banco de dados (BD), modeladores estavam constantemente tentando melhorar a escalabilidade e interoperabilidade, a fim de lidar com o grande avanço comercial de BD. Em outra frente, linguagens orientadas à objeto forçaram os programadores a usar *frameworks* de modelagem para elaboração de projetos de grande escala. Essas três vertentes juntas encontraram uma convergência na atividade de fazer modelos do mundo. Para a representação de conhecimento sobre o mundo, representam-se

²⁰ Adota-se aqui uma distinção entre linguagem procedural (ou imperativa) e declarativa. Na linguagem procedural o programador define um conjunto de ações que, quando completadas, fornecem o resultado proposto. São exemplos as linguagens C, C++ e Pascal. Na linguagem declarativa, o programador define uma série de parâmetros ou condições que devem ser cumpridos, sem definir como a computação é feita.

as propriedades estáticas do mundo ou o comportamento e as ações dos objetos (SMITH; WELTY, 2001).

A herança da comunidade de representação de conhecimento trouxe consigo a ênfase em manipulação simbólica e nos algoritmos baseados em lógica, inspirados nos primeiros pesquisadores em IA (*Good Old Fashioned Artificial Intelligence*). Uma vez que ontologias dizem respeito à existência, ou ao menos a um modelo do estado de coisas do mundo, não existe *a priori* uma definição sobre qual é a melhor representação. Após alguns experimentos com graus variáveis de sucesso, o desenvolvimento de linguagens em lógica descritiva se tornou preponderante. A preferência pela representação em lógica descritiva deveu-se especialmente a questões de implementação, uma vez que esta lógica possui características que garantem decidibilidade das inferências, ou seja, uma resposta correta é sempre garantida.

Nos próximos parágrafos descreveremos princípios de lógica descritiva utilizando notação lógica. Devido à miríade de notações possíveis, decidimos por utilizar dois conjuntos de símbolos: *Manchester syntax* (HORRIDGE *et al.*, 2006) para lógica descritiva e símbolos especiais para expressões em lógica de primeira ordem (LPO).

Lógicas descritiva (LD) são fragmentos da LPO que permitem a definição formal e explícita de termos em uma ontologia. A adoção da lógica descritiva permite a representação de conhecimento e a inferência automatizada, principalmente para verificar a consistência de classificação nas relações é-um. A representação da lógica descritiva pode ser dividida em duas partes: o T-Box e o A-Box. O T-Box representa o conhecimento geral, ou intensional (BAADER *et al.*, 2003), e a declaração mais comum é a definição de conceitos, ou seja, a definição de novos conceitos através de conceitos previamente estabelecidos. Por exemplo, uma mulher é definida como:

$$\text{Mulher} \equiv \text{Pessoa} \cap \text{Fêmea}.$$

Neste caso, usamos o operador AND (\cap) para dizer que para todo X, se X é uma Pessoa e X é Fêmea, então X é uma Mulher. Em outras palavras, ser uma Mulher significa ser IGUALMENTE uma Pessoa e Fêmea.

$$\text{Pessoa} \equiv \text{Homem} \cup \text{Mulher}$$

O operador OR (\cup) é usado para afirmar que, para todo X, se X é ou um Homem ou uma Mulher, então X é uma Pessoa. Em outras palavras, ser uma Pessoa é ser OU um

Homem ou uma Mulher. Adicionalmente, podemos usar o operador NÃO (\neg) para representar disjunção, ex. Mulher é uma Pessoa que NÃO é um Homem.

As duas afirmativas anteriores são chamadas aqui de afirmativas sobre classe, ou afirmativas T-Box, uma vez que versam sobre as definições da classe. Classes podem ser equivalentes a outras classes (símbolo \equiv ou **EquivalentTo** em Manchester Syntax) ou podem ser subclasses (símbolo \subseteq ou **SubClassOf** em Manchester Syntax). Afirmativas em T-Box podem ser quantificadas, ou seja, podem se referir a pelo menos um membro ou a todos os membros de uma classe. Para estas afirmativas usaremos operadores adicionais, denominados ForAll (ONLY em Manchester Syntax e \forall em notação de LPO) e ForSome (SOME em Manchester Syntax e \exists em LPO). A quantificação permite as seguintes expressões:

Pessoa subClassOf temParte SOME Cérebro – que significa que todo membro da classe Pessoa também é membro da classes de entidades que possui pelo menos um Cérebro como parte.

Pessoa subClassOf temPai ONLY Pessoa – que significa que todo membro da classe Pessoa também é membro da classe de entidades que possui apenas como pai uma Pessoa. Porém, note que um membro que não tem pai ainda pode ser uma Pessoa.

Os conceitos do T-Box devem ser não circulares, ou seja, os conceitos não podem ser definidos por eles próprios (mulher \equiv pessoa \cap fêmea) e apenas uma definição para um conceito é permitida.

Já o A-Box contém o conhecimento extensional sobre os indivíduos, atribuindo a eles o conceito a que pertencem ou o papel que desempenham. Por exemplo, a afirmação que Maria é uma pessoa do sexo feminino é feita por:

Pessoa \cap Fêmea (MARIA).

Infere-se, portanto, que Maria é instância do conceito “Mulher”, como definido acima. Já a relação de papel que Maria tem como filho o José é exemplificada por:

temFilho(MARIA, JOSÉ)

Lógica descritiva é o arcabouço mais comum para expressar axiomas ontológicos, representando a busca por expressividade e viabilidade computacional. As linguagens

derivadas da LD diferem em expressividade: enquanto linguagens pouco expressivas levam a subespecificação (o que implica na interpretação inadequada de modelos da ontologia), linguagens muito expressivas levam a processos de inferência computacionalmente caros e inviáveis do ponto de vista prático.

Atualmente, estas linguagens são tomadas praticamente como sinônimos de ontologia. A mais comum dentre elas é a *Web Ontology Language* (OWL-DL), que é fruto de um trabalho conjunto resultando em um compromisso, conforme citado acima, entre expressividade e decidibilidade. Além disso, a OWL tem amplo suporte de ferramentas para classificação, como o RACER²¹, Fact++ (TSARKOV; HORROCKS, 2006) e Pellet (SIRIN *et al.*, 2007).

3.1.4.1 Raciocínio em Lógica Descritiva

Por ser um subconjunto de Lógica de Primeira Ordem, a inferência em Lógica Descritiva segue um padrão similar de inferência indutiva. Apesar da discussão sobre padrões de inferência fugir do escopo do presente trabalho, faremos um resumo breve para leitores sem formação em lógica. Uma definição de classe é baseada na definição de condições necessárias e condições necessárias e suficientes. Por exemplo, considere:

Pessoa subClassOf temParte SOME Cérebro

Significa que se X é uma pessoa, então X possui pelo menos um Cérebro. É uma condição necessária para todo membro da classe Pessoa também possuir um Cérebro. Porém, nem toda entidade com Cérebro é uma Pessoa, e portanto esta condição não é suficiente – em outras palavras, a definição de ser uma Pessoa não significa somente ter um cérebro. Essa condição é válida para todas as subclasses (ex. Todo Adulto precisa de ter pelo menos um cérebro, uma vez que Adulto é um tipo de Pessoa). Já neste outro exemplo:

Pai EquivalentTo Homem AND temFilho SOME Pessoa

Isto significa que todo Homem que tem como filho pelo menos uma Pessoa é um Pai. A condição é necessária e suficiente – ser um Pai significa ser um Homem com um Filho. Existem ganhos adicionais de poder de inferência ao usar condições necessárias e

²¹ <http://www.sts.tu-harburg.de/~r.f.moeller/racer/>

suficientes, já que qualquer classe que cumpre as condições pode ser inferida como uma subclasse. As condições também propagam para as subclasses.

A codificação da ontologia em uma linguagem lógica não é suficiente para inferências adequadas. De acordo com Schulz e Hahn, as inferências corretas em ontologias biomédicas por classificadores lógicos dependem de passos metodológicos no desenvolvimento das ontologias, os quais são denominados de desiderata para inferência (SCHULZ; HAHN, 2005). Por um lado, erros de inferência são causados pela desconsideração, por parte dos autores da ontologia, da pressuposição do mundo aberto (Open World Assumption²²) (BAADER *et al.*, 2003). Neste caso, todos os modelos são aceitos a menos que sejam explicitamente falsificados, o que exige um grande esforço de modelagem. Por exemplo, ao afirmar em uma ontologia que a Cachalote é uma baleia, não é subentendido que a Cachalote não seja um peixe, a menos que seja afirmado que baleias não são qualquer outro animal que não seja baleia. Este problema foi exemplificado durante o alinhamento entre a rede semântica da *Unified Medical Language System* (UMLS) – originada da união de várias terminologias e, portanto, contendo termos ambíguos ou vagos – e o BioTop – criada desde o início como uma ontologia, com ligação para a DOLCE, BFO e para a ontologia de relações da OBO. Schulz e colaboradores reportam baixa concordância entre o inferência automática e a avaliação de um especialista ao definir relações entre os conceitos (SCHULZ *et al.*, 2009a). Ou seja, apesar de todos os esforços na modelagem, a capacidade da ontologia em detectar modelos incorretos foi significativamente menor do que a capacidade de um especialista em encontrar erros de modelagem.

Outro problema comum encontrado na biomedicina são variações das entidades descritas no domínio de saúde. Por exemplo, um infarto do miocárdio normalmente é acompanhado por uma dor intensa no peito, mas existem casos em que a dor se manifesta em outros locais, e mesmo casos sem qualquer dor. As variações são comuns, devido a doenças, procedimentos externos (medicamentos ou cirurgias) ou variações sem causa definida. Inferências lógicas exigem que as afirmativas sejam precisas em sua descrição e,

²² A Open World Assumption, ou pressuposto de mundo aberto, declara que todas combinações de mundos são possíveis, a menos que explicitamente dito em contrário. Por exemplo, a afirmação que a baleia é um mamífero não torna inválida a afirmação que a baleia é um peixe, e menos que exista um outro axioma que diga que qualquer membro da classe mamífero não é membro da classe peixe.

portanto, relações devem ser descritas de acordo com sua validade (mandatório, opcional e inválido). Outras características desejáveis que devem ser representadas em lógica são a capacidade de negação, propriedades formais das relações (transitividade, reflexividade, simetria), e formalização de relações topológicas (*parte_de* e *localizado_em*, as quais são apresentadas na sequência).

3.1.5 Ontologias e a Web Semântica

Desde seu surgimento no final da década de 1980, a *World Wide Web* (WWW) cresceu enormemente, com a criação de bilhões de páginas, além de mecanismos de busca de documentos capazes de localizar em segundos os *links* úteis para busca. Porém, a WWW atual consiste primariamente em documentos legíveis exclusivamente por seres humanos. Em um artigo de 2001, Berners-Lee e outros propuseram uma nova visão da Web, na qual não apenas documentos, mas dados, são ligados em rede (BERNERS-LEE *et al.*, 2001). A Web Semântica é uma rede de informações tratáveis por máquinas, através do uso de teorias semânticas para manipulação dos símbolos (SHADBOLT *et al.*, 2006).

Para a realização da Web Semântica, os dados devem estar disponíveis de forma tratável, ou seja, os símbolos de representação devem ser manipuláveis por algoritmos de computador. Portanto, a motivação inicial para criação de ontologias, que sistemas baseados em conhecimento são normalmente fundamentados em "plataformas de hardware, linguagens de programação e protocolos de rede heterogêneos" (GRUBER, 1995, p.907), torna-se ainda mais problemática. Como Uschold e Jasper (JASPER; USCHOLD, 1999) afirmam: "fundamentalmente, ontologias são utilizadas para melhorar a comunicação entre humanos ou computadores". As ontologias se configuram, nesse contexto, o artefato perfeito para possibilitar a troca de dados na Web.

A definição precisa do significado do termo fez das ontologias parte integrante da Web Semântica. Bancos de Dados relacionais e seus modelos de dados complexos e proprietários não podem ser usados para publicar dados de forma automaticamente tratável. Ontologias, porém, podem ser usadas para anotar dados e expressar afirmativas simples compartilháveis, desde que as ontologias fossem elas mesmas, também compartilhadas. O sucesso dessa empreitada depende da disseminação e aceitação dos termos e conceitos propostos para uso comum, independente da complexidade dos modelos. De fato, a ênfase

na troca de informações tem promovido ontologias cada vez mais simples, como a FOAF (*Friend of a Friend*²³), que descreve termos como “nome”, “possui *Website*”, etc, utilizando a linguagem natural como único balizador para construção das terminologias. Cabe aqui lembrar, que o termo ontologia tem sido usado de forma diversa, conforme mencionado na seção 3.1.1.

A criação de ambiguidades é inevitável quando se usam termos de linguagem para representar conceitos. Ontologistas entendem a importância de indicar precisamente o que os termos querem dizer, usando uma terminologia específica, e enfatizam a necessidade de restringir a interpretação, ou seja, a necessidade de considerar a semântica. Uma ontologia comum define o vocabulário corrente para que as informações (consultas e afirmações) possam ser trocados entre os agentes envolvidos, sejam pessoas ou máquinas. Neste contexto, o compromisso ontológico busca uniformizar a interpretação dos termos. Em outras palavras, usando um conjunto de termos e formalizando-se as suas relações, o projetista da ontologia faz um compromisso com uma visão da realidade de forma a evitar interpretações equivocadas. Assim, o papel do compromisso ontológico é justamente determinar qual é o significado adotado para o termo (GÓMEZ-PÉREZ *et al.*, 2004).

Esta seção explorou a transformação das terminologias médicas através da utilização de princípios ontológicos, a importância das relações formais e da lógica descritiva para inferências, bem como a visão da Web Semântica, que pretende usar terminologias formais para trocar informação na Web. De fato, ontologias têm grande potencial para a representação e troca de informações na saúde (BODENREIDER; STEVENS, 2006; CIMINO; ZHU, 2006; SICILIA *et al.*, 2009; CEUSTERS *et al.*, 2011). Ainda assim, acredita-se na existência de limitações e restrições para que tal representação seja adequada à prática médica. Tais questões são exploradas nas seções seguintes.

²³ <http://www.foaf-project.org/>

3.2 O PAPEL DAS ONTOLOGIAS FORMAIS – ABORDAGEM CRÍTICA DO REALISMO COMO TÉCNICA DE MODELAGEM

Como visto até aqui, ontologias nasceram em um ambiente turbulento e multifacetado. Na comunidade de representação de conhecimento, a representação do senso comum é um motivador importante para a criação de ontologias. Além disso, a necessidade de reuso e uso para comunicação de agentes e integração de bases de dados requerem grandes esforços de modelagem. Como demonstrado durante a criação do CYC (LENAT, 1995) e outras bases de conhecimento nos anos 90, há a necessidade de modelar independentemente da tarefa pretendida (VICKERY, 1997). No entanto, a popularização das ontologias criou um número enorme de artefatos, com diferentes linguagens e técnicas de modelagem. Ao mesmo tempo em que prometiam facilitar a integração de bases de dados heterogêneas através da definição de termos comuns, o desenvolvimento de ontologias livre e sem orientação acabou resultando em uma Torre de Babel (SMITH; CEUSTERS, 2003). O aporte de teorias filosóficas para a engenharia ontológica pretende diminuir este problema, ao fazer distinções estritas sobre o que pode e o que deve ser representado, e por que.

Nesta seção, descreve-se o conceito de ontologia formal e a influência do estudo filosófico na criação de artefatos ontológicos. Em particular, detalha-se a metodologia do realismo ontológico, suas bases filosóficas, vantagens e limitações.

3.2.1 Filosofia, Ciência da Informação e Ciência da Computação

Entre os primeiros autores a reconhecer a interseção entre a filosofia e IA estão McCarthy e Hayes (MCCARTHY; HAYES, 1969), curiosamente os primeiros a reconhecer a importância das ontologias na ciência da computação. Porém, o uso de abordagens filosóficas explicitamente orientadas ao desenvolvimento de ontologias só aconteceu na década de 90, com a introdução das ontologias formais (GUARINO, 1995; GUARINO, 1998) e a experiência de grandes projetos utilizando ontologias para representação de conhecimento, como o projeto CYC. Smith aponta que “para criar representações efetivas é

uma vantagem saber algo sobre as coisas e os processos que se pretende representar²⁴ (SMITH, 2003b, p.16). Nessa afirmação reside o objetivo do presente estudo ao abordar a disciplina filosófica da Ontologia Aplicada, que descende dos estudos em Ontologia Formal. Cabe porém citar que desenvolvimento não foi linear e nem consensual em todas as comunidades ontológicas, justificado pela falta de consenso sobre as doutrinas entre os próprios filósofos. De fato, tal falta de consenso faz parte da própria natureza do estudo filosófico.

A filosofia tem um longo histórico de estudiosos à busca do “que existe”, como por exemplo: Aristóteles, Kant, Husserl, Brentano, Wittgenstein, Peirce e vários outros. Não é possível, assim, distinguir uma única corrente uniforme. Na FIG. 2 podemos distinguir alguns autores dos séculos XIX, XX e XXI segundo suas correntes de estudo ontológico.

²⁴ to create effective representations it is an advantage if one knows something about the things and processes one is trying to represent

- Ontologia formalizada: relativa à codificação propriamente dita das outras configurações das ontologias.

O uso da ontologia como artefato para processamento requer a precisão na definição das categorias e suas relações, ou seja, que a ontologia formal seja consistente, para que sua formalização seja consistente. Para tal, é necessário definir a natureza das categorias, ou universais (caso se acredite na sua existência). Dentre as diversas possibilidades, destacam-se três interpretações possíveis: nominalismo, conceitualismo e realismo. A descrição é um tanto simples e serve apenas aos propósitos dessa tese, não representando a extensão da discussão nem sua profundidade.

O nominalismo argumenta que os universais, caso existam, são expressões linguísticas, ou seja, “além das expressões predicadas que ocorrem ou podem ocorrer na linguagem, não existem outros universais” (COCCHIARELLA, 1989, apud POLI, 2010). O nominalismo é extensionalista, uma vez que não admite entidades subjacentes que relacionem dois objetos. Por exemplo, o universal “lápiz” é apenas uma nome que representa todos os lápis existentes.

O conceitualismo argumenta que um universal resulta da capacidade cognitiva de categorizar nossas experiências. O conceitualismo nega a existência de tipos ou universais independentes do pensamento. O conceito é relacionado a uma terminologia, sendo o significado ou sentido de um termo. Outra interpretação é aquela de que conceitos são relações abstratas (regras ou funções) que permitem a interpretação adequada do termo de acordo com o contexto (MERRILL, 2009). Nesta definição, o termo “o rei da França” é uma regra que permite identificar o particular “rei da França” adequado para compreensão da sentença.

O realismo argumenta que os universais são entidades independentes dos objetos individuais. Para o realista, os universais não dependem da minha percepção e cognição para sua existência – mesmo que nunca houvesse um ser humano, uma banana continuaria a ser uma fruta, assim como uma maçã. O realismo se baseia no conhecimento da verdade sobre as coisas, sendo a verdade fundamentada em fatos, independentemente do uso linguístico de uma expressão ou mesmo de um modelo mental sobre a entidade. Uma vez que a realidade não pode ser completamente conhecida pelas pessoas comuns, o realista olha para a ciência em busca de respostas sobre a verdade.

Não está no escopo deste trabalho a discussão sobre correntes filosóficas em relação a sua consistência e mérito, mas apenas discutir as contribuições para a criação de ontologias como artefatos para sistemas de informação. Na próxima seção, exploram-se os desdobramentos da corrente realista na engenharia ontológica, justificando o motivo de sua escolha como eixo de modelagem para ontologias de representação de informações em biomedicina.

3.2.2 Realismo ontológico

“Realismo” é um termo abrangente e controverso. O realismo ontológico é uma das metodologias mais utilizadas atualmente no desenvolvimento de ontologias biomédicas (SMITH *et al.*, 2007; SCHULZ *et al.*, 2009b; SMITH; BROCHHAUSEN, 2010; SMITH; CEUSTERS, 2010), tendo sido desenvolvido e aperfeiçoado no decorrer da última década. O realismo ontológico é baseado na idéia de que a forma mais efetiva de garantir a consistência mútua das ontologias ao longo do tempo é garantir que a sua manutenção acompanhe os avanços na pesquisa empírica. A metodologia propõe um método formal para garantir que os desenvolvedores de ontologias tenham consistência na modelagem, ao adotarem o compromisso ontológico com a realidade como descrita pela ciência. Seus autores descrevem a motivação para a criação da metodologia da seguinte forma:

"Infelizmente, o sucesso (de estratégias de integração de dados) tem levado à criação constante de novas ontologias, e assim fez renascer os problemas de silo que as ontologias deveriam combater. Para este fim, é evidentemente vantajoso se pudermos encontrar uma maneira de minimizar o número de ontologias que estão sendo construídas e ao mesmo tempo maximizar a sua coerência mútua. Esses objetivos só podem ser alcançados, no entanto, se conseguirmos convencer os desenvolvedores de ontologias para aceitar certas restrições comuns sobre como eles constroem suas ontologias e se pudermos encontrar uma maneira de fazer isso sem comprometer a flexibilidade que é necessária para acompanhar os avanços científicos".(SMITH; CEUSTERS, 2010, p.139)

Filosoficamente, os autores²⁵ se baseiam em princípios de outros autores que advogam princípios realistas (mesmo que a abrangência do termo não seja identificada), como Aristóteles, Ingarden, Chisholm, Johansson e Lowe (SMITH, 2004). As disputas filosóficas entre o realismo e o anti-realismo são seculares. Entretanto, segundo os criadores do que denomina-se realismo ontológico, o uso e o benefício da metodologia não depende da adoção da postura realista: apenas é necessário que seus princípios sejam seguidos durante o desenvolvimento da ontologia (SMITH; CEUSTERS, 2010).

O pilar principal se baseia na teoria de universais (ou tipos) e particulares (ou instâncias). Apesar da variedade de definições, consideram-se aqui que particulares são as coisas que podem ser descritas com base na observação feita no mundo real, como uma clínica ou laboratório; já os universais representam aquilo que os particulares correspondentes possuem em comum, sendo invariantes na realidade (SMITH, 2004; 2006). Como o realismo ontológico se baseia na realidade e acredita que a melhor forma de descrevê-la é através da ciência, os universais são consideradas as entidades da realidade usadas na formulação de teorias científicas.

Faz parte da aproximação com a realidade a classificação dos particulares e universais de acordo com suas similaridades, o que implica em relações hierárquicas entre as entidades. Por exemplo, a entidade baleia (universal) possui características similares ao universal peixe, definidos pelo universal animal; estas características são essenciais para a existência destes universais. Ao mesmo tempo, o habitat destes animais também é uma característica que os aproxima e seria uma possibilidade de classificação, assim como a presença de estruturas funcionais (barbatana e olhos) e a coloração. Uma vez que a criação irrestrita de conjuntos e classes (ex. animais que vivem no mar e possuem cor azul) é um importante motivo para inconsistência das ontologias, o realismo ontológico restringe os conjuntos àqueles definidos pela comunidade científica através das teorias científicas. Enquanto os universais são agrupados em virtude do *que são*, outros conjuntos são agrupados em virtude de *como são* (SMITH; CEUSTERS, 2010).

Uma vez que a distinção precisa entre universais e outros conjuntos é difícil, outra característica da metodologia é o uso de uma ontologia de alto-nível para orientar a

²⁵ os mais ativos são os professores Barry Smith e Werner Ceusters, da *State University of New York at Buffalo*

categorização dos universais. Ontologias de alto nível são ontologias que descrevem categorias gerais, de acordo com alguma metodologia proposta, que orientam a criação de novas classes. Exemplos de ontologias de alto nível são a DOLCE, a BFO, a CYC, dentre outras. Para assegurar consistência na modelagem, a metodologia propõe uma abordagem *top-down* no desenvolvimento de ontologias. No caso da BFO, foram propostas categorias estruturantes, compostas por universais genéricos, especificamente “continuantes independentes”, “continuantes dependentes”, e “ocorrentes”. Esta divisão se baseia nos conceitos de SNAP e SPAN (GRENON *et al.*, 2004). SNAP são universais que têm um início e fim determinado, chamados ocorrentes, e englobam processos (por exemplo, a vida de um organismo, o nascimento) e regiões espaço-temporais (os anos 80, o dia de hoje). SPAN, ou continuantes, são universais que representam as entidades que mantêm sua identidade com o decorrer do tempo (por exemplo, os seres e os objetos). Os continuantes podem ser dependentes (ex. qualidade como as cores, que são inerentes aos objetos), independentes (ex. a mesa) ou regiões espaciais (ex. um ponto). O uso de uma ontologia de alto-nível por duas ontologias diferentes aumenta a possibilidade que seus universais sejam compatíveis e integráveis, visto que os modeladores terão uma visão compatível sobre o que é um evento, um papel, uma relação, etc.

Além de categorias de alto nível, as ontologias de cunho realista possuem relações bem definidas para manter a consistência interna da ontologia. Destacam-se as relações de instanciação e as relações mereológicas:

- Relações de instanciação: A instanciação relaciona um indivíduo à sua classe correspondente. Usando a terminologia realista, conecta o particular ao seu universal correspondente. Ao formalizar universais e particulares, é importante distinguir a teoria realista da teoria dos conjuntos. Isto pode ser feito levando em consideração a dimensão temporal: enquanto os conjuntos são identificados por suas extensões, os universais permanecem constantes mesmo que suas instâncias sofram alterações. Por exemplo, o universal *Homo sapiens* preservou sua identidade através de um longo período de tempo, apesar de praticamente todas as suas instâncias mudarem completamente a cada 100 anos (SCHULZ; HAHN, 2007).
- Relações mereológicas: A relação de “parte-de” é uma descrição comum de entidades do domínio biomédico, que possui um histórico importante de descrições anatômicas e estruturais. O significado preciso das relações

“parte-de” e “tem-parte” não é claro para todo continuante do contexto biológico. Porém, todo continuante particular ocupa uma região espacial única em todo momento no tempo de sua existência e, portanto, a localização espacial é normalmente bem determinada para a maioria das afirmativas anatômicas (SCHULZ *et al.*, 2007).

A relação “localizado_em” pode ser definida como “Entidade a é 'localizado-em' entidade b no tempo t se e somente se a região ocupada por a em t é 'parte-de' a região por b em t” (SCHULZ *et al.*, 2006). Três critérios principais podem ser usados para distinguir as relações. O primeiro é se a parte é uma parte biológica do todo – ou seja, não é um artefato feito pelo homem, como uma válvula cardíaca artificial. Este critério não se aplica a micromoléculas (açúcar, lipídios) e algumas organelas celulares que contêm conteúdo genético específico. O segundo critério é se a parte é sempre localizada no todo durante seu ciclo de vida. O terceiro é se a parte cumpre uma função biológica essencial do todo. Estes critérios não são completamente livres de ambiguidade, e é possível que a propagação de inferências seja mais segura quando as estruturas estão relacionadas por critérios estritamente espaciais (SCHULZ *et al.*, 2007). A relação “parte_de” também é usada para descrever ocorrentes, sempre que compostos de sub-ocorrentes, ao invés de relacionar estes sub-ocorrentes como condições de existência necessárias e suficientes para o todo.

O sucesso da metodologia até aqui descrita pode ser comprovada, especialmente no domínio da biologia, pelo número de ontologias desenvolvidas e utilizadas. A OBO *Foundry* (*Open Biomedical Ontologies*) é o principal repositório destas ontologias, mantendo as ontologias acessíveis, ortogonais (os universais contidos em uma ontologia são referenciados por outra e não são recriados), escritos em sintaxe adequada e identificadas corretamente (SMITH *et al.*, 2007). São particularmente importantes as seguintes ontologias:

- *Gene ontology*: é uma ontologia que descreve as características fundamentais dos genes e seus produtos (HILL *et al.*, 2008). É dividida em 3 partes: componentes celulares; funções moleculares, e processos biológicos. Essa divisão corresponde às categorias primitivas do BFO, continuante independente, continuante dependente e ocorrente. Foi uma das primeiras

ontologias desenvolvidas utilizando os princípios realistas e tem uso disseminado entre pesquisadores.

- *Foundational Model of Anatomy*²⁶: descreve a anatomia do ser humano de forma prototípica (as variações não são contempladas), incluindo as relações mereo-topológicas entre os órgãos. Usada em sistemas para ensino e inferências.
- *Cell Ontology*: descreve os tipos celulares de vários seres vivos, incluindo animais, procariotas, células e fungos (BARD *et al.*, 2005). Tem como objetivo principal a criação de um vocabulário controlado para anotação de fenótipos celulares e expressão genética. Seu uso permite questões como, por exemplo, a listagem de todos os tipos de célula em ratos que expressam um determinado gene..
- *Protein Ontology*²⁷: descreve os relacionamentos entre proteínas e as classes que representam a evolução das proteínas, para representar as formas proteicas de um locus genético e interconectar a outras ontologias. Dessa forma é possível correlacionar as ontologias a outras proteínas análogas evolutivamente e gerar novas hipóteses através da integração de dados e inferências automáticas.
- CHEBI (*Chemical Entities of Biological Interest*)²⁸: descreve entidades moleculares de interesse biológico, focando na descrição de compostos químicos “pequenos”. Estão descritas nesta ontologia, por exemplo, glicose e seus vários análogos e variações.

Além destas, inúmeras outras ontologias estão disponíveis ou sob avaliação, como a *Disease Ontology* (ontologia de doenças humanas), a *Ontology for Clinical Investigations* (descreve ensaios clínicos e outros estudos clínicos), etc.

²⁶ <http://fma.biostr.washington.edu/>

²⁷ <http://pir.georgetown.edu/pro/>

²⁸ <http://www.ebi.ac.uk/chebi>

Apesar de intensamente utilizada por diversos pesquisadores ao redor do mundo, o uso da abordagem realista na engenharia ontológica parece possuir limitações e tem recebido críticas. Na seção 3.2.4, apresentam-se brevemente as críticas mais relevantes e o motivo da escolha do realismo para a presente análise, apontando as limitações das ontologias realistas identificadas na literatura.

3.2.3 Karl Popper e a teoria dos três mundos

Karl Popper (1902-1994) foi um dos maiores filósofos da ciência do século 20. Uma abordagem útil combinando realidade, cognição e representações foi proposta por Popper em sua teoria dos três mundos. Popper propõe uma visão pluralista do universo que reconhece pelo menos três mundos diferentes, mas interativos (POPPER, 1978).

Segundo esta teoria, existe um mundo que consiste em corpos físicos, tais como pedras, plantas e animais, que é chamado de mundo 1. O mundo 1 pode ser dividido em mundo *dos objetos físicos não-vivos* e o mundo das *coisas vivas ou objetos biológicos*. Existe o mundo mental e psicológico, chamado mundo 2, que inclui pensamentos, percepções e observações, ou seja, os processos mentais e psicológicos e experiências subjetivas. No mundo 2, podemos distinguir experiências conscientes de sonhos, ou distinguir a consciência humana da consciência animal. Há também um outro mundo, chamado mundo 3, que inclui todos os produtos da mente humana, como idiomas, teorias científicas, construções matemáticas, sinfonias e esculturas. Enquanto um bloco de mármore pertence ao mundo 1, a criação de um artista de uma escultura usando este bloco é uma manifestação no mundo 3.

A partir de uma perspectiva ontológica, pode-se afirmar que o mundo 2 e mundo 3 são produtos evolutivos do mundo 1, ou seja, eles não podem sequer existir sem o mundo 1. No entanto, o mundo 2 e 3 podem atingir um status independente que lhes permite influenciar causalmente entidades do mundo 1 através de um mecanismo de feedback. Os mundos 1 e 2 estão, portanto, em interação causal e o mundo 3 influencia o mundo 1 através de mundo 2. Do ponto de vista epistemológico, a cognição humana é capaz de obter o conhecimento de mundo 1, bem como dos mundos 2 e 3 dependentes da mente (NIINILUOTO, 1999).

A teoria do três mundos de Popper tem sido aplicado a investigações em ciência da informação em saúde (BAWDEN, 2002). No domínio da saúde, o mundo 1 consiste em entidades, tais como dores, feridas e bactérias, para mencionar apenas alguns exemplos, todas elas definidas no lado do paciente. No mundo 2, podemos encontrar as representações cognitivas do mundo 1, como observações, interpretações e crenças, definidas tanto do lado de pacientes e médicos. O mundo 3 é composto de concretizações de representações cognitivas do mundo 2, por exemplo, artefatos de informação como terminologias, sistemas categóricas e prontuários médicos. Além disso, os diagnósticos nas mentes dos médicos (mundo 2) e prontuários médicos (mundo 3) estão relacionadas a distúrbios e doenças (mundo 1), através da relação de intencionalidade (SCHULZ; KARLSSON, 2011).

Abordagens similares que dividem a realidade em níveis podem ser encontrados na literatura contemporânea. Smith et al (SMITH *et al.*, 2006) definem três níveis de entidades, nomeadamente:

Nível 1 - objetos, processos, qualidades, estados na realidade;

Nível 2 - representações cognitivas dessa realidade nas mentes das pessoas;

Nível 3 - concretizações dessas representações cognitivas em artefatos de representação compartilháveis.

Essa divisão tem sido usada com sucesso no desenvolvimento de ontologias, em grande parte por uma separação de uma categoria especial para artefatos de informação que engloba entidades nos níveis 2 ou 3. Portanto, a distinção entre um diagnóstico (resultado de um processo de diagnóstico, dirigido a alguma doença) e da própria doença é explicitada. Berzell (2010) também explora a teoria dos três mundos de Popper ao colocar ontologias no terceiro mundo, que ele chama de "domínio do conhecimento proposicional objetivo" (Berzell 2010). Enquanto, do ponto de vista ontológico, Popper permite entender melhor a relação entre as entidades pertencentes ao mundo, do ponto de vista epistemológico, a teoria propõe que cada conceituação revela discrepâncias entre a realidade e as teorias sobre a realidade. Em outras palavras, ao colocar uma entidade (ex. André) e uma informação sobre uma entidade (ex. o nome "André", que denota uma pessoa na realidade) na realidade, a teoria assume que nem sempre a informação reflete a verdade. De fato, nossa hipótese básica é que os prontuários médicos representam tanto a realidade

quanto o processo cognitivo de médicos que interpretam a realidade. O assunto em questão é a diferença entre a manipulação de teorias que refletem a realidade e a manipulação de teorias que refletem os mundos cognitivos e informacionais.

Embora esta teoria de Popper tenha sido criticada (Bawden 2002), há pontos de vista favoráveis em que ela é considerada como um modelo útil para compreender a informação que descreve processos cognitivos (Abbott, 2004). Assim, pode-se encontrar adições e melhorias para as concepções de Popper, que propõem outras subdivisões das camadas originais (Bhaskar, 1978; Niiniluoto 1999) ou outra subdivisão de níveis de realidade em um estrato material, estrato psicológico e estrato social (Poli 2010).

Naturalmente, esta abordagem também mostra várias deficiências, especialmente quando se leva em conta que a maioria das representações de situações clínicas fazem uso da linguagem natural. Embora esta estrutura permita distinções entre os estados de conhecimento diferentes do médico, como a diferença entre um direito e um diagnóstico provisório, certamente não é capaz de transmitir nuances da comunicação clínica, como a preocupação de um paciente.

A teoria dos três mundos de Popper e sua distinção entre conhecimento objetivo e subjetivo proporcionam uma abordagem abrangente ao desenvolvimento de ontologias (entendidas como artefatos computacionais). Nossa motivação para procurar teorias alternativas para descrever entidades epistemológicas surgiu a partir da crescente evidência na comunidade de desenvolvimento de terminologia/ontologia que a junção invisível de informações e realidade cria problemas para inferências lógicas e consistência interna de um modelo. A teoria dos três mundos acrescenta uma nova perspectiva (filosófica) a este assunto.

Segundo Popper, as teorias descrevem algum nível de realidade, mas a manipulação posterior destas teorias é regida por regras específicas para o domínio ou nível de realidade. Portanto, como é possível afirmar que os seres humanos têm corações assim como todo mamífero (teoria sobre o mundo 1), também é possível afirmar que 7 é um número primo e fazer novas descobertas matemáticas exclusivamente manipulando as teorias sobre números. Isso sugere que as relações entre as entidades em tais teorias podem ser completamente distintas. Embora várias relações sejam adequadas para descrever entidades do mundo 1 (ex. relações mereológicas como em "coração é parte do ser humano"), a relação entre a cor normal da pele e a verdadeira cor da pele exige saber o

contexto (busca por uma doença) e uma representação complexa de combinações possíveis de cores da pele e alterações causadas por anormalidades fisiológicas. Essa percepção também sugere que devemos ser muito cautelosos ao tentar logicamente definir termos que só podem ser definidos por meio de analogias, imagens, padrões, etc

Ao afirmar que as teorias são criadas através de processo cognitivo, o arcabouço Popperiano torna mais claro que as relações entre as entidades do mundo 3 (coração representado em uma ontologia) e entidades do mundo 1 que estas denotam (coração no mundo real) dependem dos processos cognitivos que fazem esta análise. Isso significa que os modelos são teorias sobre o mundo com base em alguma perspectiva. Ontologias realistas são modelos vistos através de método científico rigoroso e análise ontológica. Assim, são coerentes e consensuais. No entanto, devido às limitações cognitivas ou técnicas, ou para a comunicação e compreensão, precisamos de simplificações (genericamente chamadas aqui de heurísticas) para criar teorias reduzidas. Nesta pesquisa, estamos particularmente interessados em teorias que exigem interpretação subjetiva de entidades do mundo um. Por exemplo, a afirmação "este paciente tem suspeita de pneumonia" nos diz que o autor da declaração acredita que o paciente tem uma pneumonia, e provavelmente se poderia inferir que há alguma evidência externa de pneumonia que levou a essa suspeita, mas não permite inferir que existe alguma infecção.

Finalmente, a teoria de Popper destaca a importância de constantemente rever as teorias atuais, uma vez que podem ser substituídas por teorias que explicam fenômenos mais estreitamente relacionados com entidades observados do mundo 1. Na medicina, muitas entidades não podem ser diretamente experimentadas pelos profissionais de saúde e são definidas de acordo com critérios mais ou menos objetivos. Como tal, a existência da entidade depende de encontrar evidências indiretas que satisfaçam os critérios, os quais podem ser alterados com o surgimento de novas evidências. Um paciente diagnosticado com autismo há 10 anos pode, simplesmente, deixar de ter a doença devido à mudança de critérios diagnósticos. Essa volatilidade de definições será importante ao representar informações do prontuário médico.

3.2.4 Vantagens e limitações do realismo ontológico

3.2.4.1 Realismo versus conceitualismo

A abordagem realista possui inúmeras críticas e limitações conforme recentemente descritos por Gary Merrill (MERRILL, 2010a; 2010b). Na ciência, a identificação de verdades é por si só um problema metafísico, visto que a ciência moderna consiste na aproximação da realidade, e não sua descrição exata. Sendo assim, a própria co-existência de universais independentes do pensamento e a necessidade de revisão destes universais com o avanço da ciência são problemáticas, adquirindo caráter mais epistêmico que realista. Rector corrobora com esta visão, e argumenta:

“(É) essencial que ontologias em sistemas de informação baseados em conhecimento sejam capazes de descrever entidades independentemente de sua existência no mundo real. Na verdade, frequentemente é necessário descrever entidades precisamente para dizer que elas não existem”²⁹ (RECTOR, 2010, p.6)

Questiona-se também que a presença de universais não é condição *sine qua non* para o desenvolvimento de ontologias. Desde que associado a boas práticas da filosofia da linguagem, lógica e semântica, é possível criar ontologias consistentes que não tenham reflexo como universais. Esta discussão metafísica, como já mencionado anteriormente, não será objeto deste trabalho. Porém, é pertinente avaliar outras alternativas metodológicas que utilizem princípios filosóficos para desenvolvimento de ontologias.

A adequação do realismo praticado na metafísica para o desenvolvimento de ontologias pode ser contraposto ao conceitualismo. Uma defesa consistente do conceitualismo pode ser vista em Cimino (1998; 2006). Segundo este autor, o uso de conceitos deve ser regido por uma série de mandamentos que garantem que cada conceito tenha seu significado explícito. O conceitualismo é mais próximo da linguagem corrente médica, pois é muito comum o uso de termos pouco definidos em situações clínicas novas e

²⁹ Note that, contrary to the way in which some realist philosophers formulate their criteria for “ontologies” in their sense for their purposes, it is essential that “ontologies” in “knowledge driven information systems” be able to describe entities whether or not they exist in the “real world”. Indeed, it is often necessary to describe entities precisely to say that they do not exist.

não completamente compreendidas. Na prática médica, os diagnósticos são geralmente presuntivos e baseados em dados incompletos - nesta situação, é muito difícil denominar um particular específico. As afirmativas são passíveis de revisão posterior e não representam verdades, mas a opinião do médico. Nesta situação, universais podem induzir a erros de avaliação e inferência. Ao mesmo tempo, Cimino argumenta que é possível a coexistência controlada entre conceitos e universais, para que a terminologia tenha a abrangência necessária ao uso, mantendo sua consistência.

Apesar da utilização de conceitos em praticamente todas as terminologias atuais, existem erros recorrentes que não foram corrigidos apesar do conhecimento (e consenso) sobre boas práticas. Isto ocorre, argumenta Smith (2006), porque a própria definição sobre o que é conceito é vaga. Conceitos são usados no lugar de propriedades, objetos e idéias sem distinção. A criação de ontologias em torno de conceitos é baseada no uso linguístico de termos e, portanto, sujeito a ambiguidades e erros de entendimento por indivíduos diferentes. Ontologias são artefatos direcionados primariamente para uso por computadores e qualquer ambiguidade linguística tem grande impacto em seu desenvolvimento.

Outra questão importante é a desconexão entre linguagem e ontologia. A abordagem conceitual irrestrita no desenvolvimento de terminologias e ontologias é sujeita à criação de termos através de seu uso corrente na linguagem. Como exemplo, os autores citam “AIDS está se espalhando rapidamente pela Ásia” e “AIDS é causada pelo vírus HIV”. Neste caso, o termo AIDS se refere a um conjunto na primeira frase, e a um universal na segunda frase (ou conceito, na visão conceitual). O uso da linguagem na formação de termos também cria conjuntos arbitrários que não se referem nem a conceitos ou universais, mas a classes (ex. “PessoasDeCabelosBrancosComMaisDe50AnosDeldade”).

Ao restringir o compromisso ontológico à realidade como descrita pela ciência, a capacidade de obtenção de consenso é significativamente maior em relação ao consenso entre duas conceitualizações. Esta vantagem é particularmente importante na descrição de ciências naturais como a biologia, em que o volume de dados é muito grande, mas existe grande consistência nas observações realizadas por diferentes organizações. Mesmo no domínio médico, declarações anatômicas e fisiológicas apresentam consenso quando formalizada como universais. Uma abordagem conceitual irrestrita pode gerar termos como “PressãoArterialElevada” cuja definição seria completamente arbitrária e poderia variar de acordo com a comunidade observadora, impedindo a interpretação adequada. O valor que determina que a pressão está elevada depende das doenças do paciente, e da capacidade

de mensuração. Além disso, como é muito difícil medir a pressão exata em ambulatório (ex. 117,5 mmHg), os valores convencionados são quase sempre múltiplos de 5 – o que não necessariamente se relaciona com a realidade, mas com a possibilidade de mensuração. A estas divisões derivadas de convenções sem uma relação intrínseca e direta com o objeto se dá o nome de limites *fiat* (*fiat boundaries*) (SMITH; VARZI, 2000).

Técnicas para avaliação da robustez da modelagem ontológica, como o OntoClean (GUARINO; WELTY, 2002) sugerem a avaliação dos termos através de critérios como essência, rigidez, identidade e unidade, mantendo o uso adequado de relações de subsunção (relações do tipo *é_Um*). Igualmente, restrições lógicas ao uso de termos que seriam mal-interpretados é essencial, como o exemplo “não-fumante”, que se aplica a pessoas, mas também a mesas e até ao próprio cigarro, que é por si só não fumante. A abordagem top-down, especialmente baseada em uma ontologia formal de alto-nível como o BFO ou a DOLCE, aumenta a consistência na modelagem.

3.2.4.2 Ontologia e epistemologia

A realismo ontológico possui limites estritos para o que pode ser representado em uma ontologia. Os universais devem seguir as teorias científicas para assegurar a proximidade com a realidade. Além disso, os particulares devem obrigatoriamente existir na realidade, sendo restrita a representação de entidades que não podem ser diretamente experimentadas (*prima facie*) através de evidência experimental. Segundo Smith e Ceuster, este é o princípio da instanciação que diz que “um termo só deve ser incluído em uma ontologia de referência se existir evidência empírica que as instâncias a que o termo se refere existem na realidade”³⁰. (SMITH; CEUSTERS, 2010, p.152)

Enquanto a definição dos objetivos e limites de uma ontologia é importante para garantir a consistência, estes limites restringem o escopo do que pode e não pode ser representado. A primeira restrição do realismo é o compromisso ontológico com a realidade, e não com mundos possíveis. É vedada a representação de entidades mitológicas (unicórnios) e entidades de ficção (o violino de Sherlock Holmes). Enquanto essencial para a

³⁰ A term should be included in a reference ontology only if there is experimental evidence that instances to which that term refers exist in reality.

representação de alguns domínios, a documentação da prática médica não requer representação de entidades inexistentes.

Uma limitação mais relevante para a documentação médica é a distinção entre a realidade e o conhecimento sobre a realidade – em outras palavras, a distinção entre ontologia e epistemologia. Epistemologia é estudo de como entidades cognoscentes conhecem a verdade sobre algum evento da realidade. A diferença entre os dois termos pode ser evidenciada ao se avaliar como se definem entidades na ontologia e na epistemologia. A ontologia diz respeito a objeto, processo, evento, todo, parte, determinação, dependência, composição, etc. Epistemologia diz respeito a crença, verdade, probabilidade, confirmação, conhecimento e suas variações (POLI, 2010), enquanto ontologia é uma teoria das coisas, epistemologia é a teoria do conhecimento.

A dependência mútua entre a existência da entidade e o saber sobre a entidade frequentemente torna a distinção entre ontologia e epistemologia pouco nítida. Bodenreider e outros descrevem algumas categorias de termos epistemológicos que podem ser comumente identificadas em terminologias biomédicas (BODENREIDER *et al.*, 2004), dentre as quais consideramos relevantes as seguintes³¹:

- Termos contendo critério de classificação: é comum a criação de termos que não representam universais, mas pretendem transmitir outros tipos de informação. Variações de entidades patológicas são um caso comum. “Convulsão febril” (convulsão causada por febre, que pode ocorrer em crianças sem disfunções neurológicas) e “convulsão afebril” (normalmente causada por disfunções neurológicas estruturais, como lesões cerebrais) são variações para a mesma entidade “convulsão”, mas transmitem informações que induzem a conclusão sobre a causa do evento. Também dentro desta categoria estão os termos que adicionam à definição da entidade informações que não possuem relação direta com a existência desta entidade. Por exemplo, a distinção entre “Tuberculose pulmonar, sem menção de confirmação bacteriológica ou histológica” e “Tuberculose pulmonar, com

³¹ Estes autores citam ainda a categoria de termos criados para se obter uma partição completa do domínio, como “Outros transtornos especificados dos glóbulos brancos”. Porém, esta categoria não é pertinente pois o foco do estudo são os aspectos epistemológicos da documentação médica, e não os erros de desenvolvimento dos vocabulários atuais.

confirmação somente por cultura” não se refere a nenhuma diferença entre as doenças a que se referem, mas à forma como o diagnóstico foi feito.

- Termos que referem a modalidade, incerteza e imprecisão: durante a documentação da situação clínica, é praticamente impossível que um profissional consiga determinar com exatidão o estado de um paciente. Normalmente o diagnóstico e a terapêutica recomendada são baseados na probabilidade de uma determinada doença de acordo com a semelhança da condição com a descrição da patologia pela literatura científica, somada a dados estatísticos da frequência da decisão. Portanto, é comum que o profissional atribua um valor de probabilidade, ex. “possível câncer”, “provável câncer”, “diagnóstico definitivo de câncer”. Outros termos referentes a modalidade versam sobre a causa da patologia (ex. “diarréia provavelmente de origem infecciosa”) e prognóstico (ex. “câncer provavelmente benigno”). Termos que se referem a imprecisão também são comuns na medicina, especialmente em terminologias utilizadas para fins epidemiológicos como o CID-10. Um exemplo claro são os marcadores “não especificado”, que não adicionam informação sobre a doença mas sobre o que se sabe sobre aquela instância específica da doença.
- Termos relativos a normalidade e divisões arbitrárias (*fiat boundaries*): alguns termos usados na documentação médica pretendem transmitir instruções sobre como a informação deve ser interpretada, e não sobre a entidade em si. Por exemplo, um neonato com a frequência cardíaca de 110 batimentos por minuto tem a “frequência cardíaca normal”. Porém, um adulto com esta frequência apresenta “taquicardia”. Além de induzir a interpretação, as divisões frequentemente se baseiam em dados estatísticos e populacionais, podendo apresentar variação entre países diferentes. Por fim, é importante notar que parte do conhecimento médico é baseada em eventos históricos que tiveram definição quase arbitrária.

As categorias acima se referem à terminologia existente em vocabulários médicos, destinados à descrição clínica do paciente. Porém, uma ontologia deve representar o conhecimento médico e não somente as instâncias. Nesta tarefa, também podemos encontrar dificuldades de representar o conhecimento básico da medicina de acordo com a metodologia realista. Schulz e outros (2009b) argumentam que a tentativa de codificar conhecimento default (conhecimento tipicamente verdadeiro a menos que existam

evidências em contrário, como a afirmação “pássaros voam”) ou probabilístico utilizando ontologias da forma com existem atualmente tem grandes chances de produzir modelos incorretos. Ontologias não podem expressar afirmativas probabilísticas uma vez que considera-se o princípio que “todas as propriedades associadas a um tipo (universal) em alguma ontologia deve ser verdadeiro para todas as instâncias de um tipo particular”³²(STENZHORN *et al.*, 2008, p.3770).

Portanto, conhecimento probabilístico como fatores de risco para doenças ou chance de mortalidade não podem ser expressas de forma simples e intuitiva. Ao dizer que fumar aumenta a chance de câncer de pulmão, não dizemos nada sobre qualquer instância de câncer. Uma solução seria representar estas afirmações como disposições probabilísticas, entidades dependentes que não precisam de realização para existir, levando à criação de nós, como por exemplo “*Risco de câncer*”, na ontologia. Já a definição precisa da probabilidade não seria possível na atual concepção de ontologia (50% de risco de morte), exigindo outras formas de representação.

Por fim, o realismo impõe restrições ao uso de ontologias em sistemas de informação hospitalares, uma vez que dificulta a expressão de alguns termos administrativos comuns. Por exemplo, uma cirurgia cancelada não é e nunca foi uma instância de uma cirurgia, mas um planejamento.

A busca por mecanismos robustos de representação da prática médica são o grande motivador deste trabalho. Como vimos, o profissional de saúde representa em um prontuário diversas afirmações que não podem ser representadas adequadamente em ontologias realistas devido a restrições exigidas pela metodologia. Atualmente, a forma mais próxima para expressão destas afirmações (que estamos chamando genericamente de termos e afirmações epistemológicas, visto que estão relacionados à forma como o profissional aborda a realidade) é o uso de modelos informacionais (SCHULZ; CORNET, 2009).

Porém, não existe ainda uma metodologia consolidada para ligar ontologias e modelos informacionais (RECTOR *et al.*, 2009). Além disso, os modelos possuem inconsistências de modelagem, e requerem tecnologias específicas para seu funcionamento,

³² all properties associated with any given type in any ontology must be true for all instances of this particular type

uma vez que sua criação é guiada prioritariamente por casos de uso. Para buscar uma resposta para as limitações das ontologias mantendo os benefícios da modelagem formal e lógica, vamos explorar três alternativas: o uso de ontologias informacionais, uso de redes conceituais informais (como a linguagem SKOS) e abordagem interpretativa da hermenêutica.

3.2.4.3 Ontologias informacionais para representação de afirmativas epistemológicas

As afirmativas epistemológicas se referem à forma como conhecemos as coisas. A descrição das entidades de conhecimento e informação (considerando conhecimento algo próprio do ser cognoscente, e informação como uma entidade externa ao ser) não descreve uma entidade, mas é *sobre* uma entidade (entidade intensional, no sentido dado por Franz Brentano). Portanto, é falso representar uma “*cirurgia cancelada*” como uma “*cirurgia*”, pois esta nunca aconteceu. Uma solução é representar a “*cirurgia cancelada*” como um objeto informacional, um planejamento, que é sobre uma “*cirurgia*”. Neste caso, a cirurgia nunca será concretizada, mas o planejamento existiu durante um espaço de tempo definido e evidenciado.

No decorrer deste estudo, iremos explorar as possibilidades de representação de termos epistemológicos através de ontologias informacionais, como proposto por Schulz, e outros (SCHULZ *et al.*, 2010a). Para tal, utiliza-se a *Information Artifact Ontology* (IAO)³³. A IAO é uma ontologia que representa as entidades informacionais através de artefatos. A criação desta ontologia foi motivada pela necessidade de representar entidades comuns na pesquisa científica como publicações, resultados e bases de dados. Outros exemplos de artefatos informacionais são número serial, endereço de e-mail, jornal, nota de laboratório.

Pode-se perceber a interseção entre a ontologia informacional IAO e os modelos de informação na medicina. Estes modelos (como o HL7 e o OpenEHR) pretendem padronizar a forma como os registros em medicina são feitos, definindo meta-dados e terminologias para a criação destas mensagens. Portanto, um metadado como “Diagnóstico primário” não denota diretamente a doença, mas afirma que tal diagnóstico é o caso, ou seja, o

³³ <http://code.google.com/p/information-artifact-ontology/>

diagnóstico é sobre a doença do paciente. De fato, “diagnóstico” é uma entidade informacional bem definida e consensual e poderia ser representada desta forma, com todos os benefícios da representação ontológica (re-uso, inferências consistentes, consenso).

3.2.4.4 Redes conceituais

Nas seções acima, defendemos as vantagens do realismo sobre o conceitualismo como base para uma metodologia consistente de desenvolvimento de sistemas terminológicos, particularmente ontologias. Apesar de oferecer consistência e coerência, o realismo exige que a representação seja restrita pela realidade como descrita pela ciência. Este trabalho avaliará o uso de conceitos para representação de informações epistemológicas, ampliando o escopo coberto pela representação. Para tal, seguiremos as recomendações descritas por Klein e Smith. Para definir melhor o significado de conceito e evitar o uso inadequado do termo, usaremos a definição destes autores: “conceito deve ser usado exclusivamente para referir 1) ao significado do termo geral correspondente, este termo sendo 2) único e 3) consensual entre pessoas da mesma área de conhecimento” (KLEIN; SMITH, 2010, p.722).

O uso de conceitos na representação de informações clínicas, apesar de ser mais sujeito a inconsistências, permite a representação mais próxima da linguagem, já que representa o significado do termo, e não denota uma entidade (universal ou particular). A definição de cada termo pode ser feita através do uso de linguagens formais ou através de descrição em linguagem natural, de acordo com homogeneidade da interpretação do significado do termo (conceitos como cima e baixo são muito intuitivos e interpretados de forma homogênea, enquanto o termo AIDS requer explicações mais precisas). O relacionamento entre conceitos deve ser feito através de relações semânticas “mais_abrangente” e “mais_especializado”³⁴, considerando se o significado dos conceitos. Adicionalmente, podemos considerar a relação “relacionado”, proposta na definição do padrão W3C SKOS³⁵ (*Simple Knowledge Organization System*).

³⁴ Referindo-se a “broader_than” e “narrower_than”, respectivamente

³⁵ <http://www.w3.org/2004/02/skos/>

3.2.4.5 Hermenêutica

As duas abordagens mencionadas para solucionar as limitações das ontologias realistas buscam soluções práticas para representação do conhecimento mantendo a consistência proposta pelo realismo. O realismo ontológico busca responder à multiplicidade de modelos conceituais e linguagens com restrições sobre o quê pode ser representado sem ambiguidades. A hermenêutica propõe uma quebra paradigmática radical, ao afirmar que a definição do ser depende de uma interpretação constante do ser cognoscente e, portanto, não pode ser definida a priori.

O estudo da hermenêutica aqui apresentado é baseado nos trabalhos de Fonseca e Martin (FONSECA; MARTIN, 2005a; FONSECA; MARTIN, 2005b; FONSECA; MARTIN, 2009). Por sua vez, estes autores se referem à hermenêutica de *Gadamer*, que afirma que o significado de uma representação ou um símbolo é aberto a interpretação. Dois conceitos citados pelos autores são particularmente relevantes para o desenvolvimento de ontologias: quebra e fusão de horizontes.

O conceito de quebra (*breakdown*), criado por Heidegger e transposto para a inteligência artificial por Winograd e Flores (1987), sugere que o conhecimento sobre um objeto não se dá através do contato com o mesmo, mas através de uma situação inesperada na interação com este objeto. Um exemplo comum é o ato de martelar um prego. Se a ação ocorre como esperado, o martelo não é visto como um objeto com propriedades definidas (função, categoria), mas é integrado ao conjunto maior de ações que compõe o martelar (*ready-at-hand* ou *zuhanden*). Porém, se ao invés de acertar o prego a pessoa acerta o dedo, ocorre uma quebra, e o objeto se torna o centro da atenção (*Present-at-Hand* ou *vorhanden*). A implicação para o presente trabalho é a sugestão da criação de espaços para quebras na comunicação, em que usuários e criadores das ontologias pudessem promover a resolução de inconsistências em busca de ontologias mais naturais (*ready-at-hand*).

O conceito de fusão de horizontes propõe que a compreensão entre dois comunicadores requer que seus horizontes – uma perspectiva limitada mas aberta, composta por um conjunto de pressupostos e valores – sejam unidos. Para a fusão, duas condições devem ser atendidas. Os comunicadores precisam saber qual é o conteúdo de conhecimento comum, incluindo as dimensões em que os objetos da conversação permanecerão constantes ou exibirão variação previsível. Em segundo lugar, os dados

devem ser ordenados segundo as dimensões em que exista variação no significado. São estas dimensões que estão em questão e que devem ser aproximadas para a fusão. (FONSECA; MARTIN, 2005a)

Acredita-se que estudo da proposta de ontologia hermenêutica resulta em contribuições relevantes na avaliação das vantagens da representação dos termos epistemológicos por ontologias informacionais e conceitos, por levar em consideração a importância da interpretação do receptor da mensagem e sua forma de conhecer a realidade através da mensagem. No caso da comunicação médica é essencial que toda mensagem clínica seja corretamente interpretada e armazenada, de forma a ser recuperada consistentemente. O objetivo do estudo da hermenêutica neste trabalho é alargar a compreensão sobre a natureza do processo de comunicação e significação, para que as insuficiências das abordagens estudadas sejam prontamente reconhecidas para que uma solução seja desenvolvida. Por fim, concordamos com a afirmação de Hjørland:

“Cientistas da informação devem se perguntar a pergunta pragmática: Dado os diferentes interesses e paradigmas da área, quais tipos de interesse um sistema específico deve apoiar? Qual diferença faz se alguns tipos de relações semânticas são usados em detrimento de outras? Talvez a tarefa mais importante do profissional da informação seja fazer visíveis os diferentes interesses e paradigmas para que o usuário faça uma escolha informada”³⁶ (HJORLAND, 2007, p.389).

³⁶ Information scientists should ask the pragmatic question: Given the different interests and paradigms in the field, what kinds of interest should this specific system support? What difference does it make whether some kinds of semantic relations are used at the expense of others? Perhaps the most important task of the information professional is to make the different interests and paradigms visible so that the user can make an informed choice

3.3 A TROCA DE MENSAGENS EM BIOMEDICINA – SEMIÓTICA E MODELO DE INFORMAÇÃO EM SAÚDE

Como discutimos acima, podemos distinguir duas correntes que propõem o uso de ontologias (como artefato de sistemas de informação): 1) a corrente conceitual, que propõe o uso de ontologias para representar o conhecimento sobre algum domínio através de axiomas lógicos, permitindo inferência automatizada; 2) a corrente realista, que propõe o uso de ontologias para representação da realidade, garantindo que a representação é consistente e robusta. O objetivo deste trabalho situa-se entre estas duas correntes, pois a complexidade do domínio impede que o significado dos termos e a relação entre eles sejam consistentes sem uma representação adequada do conhecimento médico. Enquanto a ontologia realista é particularmente adequada para garantir robustez e consenso, os princípios filosóficos subjacentes não permitem a representação de termos fora do domínio do real, que chamamos anteriormente de termos epistemológicos.

Nesta seção, pretendemos avaliar a contribuição de outras formas de comunicação entre sistemas de informação em saúde que levam à interoperabilidade entre estes sistemas. Para tal, traçaremos um paralelo entre o estudo da semiótica e o estudo da interoperabilidade, comparando as noções de sintaxe, semântica e pragmática. Em seguida, faremos uma exposição dos modelos de informação em saúde, ferramentas importantes na atualidade para a comunicação de sistemas de informação.

Não faz parte do escopo deste trabalho explorar minuciosamente um assunto denso como as teorias de significado e, portanto, nos limitaremos a demonstrar algumas teorias relevantes ao nosso estudo.

3.3.1 Semiótica e significado

Desde sua origem, as ontologias (como artefato computacional) tiveram grande influência da comunidade da representação de conhecimento, oriunda por sua vez da inteligência artificial. O centro desta corrente é baseado na modelagem de eventos do mundo real, exteriorizando o modelo através do uso de símbolos e suas regras de manipulação (GOFAI – *Good Old Fashioned Artificial Intelligence*). Ao tratar de interoperabilidade, é essencial explorar o estudo dos sinais e suas implicações para o desenvolvimento de ontologias. Sinais são padrões de representação, que tomam materialidade sob a forma de códigos numéricos, palavras, padrões de luz e sombra (desenhos), objetos estáticos e em movimento, dentre outras formas. Três termos se destacam sobre este tema: Sintaxe, Semântica e Pragmática.

Sowa (2000), baseado em Peirce, define os três termos como ramos da semiótica (estudo dos sinais, ou signos) da seguinte forma:

- Sintaxe: estudo que relaciona um signo a outros signos;
- Semântica: estudo que conecta os signos às coisas no mundo, e padrões de signos a padrões correspondentes que ocorrem entre coisas às quais o signo se refere;
- Pragmática: estudo que relaciona signos aos agentes que os utilizam para se referir a coisas no mundo, e para comunicar suas intenções sobre essas coisas a outros agentes, que podem ter intenções semelhantes ou diferentes a respeito das mesmas coisas ou coisas diferentes.

A semiótica explica o significado através da relação entre o símbolo, o referente, e a referência (ou pensamento). O modelo linguístico mais comum é o descrito por Ogden e Richards (Figura 3). O símbolo ou sinal³⁷ é um padrão de dados identificado e

³⁷ Peirce divide sinais em três tipos: ícones (se associa ao objeto através de semelhança estrutural, como a foto de um pão e uma padaria); índice (denota o objeto por uma conexão real entre o sinal e o objeto, como o cheiro de pão e uma padaria); e símbolo (denota o objeto por um sinal arbitrariamente selecionado, como a palavra “padaria”). Como estes termos

individualizado pelo observador. O conteúdo do símbolo evoca uma associação entre o símbolo e o referente, sendo o símbolo uma forma de representação da entidade referenciada. A terceira entidade, chamada de pensamento, referência ou conceito, relaciona o símbolo ao seu referente.

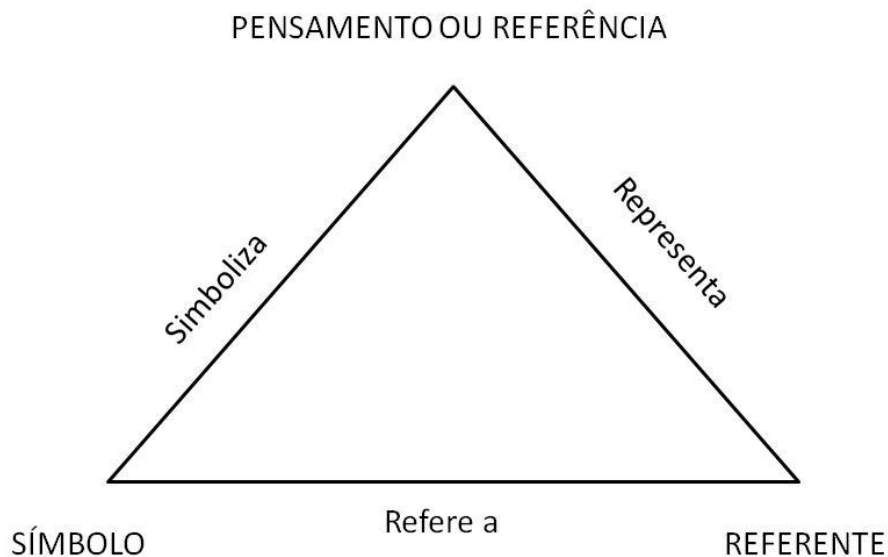


Figura 3 – Triângulo semiótico de Ogden e Richards

A sintaxe é classicamente considerada o estudo das regras de formação e manipulação de símbolos. A sintaxe na análise linguística se refere à disposição de cada termo na frase e das frases no discurso, sem nenhuma relação direta com o significado individual ou coletivo dos termos. A semântica, por outro lado, pretende definir o significado dos termos e das frases.

Assim como a busca da resposta para a pergunta ontológica “o que é ser”, o estudo da linguagem e do processo de significação é milenar. A palavra significado pode ser compreendida de várias formas. Sellars apud Rapaport (2002) aponta algumas interpretações para a palavra significado:

- Significado como tradução:

aparecem em diferentes contextos fora dos trabalhos de Peirce, não faremos esta distinção e usaremos símbolo e sinal indistintamente no sentido mais abrangente.

- “redondo” significa circular
- “*cheval*” significa cavalo
- Significado como sentido³⁸
 - “redondo” expressa o conceito Circularidade
 - “*cheval*” expressa o conceito de Ser Cavalo
- Significado como o ato de nomear
 - “redondo” nomeia o conceito Circularidade
 - “*cheval*” nomeia cavalos
- Significado como conotação
 - “*cheval*” conota a propriedade de possuir quatro patas
 - “Parigi” conota a propriedade de ser a capital da França
- Significado como denotação:
 - “redondo” denota coisas circulares

É possível identificar dois paradigmas dominantes no estudo do significado (SPEAKS, 2010): o paradigma lógico-positivista e o pragmático. O lógico-positivista, representado por teóricos como Carnap e Wittgenstein (jovem) sugere que as expressões são substitutos das entidades e seus significados são as entidades a que eles referem. Esse paradigma é composto por várias teorias, que chamaremos genericamente de teorias proposicionais e corresponde à semântica. O paradigma pragmático, representado por teóricos como John Dewey e Wittgenstein (tardio) sugere que as expressões são ferramentas para interação e seus significados são suas funções embutidas na interação, permitindo a continuidade satisfatória desta interação. Chamaremos as teorias resultantes deste paradigma de teorias fundamentais do significado baseado em (SPEAKS, 2010).

As teorias proposicionais afirmam que a linguagem é composta de proposições que podem ser verdadeiras ou falsas. Estas teorias têm grande proximidade da teoria de referência que, ao invés de relacionar as expressões ao significado, relacionam as expressões às suas contribuições para determinar o valor verdade das expressões. Por exemplo, ao afirmar que “Dilma Roussef é a presidente do Brasil”, é possível determinar qual é a entidade a que o nome “Dilma Roussef” se refere, e avaliar a condição verdade da

³⁸ No original “Meaning as sense”

expressão. Como Carnap coloca, “Saber a condição verdade de uma sentença é saber o que é afirmado por ela (...)– em termos comuns, seu “significado”³⁹. (CARNAP, 1910) apud (CREATH; FRIEDMAN, 2007, p.171). Na visão de alguns autores desta corrente, a linguagem natural poderia ser simplificada e re-escrita em lógica formal, podendo ser tratada matematicamente.

A criação de significado se dá conforme a atribuição adequada de um referente a um símbolo, já que o conteúdo do símbolo determina o referente. O processo de identificar o referente nem sempre é óbvio. Um exemplo comum são os termos “estrela vespertina” e “estrela matutina” se referem à mesma entidade no mundo real, o planeta Vênus. Porém, intuitivamente os dois termos têm significados diferentes, transmitindo mensagens sobre o momento em que o astro celeste fica mais visível. Outro problema da teoria é a necessidade da interpretação do contexto para algumas expressões, como “eu”, “agora” e “há 100 anos”, visto que o referente se altera de acordo com condições externas à frase. Podemos conceber o processo de identificação do referente através de duas formas: intensão e extensão.

Intensão é o conjunto de funções ou regras que permitem o reconhecimento do referente, mesmo que o referente exato não possa ser reconhecido. Por exemplo, nós somos capazes de interpretar a expressão “o maior carro azul do mundo”, mesmo sem saber exatamente qual entidade individual é referida na expressão. A frase transmite um conjunto de regras que permite a identificação da entidade caso saibamos o tamanho e a cor de todos os carros do mundo. Entidades intensionais não precisam existir no mundo real, uma vez que a regra permite a compreensão independente do conhecimento sobre o verdadeiro referente. São entidades intensionais as entidades ficcionais, os objetos não existentes (unicórnios), os objetos impossíveis (quadrados redondos), os objetos do pensamento, as proposições e os objetos arbitrários (RAPAPORT, 2009).

Extensão é a identificação do referente através da criação de um conjunto de entidades que correspondem ao termo ou símbolo. Um termo pode ser definido através de uma definição ostensiva, ou seja, por um processo de apontar todas as entidades que o

³⁹ “Since to know the truth conditions of a sentence is to know what is asserted by it, the given semantical rules determine for every sentence of [a constructed language~what it asserts—in usual terms, its “meaning” – or, in other words, how it is to be translated into English”

termo referencia – independente das regras que estejam embutidas no pensamento do observador.

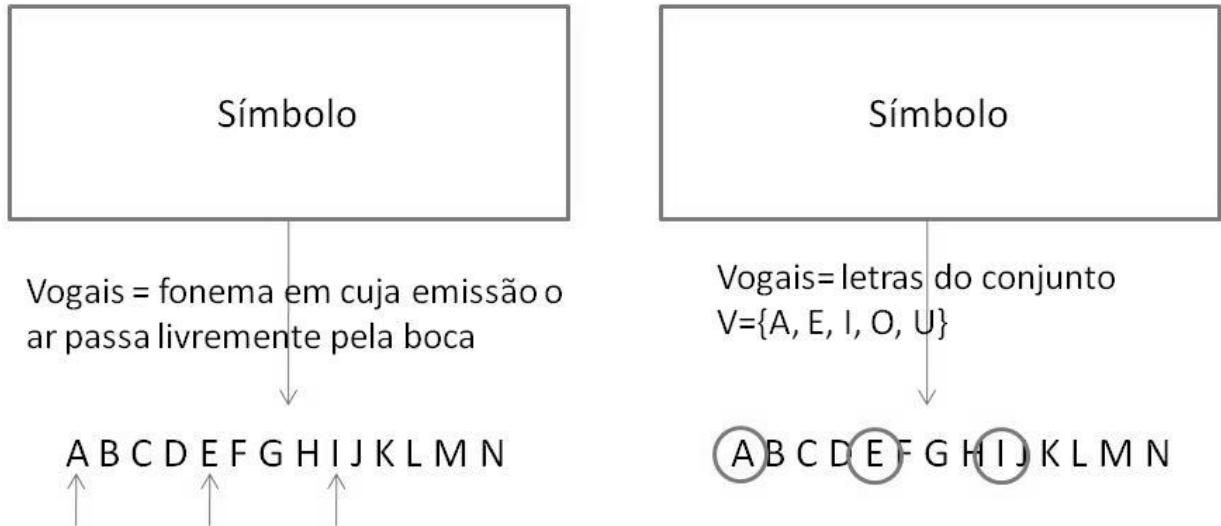


Figura 4 – Representação de intensão e extensão

As teorias fundamentais do significado (pragmática da semiótica) são baseadas na idéia que a linguagem é uma ferramenta abrangente de comunicação e, portanto, teria uma função de coordenar as interações entre as pessoas. Wittgenstein exemplifica:

“Água!

Fora!

Ai!

Ajuda!

Perfeito!

Não!

Você ainda está disposto a chamar estas palavras de “nomes de objetos?”⁴⁰

(WITTGENSTEIN, 1997, p.13)

Nesta concepção, existe uma diferença entre o significado, ou conteúdo de uma expressão linguística (ou seja, aquilo que a expressão descreve) e o que a pessoa quis dizer esta frase. Na expressão publicitária “Não é uma Brastemp” ao se referir a

⁴⁰ “Wasser! Fort! Au! Hilfe! Schön! Nicht! Bist Du nun noch geneigt, diese Wörter “Benennungen von Gegeständen“ zu nennen?”

eletrodomésticos de marcas diferentes, o objetivo do publicitário não é fazer uma afirmação corriqueira sobre o produtor do equipamento, mas sugerir implicitamente a diferença na qualidade entre os produtos. Um dos principais teóricos desta corrente é *Paul Grice*. Segundo Grice, o significado pretendido da fala poderia ser analisado da seguinte forma:

a significa *p* ao dizer *x* se e somente se *a* pretende ao dizer *x* que:

1. Sua audiência acredita em *p*,
2. Sua audiência reconheça sua intenção e
3. 1) ocorra baseado em 2);

O exemplo torna clara a divisão entre semântica e pragmática no domínio das ontologias. Porém, Rapaport (2002) questiona a divisão entre sintática e semântica. Rapaport argumenta que o comportamento linguístico adequado (como julgado por um observador externo) de um sistema de inteligência artificial não requer o contato com o mundo através da percepção sensorial. Exemplificando com um sistema de inteligência artificial, ele aponta que semântica não necessariamente é uma correspondência entre a linguagem e o mundo, mas entre a linguagem comum e a linguagem particular de um sistema – a linguagem que representa o conhecimento sobre as coisas do mundo.

Nesta concepção, a frase “*red* quer dizer vermelho” não diz nenhuma informação sobre o termo “*red*” que não poderia ser formulado de outra forma. Porém, se alguém quer saber o que “*red*” quer dizer, basta dizer que significa “vermelho” e as mesmas associações serão transmitidas ao novo termo, ou seja, houve apenas uma manipulação sintática que associou dois símbolos a um mapa mental. Segundo este autor, “falar sobre “o” sentido de uma frase não é adequado; o contexto deve ser levado em conta. Mas o sentido geral do argumento se mantém: Sentidos de sentenças são dados por pensamentos, não por condições verdade” ⁴¹ (RAPAPORT, 2002, p.19).

Apesar da divergência com a definição de Carnap apontada acima, os símbolos e a interpretação podem ser individualizadas a partir da necessidade de uma resposta para

⁴¹ Of course, to talk of ‘the’ meaning of a sentence is misleading; context needs to be taken into account. But the broader point holds: Meanings of sentences are provided by thoughts, not by truth conditions.

continuidade do diálogo. Ao invés dos símbolos transmitirem o significado (e do interlocutor que escuta ter que interpretar a expressão), os símbolos ativam o significado na mente do interlocutor, como que ativassem os conceitos (RAPAPORT, 2003). Para este autor, a semântica de sistemas computacionais é uma semântica sintática - é uma semântica formal em que o significado diz respeito a uma função matemática (função interpretação).

Ainda outra abordagem é apresentada por William Woods (2007), que propõe uma semântica procedural. Segundo este autor, são incorretas as seguintes noções do que é semântica (WOODS, 2010):

- A totalidade dos conceitos conectada a um conceito são sua semântica
- Qualquer coisa que um sistema faz com a entrada de dados é sua semântica
- Não há diferença entre sintaxe e semântica
- Semântica está no olhar do observador (os nomes das classes)

A idéia da semântica procedural é que a semântica de frases em linguagem natural pode ser caracterizada em um formalismo cujo significado seja definido por procedimentos abstratos que podem ser realizados. Portanto, o significado de um substantivo é o procedimento para determinar se ele é verdadeiro ou falso, e o significado de uma ação é a habilidade de realizar a ação ou dizer se esta foi realizada.

Quadro 1 – Definição de significado de acordo com as teorias apresentadas

	Semântica Sintática	Teorias Proposicionais	Semântica procedural	Pragmática
O que é significado?	Relação entre uma linguagem pública e uma linguagem privada	Relação entre o termo e o mundo	Procedimentos abstratos que possam ser realizados	Uma proposição como pretendida pela pessoa que declara

Almeida e colaboradores (ALMEIDA *et al.*, 2011b) oferecem uma revisão abrangente sobre o significado de “semântica” para a comunidade da Web Semântica. Estes autores diferenciam a semântica para humanos, que aumenta à medida em que mais informação é adicionada sobre um termo, e semântica para máquinas, que aumenta à

medida que a expressividade formal (simbólica) permite representações mais complexas com inferências corretas.

O estudo atual das ontologias (e de outras abordagens para organização e representação de conhecimento) e da lógica subjacente para sua representação é baseada quase que exclusivamente em teorias proposicionais (HJORLAND, 2007), uma vez que propõe métodos objetivos de avaliação do significado, além de permitir a aproximação da lógica com a linguagem. O objetivo proposto das ontologias é ser utilizado como artefato de representação do conhecimento e linguagem independentemente do utilizador – uma abordagem puramente pragmática oferece dificuldades ao exigir que a audiência do interlocutor atribua crenças e intenções, criando uma ambigüidade contraprodutiva. Soma-se a isto a necessidade que interlocutor e audiência compartilhem a práxis (hábitos, costumes) e modos de ser semelhantes. Citando novamente Wittgenstein, “se um leão pudesse falar, nós não poderíamos entendê-lo”⁴² (WITTGENSTEIN, 1997, p.223)

Para para a comunidade de integração de sistemas, o significado de semântica, a relação entre semântica e sintaxe e as soluções propostas para a comunicação entre computadores têm conotação bem mais prática. É um consenso bem estabelecido que a comunicação entre diversos sistemas de informação em saúde é um pré-requisito para atingir os objetivos de melhoria dos serviços de saúde. O cuidado em saúde é tipicamente multidisciplinar e realizado em vários locais diferentes (hospital, centro de saúde, laboratórios e centros de diagnóstico por imagem, etc.), e cada ponto de atenção precisa de todas as informações pregressas para tomar a conduta correta. Porém, a complexidade do domínio aumenta a complexidade das mensagens e dificulta a troca de informações tratáveis por computador. A interoperabilidade semântica é advogada como a principal solução para este problema. A multiplicidade de interpretações para o termo "semântica" gera grande número de definições para a interoperabilidade. Interoperabilidade semântica pode ser definida como:

"A capacidade, facilitada por aplicações e sistemas de tecnologias da informação e comunicação (TIC), de trocar, compreender e agir em informação de conhecimento de pacientes/cidadãos e outras fontes relacionadas à saúde, entre profissionais de saúde, pacientes e outros atores e organizações linguística e culturalmente diferentes,

⁴² Wenn ein Löwe sprechen könnte, wir könnten ihn nicht verstehen.

*dentro e através de jurisdições do sistema de saúde de maneira colaborativa*⁴³ (STROETMANN et al., 2009, p.10).

Nesta seção veremos a definição de semântica e sintaxe na comunidade de integração de dados para posicionar os modelos de informação e ontologias nas categorias adequadas.

Ao avaliar o grau de compreensão de uma expressão transferida de um sistema para outro, existem várias possibilidades de interoperabilidade (EUZENAT J., 2001):

- Codificação: ser capaz de segmentar a representação em caracteres
- Léxica: ser capaz de segmentar a representação em palavras (ou símbolos)
- Sintática: ser capaz de estruturar a representação em sentenças estruturadas (ou fórmulas ou afirmativas)
- Semântica: ser capaz de construir o significado proposicional de uma representação
- Semiótica: ser capaz de construir o significado pragmático

A interoperabilidade sintática é a interoperabilidade no nível da aplicação que permite que múltiplos componentes de software co-existam, mesmo que suas linguagens de implementação, interface e plataforma de execução sejam diferentes. Interoperabilidade semântica pode também ser definida como a faculdade de interpretar as anotações no nível semântico, ou seja, atribuir a cada pedaço de conhecimento importado a interpretação correta ou conjunto de modelos. A distinção entre interoperabilidade sintática e semântica é mais bem descrita através de um exemplo (PARK; RAM, 2004, p.596):

Vamos assumir que temos um dado de valor “37”. O que “37” significa para nós? Esta informação pode ser inútil a menos que saibamos seu significado. Se “37” esta armazenado na base de dados de temperatura, nós podemos assumir com segurança que “37” é sobre temperatura. Se está armazenado no atributo da idade da pessoa, então podemos facilmente concluir que “37” é a idade da pessoa.

⁴³ the ability, facilitated by Information and Communication Technologies (ICT) applications and systems, to exchange, understand and act on citizens/patients and other health-related information and knowledge among linguistically and culturally disparate health professionals, patients and other actors and organizations within and across health system jurisdictions in a collaborative manner

Neste caso, uma vez que podemos identificar seu significado baseado na informação do esquema da base de dados (ou seja, pelo nome do atributo), nós trabalhamos com a “semântica fraca”. Semântica fraca pode ser identificada baseada na informação estrutural, sintática e de valor/extensional contida nas bases de dados. Continuemos a examinar as complicações associadas ao valor “37”. Quando tentamos entender a semântica do dado, nós devemos identificar o contexto em que os dados estão sendo usados. Por exemplo, em vários países asiáticos, o esquema de contagem da idade é diferente daquele usado no Ocidente. Nestes países asiáticos (ex. Coréia), no dia em que a criança nasce, sua idade é um. Após, a idade do bebê aumenta 1 ano a cada Ano Novo. Então, se um bebê nasce em 31 de dezembro, sua idade é dois em seu segundo dia de vida. Portanto, a mesma idade de uma pessoa “37” é interpretada diferentemente nos Estados Unidos e Coréia. Neste exemplo, para interpretar corretamente a semântica do valor “37”, nós devemos entender o contexto social da fonte dos dados. Entretanto, este tipo de heterogeneidade semântica não é limitada a diferenças culturais. Até em um mesmo país, a forma como uma organização define sua semântica pode ser diferente de outra organização. (...) Portanto, múltiplas interpretações do mesmo dado fazem com que o esforço de integração seja extremamente difícil, uma vez que cada fonte de informação e receptor potencial desta informação podem operar em um contexto diferente, o que leva a uma heterogeneidade semântica de larga escala ⁴⁴.

⁴⁴ Let us assume that we have a data value “37.” What does “37” mean to us? This information may be useless unless we know its meaning. If “37” is stored in the temperature database, we can safely assume that “37” is about temperature. If it is stored in a person’s age attribute, then we may easily conclude that “37” is a person’s age. In this case, since we can identify its meaning based on the schema information from a database (i.e., the name of an attribute), we deal with “weak semantics.” Weak semantics can be identified based on structural, syntactic, and value/extensional information in databases [Sheth 1995]. Let us further examine the complication associated with the data value “37.” When we attempt to understand the data semantics, we should identify the context of the data being used. For instance, in many Asian countries, their age counting scheme is different from that of Western countries. In those Asian countries (e.g., Korea), on the day a baby is born, her age is one. Afterward, the baby’s age is incremented by 1 each New Year. So, if a baby is born on December 31, her age becomes two on the very next day. Therefore, the same person’s age “37” is interpreted in the United States differently than in Korea. In this example, to correctly interpret the actual semantics of the data value “37,” we must understand the social context of the source data. However, this type of

O que a interoperabilidade semântica pretende promover é a habilidade de resolver conflitos semânticos que surgem de diferentes perspectivas, significados implícitos e pressupostos, criando um ambiente informacional compatível baseado em conceitos consensuais, acordados entre as diversas unidades de negócio. Portanto, nesta visão, a interoperabilidade sintática dá a solução tecnológica, enquanto a interoperabilidade semântica cria soluções semióticas, lingüísticas, filosóficas e sociais.

3.3.2 Modelos de Informação em Medicina

A documentação médica é classicamente composta por documentos pouco estruturados, escritos à mão, e com conteúdo heterogêneo de acordo com a especialidade médica. Ao mesmo tempo, o prontuário médico (o conjunto de documentos clínicos que descreve o histórico de saúde de um determinado paciente) é uma ferramenta essencial do trabalho, não apenas para fins médico-legais de documentação, mas como estruturação de planos de cuidado e busca de informação para tomada de decisão. Inúmeras tentativas de estruturar o prontuário já foram feitas nos últimos 50 anos. Uma das primeiras foi o prontuário baseado em problemas, proposto por *Lawrence Weed* em 1969. Na proposta de *Weed*, cada problema de saúde do paciente deveria ser individualizado no processo de documentação e estruturado em quatro partes: Subjetivo, Objetivo, Avaliação e Plano (o acrônimo SOAP). Para os diversos usos do prontuário no meio eletrônico esta estruturação é insuficiente – é necessário estruturar os diagnósticos, medicamentos, exames laboratoriais, sintomas, etc., de acordo com as necessidades de uso da informação de cada sistema. Para solucionar estes problemas, várias associações criadoras de padrões propuseram estruturas, através de modelos de informação para compartilhamento.

semantic heterogeneity is not limited to cultural differences. Even within the same country, the way in which one organization defines data semantics could be different from the way another organization defines it. Different functional units in the same organization could interpret and use information differently. Hence, multiple interpretations of the same data make an integration effort extremely difficult since each source of information and potential receiver of that information may operate within a different context, which leads to large-scale semantic heterogeneity

Modelos de informação têm sido usados na informática em saúde há várias décadas. Eles podem ser utilizados em *templates* para entrada de dados do prontuário em prontuários médicos eletrônicos – seu objetivo é definir um conjunto de variáveis comuns para permitir a interoperabilidade semântica. A criação destes modelos é realizada através da definição de necessidades de comunicação e casos de uso práticos. Podem ser codificados vários tipos de informação, como afirmações contextuais sobre eventos ou entidades particulares. Na seção abaixo descreveremos três organizações de grande importância na promoção da interoperabilidade na saúde neste início de século XXI, as quais também são relevantes para os objetivos do presente trabalho (BLOBEL *et al.*, 2006): o *Health Level Seven International* (HL7), a Fundação *OpenEHR* e a *International Standards Organization* (ISO), especificamente o padrão ISO/CEN EN13606.

3.3.2.1 *Health Level Seven International* (HL7)

O HL7⁴⁵ é uma organização desenvolvedora de padrões – *Standards Developing Organizations* (SDO) – que opera no domínio da saúde, criando padrões de dados administrativos e clínicos. O nome é originário dos níveis de comunicação da *Open Systems Interconnection* – o nível sete é o nível aplicativo⁴⁶. A missão do HL7 é criar padrões para interoperabilidade que melhorem a prestação de cuidado de saúde, otimizem o fluxo de trabalho, reduzam a ambiguidade e melhorem a transferência de conhecimento entre os agentes envolvidos, incluindo os prestadores serviços de saúde, agências governamentais, agente comerciais, outras SDOs e pacientes.

Existem duas versões em uso para os padrões HL7: a versão 2 e versão 3. A versão 2 do padrão de mensagens do HL7 (Protocolo de Aplicação para Troca de Dados Eletrônicos em Ambientes de Saúde) é o padrão de troca de informações em saúde mais utilizado do mundo. Datado do final da década de 80, está atualmente na revisão v2.6. A

⁴⁵ <http://www.hl7.org>

⁴⁶ Os níveis são Físico (impulsos elétricos, luminosos ou rádio); Ligação de dados (codificação dos dados em bits); Rede (tecnologias de roteamento e switch); Transporte (controle do fluxo de dados); Sessão (estabelecimento, gestão e término de conexões entre aplicações); Apresentação (tradução entre formato de dados da rede para o formato da aplicação); Aplicação (a sintaxe dos dados é identificada, assim como as informações necessárias para a aplicação, como autenticação e privacidade).

versão 2 padroniza o conteúdo e tipos de dados de uma mensagem a partir de situações comuns na prestação de cuidados de saúde. Nestas mensagens, algumas seções são definidas e para cada seção são definidos elementos, a sequência em que devem aparecer os elementos, os tipos de dados, se são opcionais ou mandatórios.

Por exemplo, a admissão de um paciente em um hospital é comunicada através da mensagem ADT (admissão). Nesta mensagem, as três primeiras seções são MSH (que corresponde ao cabeçalho da mensagem), EVN (que corresponde ao tipo de evento) e PID (que corresponde à identificação do paciente). A seção PID é composta por vários elementos, dentre eles, o nome do paciente. O nome do paciente deve obrigatoriamente constar na mensagem, é o quinto elemento da mensagem, e deve ser representado com o tipo XPN, que contém os seguintes dados (entre parênteses os tipos de dados): |<nome da família (FN)> ^ <Primeiro nome (ST)> ^ <Segundo nome ou demais nomes próprios ou iniciais (ST)> ^ <sufixo (ex., JR or III) (ST)> ^ <prefixo (e.g., DR) (ST)> ^ <títulos (e.g., MD) (IS)> ^ <código do tipo de nome (ID)> ^ <código de representação do nome (ID)> ^ <conteúdo do nome (CE)> ^ <intervalo de validade do nome (DR)> ^ <ordem de montagem do nome (ID)>|. A versão 2 não pressupõe nenhuma linguagem específica, e as mensagens são normalmente trocadas como arquivos texto, podendo ser transmitidas em outros formatos, como XML.

A versão 2, apesar de seu sucesso na prática, foi alvo de várias críticas, como a falta de um modelo explícito, metodologia de desenvolvimento não documentada, falta de definição formal para o conteúdo e relacionamentos, ambiguidade e falta de rigor no uso de vocabulários controlados. A versão 3 (HL7 v3) pretende sanar os problemas enfatizando a redução de ambiguidade através de três princípios: metodologia formal de desenvolvimento orientada a objetos; ênfase no uso de vocabulários controlados; transporte padronizado via *eXtended Mark-up Language* (XML) (KAMINKER; CAMPOS, 2008).

A documentação do HL7 v3 é composta por vários artefatos estáticos e dinâmicos (*storyboards* – os quais relatam uma história em que o padrão é usado – diagramas de interação e atividade). Dentre os artefatos estáticos estão os vocabulários e os modelos de informação, compostos por um modelo de referência, o *Reference Information Model* (RIM), o *Domain Message Information Model* (DMIM) e o *Refined Message Information Model* (RMIM). Em cada modelo, são detalhadamente definidos os metadados de cada objeto, os seus tipos de dados e as restrições.

Um tipo especial de padrão do HL7 v3 é o CDA (*Clinical Document Architecture*) (DOLIN *et al.*, 2006; FERANTI *et al.*, 2006). O CDA é uma representação comum para documentos clínicos através de um padrão de marcação do documento que especifica a estrutura e semântica. O documento é definido por uma estrutura única, e pode conter texto, imagens, sons e outros conteúdos multimídia. Todos os documentos CDA possuem persistência, autor, completude, podem ser lidos por pessoas e podem ser autenticados. O CDA é implementado em XML.

3.3.2.2 ISO/CEN 13606

A norma ISO/CEN (Organização Internacional de Padronização / Comitê Europeu de Padronização) 13606⁴⁷ é um padrão europeu que define uma arquitetura de informação mínima para transmissão o Registro Eletrônico em Saúde (RES) de um cidadão. ISO/CEN 13606 é um padrão que descreve a estrutura de informação para representar as informações de forma a suportar a interoperabilidade de RES e sistemas de informação clínica, sejam *middleware* ou sistemas de transferência, adição, modificação de informações de saúde, ou ainda sistemas de troca de mensagens (KALRA, 2006). A norma é baseada nos requisitos ISO 18308 (Requisitos para um Registro Eletrônico em Saúde) e garante a interoperabilidade entre sistemas heterogêneos e legados. O objetivo da norma, assim como o HL7 v3, é garantir a interoperabilidade na comunicação de RES, como apontado na introdução da norma:

“O objetivo geral deste padrão é definir uma arquitetura de informação rigorosa e estável para comunicar parte ou o todo de um Registro Eletrônico em Saúde de um único sujeito de cuidado (paciente). Isto serve para apoiar a interoperabilidade de sistemas e componentes necessários para comunicar (acesso, transferência, adição ou modificação) os dados do RES via mensagens eletrônicas ou objetos distribuídos:

Preservando o significado clínico original pretendido pelo autor,

⁴⁷ <http://www.en13606.org/>

Refletindo a confidencialidade dos dados como pretendido pelo autor e paciente”(CEN/ISO EN13606 INFORMATION SITE, 2009)

O padrão é composto por cinco partes: Modelo de Referência, Especificação de Troca de Arquétipos, Arquétipos Referência, e Lista de Termos, Segurança e Modelos de Troca. O Modelo de Referência define os blocos de construção do RES, que são Pastas, Composições, Seções, Entradas, Clusters, Elementos e Valores de Dados. Pode também incluir informação contextual adicional, como autor e sujeito do registro, informação temporal, descrição dos ambiente clínico, passos do raciocínio clínico, grau de confiança, achados negativos, mudanças de opinião, e indicações que os dados foram modificados. O padrão está em uso ativo por várias organizações nacionais, como o governo Sueco (SWEDISH ASSOCIATION OF LOCAL AUTHORITIES AND REGIONS, 2010) e pelo Estado de Minas Gerais (SANTOS *et al.*, 2012).

3.3.2.3 OpenEHR

O OpenEHR⁴⁸ é uma especificação de uma plataforma aberta, detalhada e testada, que permite a interoperabilidade para troca de informações em saúde. O padrão é mantido pela *OpenEHR Foundation*, uma organização sem fins lucrativos fundada pela *University College of London* (Reino Unido) e pela *Ocean Informatics Pty* (Australia). A abordagem do OpenEHR é baseada nos resultados do projeto GEHR (*Good European Health Record*), posteriormente refinada em projetos europeus e australianos.

A principal característica do OpenEHR é sua modelagem em dois níveis, herdada da ISO/CEN EN 13606 (GARDE *et al.*, 2007). O primeiro nível é um modelo de referência, que define classes para cumprir requisitos médico-legais e funções de gestão de registros. O segundo nível consiste na metodologia dos arquétipos, o que garante a troca de informações clínicas interpretáveis (interoperabilidade semântica). A diferença entre a ISO/CEN 13606 e o OpenEHR é o nível de detalhamento. O OpenEHR especifica as

⁴⁸ <http://www.openehr.org/home.html>

definições clínicas de forma a diminuir o número de itens opcionais e garantir que as mensagens sejam compreendidas por dois sistemas sem exigir nenhum tipo de ajuste (SCHLOEFFEL *et al.*, 2006).

A relação entre os dois níveis no OpenEHR é feita explícita através de um modelo denominado por seus autores de Clinical Investigator Record (CIR)⁴⁹ (BEALE; HEARD, 2007b). Segundo os autores, o objetivo da CIR é reconhecer, identificar e modelar os tipos de informação que podem ser criados para registro e uso em sistemas de informação de saúde, e é a base para a modelagem das entradas clínicas no OpenEHR. São reconhecidos 5 tipos de informação:

- Observação: informação criada pelo ato de observação, mensuração, questionamento ou teste do paciente ou substância relacionada;
- Opinião: inferências do investigador sobre o significado das observações; inclui diagnósticos, avaliações, planos e objetivos;
- Instrução: instruções baseadas em opinião de forma suficientemente detalhadas para que possam ser executadas e cumpram uma determinada intervenção;
- Ação: registro de ações intervencionistas que já ocorreram, seja por instruções ou outros;
- Evento Administrativo: registro de um evento de negócio ocorrendo em um contexto administrativo; ex. Admissão, marcação, encaminhamento, alta, etc.

O modelo é organizado taxonomicamente e cada nível reflete um critério de categorização distinto (Figura 5). O primeiro nível divide a informação de cuidado ao paciente no passado, presente e futuro. Desta forma, História reúne a informação sobre eventos e estados passados, a Opinião reúne as impressões do autor sobre as condições

⁴⁹ Conforme discutido na seção 3.1.1, um grande número de modelos é correntemente denominado ontologia. Porém, uma vez que os preceitos de categorização ontológica – axiomatização, fundamentação filosófica na realidade - não foram seguidos, a CIR será referida como um modelo.

atuais do paciente, e Instrução reúne informações sobre as ações que serão realizadas no futuro. O segundo nível diferencia observações (branco) e ações (cinza). Por fim, são identificados sub-tipos de acordo com a forma como a opinião do autor é formada. Por exemplo, Diagnóstico identifica a patologia do paciente, enquanto Prognóstico refere-se a uma opinião sobre o curso futuro desta patologia que pode ou não se concretizar.

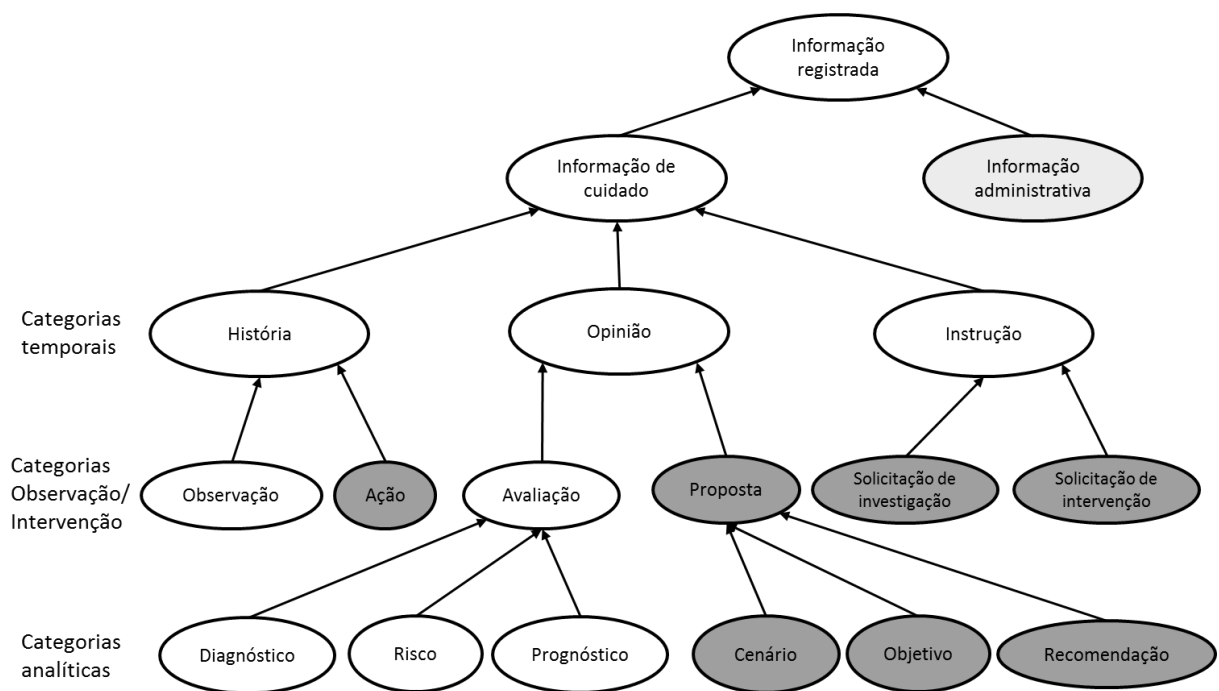


Figura 5 – A ontologia Clinical Investigator Record (CIR) – adaptada de (BEALE; HEARD, 2007b)

A CIR demonstra a conexão entre o modelo de referência (informação registrada) e a informação sobre o paciente. Podem ser vistos no modelo os cinco tipos possíveis de entradas de informação sobre o paciente de acordo com o OpenEHR: Informação administrativa, Observação, Ação, Opinião e Instrução.

Os arquétipos formam a grande distinção entre os padrões ISO/CEN 13606/OpenEHR e o HL7 CDA. Arquétipos são modelos clínicos consensuais ou conceitos específicos de domínio, ou "definições formais de uma combinação prescrita das classes fundamentais definidas no Modelo de Referência para um domínio clínico particular ou uma

organização"⁵⁰ (KALRA, 2006, p.138). Arquétipos definem informações de grande relevância clínica, regras de negócio específicas para a saúde e permitem que a informação seja estruturada de forma complexa. A estruturação da informação é feita através da escolha do melhor termo para descrever o conceito relevante, considerando que o contexto clínico é completamente especificado. Por exemplo, o arquétipo “pressão arterial” (observação) define o que é “pressão sistólica” (pico sistêmico da pressão sanguínea arterial – medida na fase sistólica ou de contração do ciclo cardíaco), define o tipo de dado para esta informação (mmHg), o estado do paciente no momento da medida, como posição (sentado, em pé), informações sobre o protocolo da medida (tamanho do manguito) e o evento em que a medição foi feita. Todas estas informações se referem a uma única medição, formando um objeto complexo clinicamente relevante. A Figura 6 demonstra o arquétipo visto pelo Clinical Knowledge Manager (OPENEHR FOUNDATION, 2011)

⁵⁰ formal definitions of prescribed combinations of the building block classes defined in the Reference Model for particular clinical domains or organizations

Symptom

English

Archetype: Symptom (openEHR-EHR-CLUSTER.symptom.v1)

Header Items

Structure: Cluster
 Occurrences: 1..1 (mandatory)
 Cardinality: 1..* (mandatory, repeating, unordered)

T Text	Symptom The symptom experienced. Comment: Coding with an external terminology is preferred, where possible.	Free or coded text
X Boolean	Nil Significant The person has not had any significant experience of the symptom.	Allowed values: {true}
T Text	Clinical Description Description of the symptom.	Free or coded text
A Slot (Cluster)	Location in Body Details of the location in the body.	Include: openEHR-EHR-CLUSTER.anatomical_location.v1 and specialisations Exclude: All not explicitly included archetypes
≡ Ordinal	Severity The severity of the symptom.	1: Trivial [The symptom is trivial and causes no problems.] [SNOMED-CT::162466003] (Symptom trivial (finding)) 2: Mild [The symptom does not interfere greatly with day to day activities.] [SNOMED-CT::162468002] (Symptom mild (finding)) 5: Moderate [The symptom interferes with day to day activities.] [SNOMED-CT::162469005] (Symptom moderate (finding)) 8: Severe [The symptom prohibits some key daily activities.] [SNOMED-CT::162470006] (Symptom severe (finding))

Previous Next

Figura 6 – O arquétipo Sintoma

Os arquétipos do OpenEHR são escritos em uma linguagem própria (ADL – *archetype definition language*) e permitem que clínicos criem novos arquétipos para que se tornem padrões acordados. Arquétipos também permitem a validação dos dados no momento da entrada de dados, alertando sobre erros de acordo com as regras clínicas. Eles expressam as informações ao indicar (HOVENGA *et al.*, 2005):

- Como a informação deve ser representada;
- O que é opcional e o que é mandatório;
- O que é um valor possível para um elemento;
- Quaisquer outras regras que precisam ser representadas.

Todos os três padrões pretendem apoiar a interoperabilidade semântica na saúde, com abordagens ligeiramente diferentes. Essas abordagens criam grandes diferenças na especificação de cada modelo, e um trabalho considerável na harmonização (mapeamento) entre os padrões. Na FIG.5 demonstra-se a comparação entre os padrões.

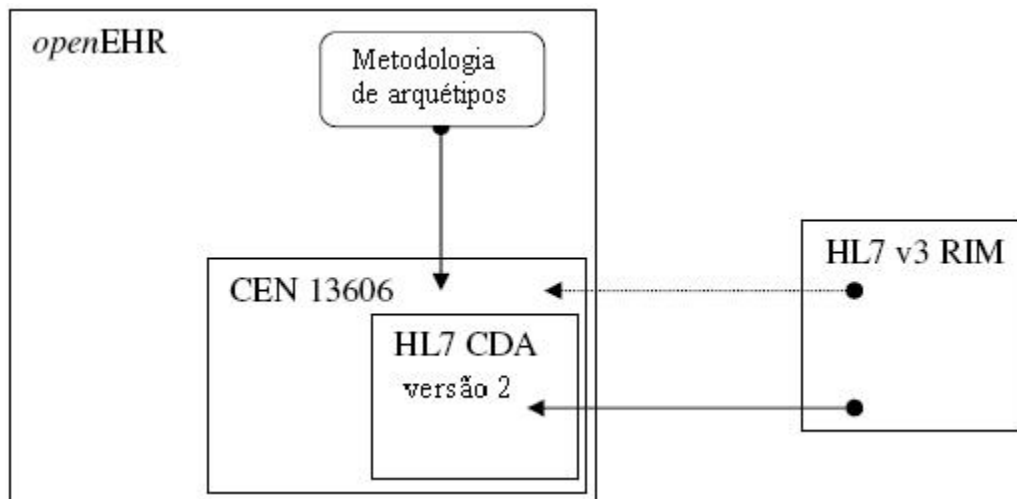


Figura 7 – Relação esquemática entre o openEHR, CEN 13606 e HL7 CDA

FONTE: (SCHLOEFFEL *et al.*, 2006)

3.3.2.4 Ontologias e modelos de informação

Como discutido acima, existe um conjunto considerável de trabalhos que apoiam o desenvolvimento e avaliação de modelos de informação. No entanto, pesquisas que misturam essas duas abordagens são relativamente incomuns, uma vez que ambas as comunidades têm objetivos e origens muito diferentes. Modelos de informação apresentam uma visão muito pragmática, enquanto a comunidade centrada em ontologias tem seu foco orientado para a investigação científica. É um princípio fundamental de modelos de informação de que os dados sobre paciente reais sejam representados utilizando terminologias médicas. O método exato para o desenvolvimento desta terminologia está fora do escopo, mas a conexão entre terminologia e modelo de informação é central para determinar o significado da mensagem. Esta questão refere-se à ligação formal entre terminologias e registros eletrônicos de saúde e é relevante para esta pesquisa. Rector e outros (RECTOR *et al.*, 2009) definem alguns requisitos para tal ligação, como:

1. Deve existir uma interface clara entre o modelo de códigos e o modelo de informação, uma "Interface de Ligação de Códigos" (Code Binding Interface).
2. A ligação deve ter expressividade suficiente para capturar (a) lista de códigos enumerados; (b) todos os subcódigos de um dado código (com ou sem a raiz); (c) todas as combinações booleanas de (a) e (b).
3. Um acordo para ligação com expressões pré e pós-coordenadas contidas no SNOMED CT ou outros sistemas de códigos posicionais.
4. Restrições mútuas entre informação e modelos de codificação deve ser explícita e testável.
5. As restrições entre informação e modelos de codificação deve ser parte de uma metodologia coerente para expressar as restrições do modelo de informação como um todo.
6. Os modelos e interfaces devem ser expressas em linguagens padronizadas com semântica bem definida e ferramentas disponíveis.

Esses autores propõem uma estrutura sofisticada para a ligação do modelo de códigos ("código para o diabetes") e do modelo de estruturas de dados ("Diagnóstico"). Nesta abordagem, o foco é sobre a validade da informação - as restrições que garantam que as estruturas de dados são válidas, isto é, a informação digitada tem o tipo de dados e escopo corretos - em vez das afirmações ontológicas que estão sendo feitas, as quais eles chamam de precisão. Por exemplo, uma restrição pode ser feita em um campo que requer preenchimento por um "código de diabetes". O significado do que é "diabetes" não seria abrangido pela abordagem, juntamente com a exata natureza da relação entre diabetes e o paciente examinado.

Martínez-Costa e outros (MARTINEZ-COSTA *et al.*, 2010; COSTA *et al.*, 2011) oferecem um ponto de vista diferente sobre o problema. Eles desenvolveram um método para a conversão do modelo de referência do openEHR e normas ISO 13606 em um modelo de referência comum expresso em OWL. Enquanto eles alcançaram a interoperabilidade entre os dois padrões, esta metodologia não abrange a real relação entre a informação (entrada, elemento de dados) e a entidade real que a informação denota (diabetes).

Santos (SANTOS *et al.*, 2010) (SANTOS, 2011), em sua tese, argumenta que as ontologias são uma parte essencial do desenvolvimento de arquétipos. Ele mostra a viabilidade do uso de arquétipos de um sistema regional de saúde e faz uma analogia entre ontologias e arquétipos - no sentido de que ambos fazem uso de um conjunto de regras

para restringir a interpretação. Enquanto a analogia é, em princípio, correta, há importantes diferenças adicionais que são relevantes para o presente trabalho, particularmente quando se estuda a ligação entre os modelos e ontologias. O primeiro se refere ao paradigma de modelagem adotada aqui (realismo ontológico), que faz uma distinção explícita entre entidades de conteúdo de informação e as entidades reais que as primeiras denotam. Por exemplo, o diagnóstico sobre a diabetes e diabetes em si. A segunda se refere ao tipo de regras, já que modelos de informação são baseados em restrições processuais e persistência em banco de dados relacional, enquanto ontologias são baseadas em definições lógicas declarativas e um conjunto de afirmações. Finalmente, estamos mais preocupados aqui com a interpretação da mensagem completa após sua construção (precisão, de acordo com Rector) do que com os métodos de validação para garantir que o médico registrou a informação certa no lugar certo.

3.3.3 Linguagem e Conhecimento

Ontologia (no sentido filosófico) não trata de linguagem, mas da natureza do ser e da existência. Ontologias (no sentido de artefatos computacionais) são vistas na academia e no meio empresarial como artefatos heterogêneos que podem representar significados linguísticos (um tipo especial de terminologia) e a representação de conhecimento (sendo um modelo do mundo em alguma perspectiva). Por um lado, define o sentido de um termo, seja através de axiomas formais e relacionamentos entre os termos, ou através da listagem extensional daquilo que o termo representa (definição ostensiva). Por outro lado, descreve a teoria do mundo, ou pelo menos uma teoria do mundo. Ao mesmo tempo, ontologias não são uma linguagem, e representam o mundo, e não o conhecimento. No presente trabalho, as ontologias são vistas como a interseção entre o conhecimento e a linguagem, para permitir comunicação, raciocínio e consistência.

A relação entre ontologia e linguagem pode ser evidenciada pela passagem de Gruber:

“Pragmaticamente, uma ontologia comum define o vocabulário com o qual as questões e afirmativas são trocadas entre

agentes. Comprometimentos ontológicos são acordos para usar um vocabulário compartilhado de uma forma consistente e coerente. Os agentes que compartilham um vocabulário não precisam de compartilhar uma base de conhecimento; cada um sabe coisas que o outro não sabe, e um agente que se compromete a uma ontologia não precisa responder a todas as questões que podem ser formuladas em um vocabulário compartilhado⁵¹.” (GRUBER, 1995, p.2)

Porém, não há semântica sem um relacionamento entre o símbolo e o mundo (teoria proposicional) ou entre os símbolos de uma linguagem comum e os símbolos de uma linguagem própria. Como garantir que os agentes interpretem corretamente um termo se o conhecimento **sobre a semântica do termo** não é comum? Almeida apresenta uma resposta:

“Finalmente, retornando à questão que originou essa discussão – a relação entre as ontologias e as linguagens – pode-se concluir (...) que uma ontologia é capaz de representar uma linguagem, na forma de uma linguagem natural controlada, pois proporciona: (a) a representação das relações entre símbolos; (b) a representação das relações entre os símbolos e as entidades presentes no mundo, expressando o significado desses símbolos; (c) a representação do contexto de utilização dos símbolos, ao explicitar os propósitos e as intenções dos agentes que utilizam os símbolos para se referir a objetos” (ALMEIDA, 2006, p.247)

Não se pretende argumentar que toda a semântica exigida para a “compreensão” de um termo esteja contida em uma ontologia. Estudos cognitivos fornecem fortes evidências

⁵¹ “Pragmatically, a common ontology defines the vocabulary with which queries and assertions are exchanged among agents. Ontological commitments are agreements to use the shared vocabulary in a coherent and consistent manner. The agents sharing a vocabulary need not share a knowledge base; each knows things the other does not, and an agent that commits to an ontology is not required to answer all queries that can be formulated in the shared vocabulary.”

que a compreensão de uma afirmativa depende de conhecimento da vida real, como conhecimento sobre eventos e hábitos comuns na sociedade, como a associação entre médico e hospital, comprador e vendedor, e celeiro e feno (HARE *et al.*, 2009). As ontologias têm apenas a capacidade de usar algumas relações formalizadas por axiomas. Além disso, ontologias conceituais podem criar associações inconsistentes e, portanto, o conhecimento expresso deve ser restrito (no presente estudo, a restrição será a realidade). Ainda assim, a capacidade de inferência nas ontologias permite respostas complexas e garante que a comunicação seja efetiva mesmo quando as regras diferem entre os agentes. Por exemplo, uma pergunta sobre o uso de algum medicamento por determinado paciente requer que ambos os agentes compartilhem o mesmo significado de “medicamento”, e que saibam que “Ácido acetil-salicílico” é um medicamento, independente do comportamento do agente ao registrar que o paciente usa AAS (alerta de preço, por exemplo).

Nesta visão, os modelos de informação carregam toda a semântica nas definições intensionais de seus termos e classes. No exemplo do arquétipo de Pressão Arterial (observação), define-se o termo “Localização” como “o lugar do corpo onde a pressão sanguínea é mensurada”⁵². A interpretação correta depende das regras de interpretação embutidas no modelo, e principalmente da adesão destas regras por parte do desenvolvedor. De fato, todas as ontologias, apesar das pretensões ao contrário, usam a linguagem humana para “representar” o mundo. Alguns elementos funcionais menores podem ser representados em alguma forma de lógica, mas a carga interpretativa mais importante é sempre realizada pela linguagem (BREWSTER; O'HARA, 2007). Ainda que as linguagens de representação quase sempre tenham íntima relação com linguagens naturais, uma vez que a criação e manutenção destas bases são feitas por pessoas (NIRENBURG; WILKS, 2001), o relacionamento formal das ontologias permite inferências lógicas determinadas pelo domínio de conhecimento.

A busca por uma representação ontológica da informação representada nos modelos de informação é baseada em duas premissas:

⁵² Disponível em: <http://www.openehr.org/knowledge/>, acessado em 05/03/2011

1. a abordagem utilizada no desenvolvimento de modelos de informação ocasiona falhas na representação dos modelos devido à falta de princípios filosóficos robustos (SMITH; CEUSTERS, 2006)
2. a representação em linguagem lógica permite a manipulação ordenada de classes e respostas adequadas e abrangentes para as questões colocadas por outros agentes, o que permite uma forma mais completa de interoperabilidade semântica (ainda que incompleta)

Concluindo, partiremos do princípio que ontologias permitem a interpretação de um termo ao mapear o termo à sua linguagem própria e categorizar o termo usando suas relações internas, ou seja, com outros termos. A linguagem própria deve ser independente da linguagem (enquanto possível) e deve seguir uma metodologia adequada (no caso, o realismo). A questão estudada será a capacidade de expressar termos não realistas mantendo a coerência interna da representação.

4 METODOLOGIA

4.1 CONSIDERAÇÕES INICIAIS

Pretende-se aqui realizar uma análise detalhada sobre a forma como a informação médica é representada em modelos de informação e em ontologias biomédicas, para avaliar as possibilidades e limitações da representação ontológica de dados do prontuário. Apresentam-se nessa seção considerações sobre a natureza da pesquisa, seus objetivos, sobre o tipo de abordagem ao problema e sobre os procedimentos técnicos. Além disso, são apresentadas para maior clareza algumas considerações teóricas não citadas na revisão de literatura, que são relevantes para o entendimento da metodologia. Considerações sobre a influência da tecnologia no percurso metodológico são também brevemente descritas;

A presente pesquisa é considerada, de acordo com Yin (1994) e Gil (1991), uma **pesquisa aplicada**, visto que tem como objetivo criar conhecimento para aplicação prática e orientado à solução de problemas específicos. A metodologia foi concebida para fornecer dados práticos para desenvolvedores e modeladores de sistemas clínicos, considerando o estado atual dos padrões de troca de informação em medicina e atual estado de desenvolvimento de ontologias biomédicas.

Em relação a seus objetivos, a **pesquisa é exploratória**, visto que tem como objetivo explorar a relação entre modelos de informação em saúde e ontologias biomédicas. Artefatos informatizados classificados nestas duas categorias possuem critérios de modelagem distintos. A presente pesquisa explora as diferenças e semelhanças entre estes critérios e o impacto destas na construção de sistemas informatizados. A análise destas diferenças e semelhanças foi facilitado pela formação do autor, graduado em Medicina e com larga experiência profissional no desenvolvimento, aquisição e uso de sistemas de informação em saúde.

A abordagem realizada utilizará elementos qualitativos e quantitativos, com clara predominância de elementos qualitativos. Os objetos de estudo são modelos de informação e ontologias biomédicas. Ambos são modelos criados por pessoas e concretizados em artefatos informatizados. Ainda que existam critérios claros para classificar tais modelos em

uma ou outra categoria, existe grande heterogeneidade entre modelos de uma mesma categoria, como demonstrado na figura 1 da seção 3.1.1. Portanto, é impossível excluir completamente a subjetividade ao analisar os modelos. Porém, existe aqui o exposto objetivo de corroborar os achados subjetivos com indícios objetivos, que tomam duas formas principais:

1. Provas ou testes utilizando linguagens lógicas, garantindo a consistência das premissas
2. Descrições quantitativas dos objetos estudados

Ontologias possuem objetivos e usos diferentes considerando diferentes comunidades. Mesmo se considerarmos o escopo de ontologias para sistemas de informação, ontologias podem ser utilizadas para processamento de linguagem natural (SIMON *et al.*, 2006), recuperação de informação - seja para indexação (NOY *et al.*, 2009) e para criação e extensão dos termos de busca (BHOGAL *et al.*, 2007) – e integração de bases de dados ou mensagens (KÖHLER *et al.*, 2003). Casos de uso também podem ser respondidos com o uso exclusivo de ontologias ou em conjunto com outros artefatos de representação de informação. A seleção dos requisitos para avaliação da metodologia de representação da informação médica proposta foi baseada na literatura já extensa sobre registros eletrônicos de saúde, web semântica e modelos de informação. No entanto, o número de casos de uso possíveis para ontologias em saúde é enorme, e não é possível responder a todos eles no âmbito desta tese. A maioria dos critérios escolhidos são relativos às vantagens esperadas do uso de uma representação lógica / ontológica de termos, principalmente a organização de informações para uso posterior. Foram exploradas questões técnicas (computabilidade) de forma superficial que, embora essencial para a implementação, estão fora do escopo deste trabalho. A seguir, são apresentadas as razões para escolha de cada requisito, as quais levaram à formulação de questões com as quais será testada a abordagem resultante.

A análise de requisitos, sejam requisitos funcionais ou técnicos, é uma parte essencial do processo de engenharia de software. No decorrer da pesquisa, observou-se que vários critérios para avaliação de ontologias estão mais relacionadas com a perspectiva de engenharia do que a perspectiva científica. Por exemplo, Almeida (2009) descreve quatro tipos de abordagem para a avaliação de ontologias (Tabela 1). Seguindo esta tipologia, pretende-se avaliar a ontologia e modelo de informação de acordo com a sua capacidade de representar as entidades de registros de saúde (segundo a tabela, comparação com fontes

de dados) e a capacidade de satisfazer os critérios, especialmente aqueles para a classificação automática e tratamento da informação (conformidade com os requisitos). No entanto, sustenta-se aqui que a avaliação final pragmática será a implementação real desses artefatos em sistemas de informação reais de trabalho, que serão alvo de pesquisas futuras. A avaliação será a mais objetiva possível, de forma a garantir a extensibilidade dos resultados e generalização dos resultados.

Tipo de abordagem	Comparação com um padrão	Avaliação do uso em uma aplicação	Comparação com fontes de dados	Conformidade com requisitos
Descrição	Comparação sintática entre uma ontologia e um padrão, que pode ser outra ontologia	Uso de uma ontologia em uma aplicação, seguida da avaliação de resultados	Comparação com uma fonte de dados coberto pela ontologia	Avaliação conduzida por pessoas que pretendem verificar a aderência de uma ontologia a critérios e padrões

Tabela 1 - Tipo de abordagem para avaliação de ontologias (ALMEIDA, 2009)

Várias tentativas de utilizar princípios ontológicos para o desenvolvimento das terminologias médicas e aplicações foram feitas até hoje. A abordagem do *Referent Tracking* baseia-se na premissa de que o sistema de informação em saúde ideal irá controlar cada entidade na realidade relacionada com o estado clínico do paciente no contínuo de saúde (CEUSTERS *et al.*, 2005; CEUSTERS; SMITH, 2006). Dentro desta abordagem, o aspecto essencial é dar um identificador universal a essas instâncias para que seja possível saber exatamente quando alguma doença específica começou a existir, de tal forma que, se um diagnóstico foi feito e alterado posteriormente, o identificador doença permaneça o mesmo. Não vamos considerar esse requisito como de suma importância, devido à incompatibilidade com a abordagem de modelo de informação, que é orientada a documentos, em vez de orientada a entidades da realidade.

Outras iniciativas (RECTOR; ROGERS, 2006; SCHULZ; KLEIN, 2008; STENZHORN *et al.*, 2008; SCHULZ *et al.*, 2010b) tentaram restringir o escopo de ontologias em sistemas

de saúde. Rector sugere quatro usos principais para ontologias em sistemas de informação em saúde (RECTOR, 2010, p.5):

1. Especificar as entidades e relações sobre as quais existe informação a ser transmitida;
2. Organizar as entidades e descrições em estruturas "tipo-de";
3. Representar as características universais destas entidades e relações – aquelas coisas que são verdadeiras por definição ou necessidade;
4. Especificar o que pode, ou não pode, ser dito sobre estas entidades ou relações.

O único uso que não será considerado é o uso número quatro, que lida com restrições no nível de entrada (quais valores podem ser usados por um médico que preenche um campo do sistema). Além disso, é importante ressaltar que ontologias são claramente diferenciadas de sistemas de representação de conhecimento neste trabalho. A melhor definição pode ser vista no trecho abaixo:

"Representação de conhecimento – que deve ser mais adequadamente referenciada como a modelagem de crenças entre cientistas – não é função de ontologias formais. Também não é o caso que ontologias formais descrevam entidades que pertençam ao domínio da linguagem humana. Estes dois tipos de artefatos representacionais representam coisas diferentes, servem a propósitos diferentes e usam diferentes formalismos" (SCHULZ et al., 2009b, p.42)⁵³

Quanto aos **procedimentos técnicos**, a pesquisa pode ser classificada em duas categorias. A primeira etapa, que faz a análise dos paradigmas de modelagem e propõe uma forma de co-existência entre os mesmos, pode ser classificada como pesquisa bibliográfica. A segunda etapa pode ser classificada como um estudo de caso, visto que pretende testar os achados bibliográficos em uma simulação prática utilizando artefatos criados como modelos de informação e ontologias disponíveis em duas bases de dados – o *OpenEHR Clinical Knowledge Manager* e o *Open Biomedical Ontologies Library (OBO)*.

⁵³ "(K)nowledge representation – which might more properly be referred to as the modeling of beliefs among scientists – is not a task of formal ontologies. Nor do formal ontologies describe entities properly belonging to the domain of human language. These two kinds of representational artifact represent different things, serve different purposes and use different formalisms."

Um outro aspecto, que não diz respeito diretamente à metodologia de pesquisa científica, mas que influencia a pesquisa é a questão da tecnologia utilizada, considerando-se que mais de uma possibilidade foi avaliada. O uso de ontologias como artefato de sistemas de informação é bastante heterogêneo, abrangendo desde uso para organização hierárquica de diferentes conceitos até representações logicamente complexas escritas em lógica de primeira ordem (ver seção 3.1.2). Considerando as hipóteses e questões levantadas na seção 4.3.4, a tecnologia utilizada para representação deve contemplar:

1. Capacidade de inferência
2. Capacidade de recuperação de dados individuais

Estes requisitos estão intimamente interrelacionadas. A primeira questão diz respeito à forma de interpretar o que foi representado pelo médico. Considerando o fragmento "o paciente relata dor no peito", pode-se representar esta informação através de pelo menos duas formas distintas:

- Sintoma: "Dor Torácica" (classe)
- Sintoma: "A dor no peito do paciente José da Silva" (instância) *Type* "Dor Torácica" (classe)

O segundo requisito diz respeito à formulação de *queries* para busca da informação armazenada. Nos sistemas baseados em bancos de dados convencionais, a busca pela informação segue o princípio do mundo fechado, ou seja, a resposta é negativa a menos que se prove o contrário. Por exemplo, se não existe menção de "Dor Torácica" para um determinado paciente, este é considerado um paciente que não teve dor torácica. Nos sistemas baseados em lógica descritiva, a inferência segue o princípio do mundo aberto, ou seja, a resposta é inconclusiva e menos que se prove o contrário. Por exemplo, se existe uma afirmação que o sintoma é do tipo "Dor Torácica", o mesmo sintoma pode ser do tipo "Tosse", a menos que algum axioma explicitamente restrinja esta possibilidade.

É importante ressaltar que a tecnologia utilizada é particularmente dependente dos requisitos de cada sistema. Algumas soluções estudadas são:

- RDBMS
- XML, RDF, OWL
- Representações de lógica de primeira ordem

Os sistemas tradicionais de bancos de dados são a escolha natural para armazenamento de quantidades significativas de dados, devido à velocidade de recuperação de informação, robustez do banco de dados (resistência a erros e perdas) e existência de linguagens e ferramentas amplamente distribuídas. A busca nos bancos de dados é realizada através de uma linguagem de busca (*Structured Query Language*, ou SQL). O SQL baseia-se no princípio de mundo fechado. Por outro lado, a realização de inferências em lógica descritiva não é nativa e deve ser controlada externamente durante a criação do comando. Já os sistemas de bancos de dados requerem ainda a criação prévia de um esquema de dados. Ainda que este esquema possa ser tão genérico quanto a representação de sujeito, predicado e objeto, cria um esforço inicial que desencorajou o uso no presente estudo. A velocidade de recuperação não é um critério importante para o desenvolvimento desta tese – em situações reais, este critério frequentemente sobrepuja todos os outros.

O armazenamento de dados em documentos escritos em linguagens de marcação oferecem a vantagem de facilidade de manipulação, ausência de esquemas pré-definidos e relações taxonômicas. Dentre as linguagens de marcação avaliadas, a OWL2 apresenta vantagens para o estudo. Permite inferências em lógica descritiva nativamente, através de ferramentas de inferência automatizada (*reasoners*). Se adotado o RDF, essa linguagem permite asserções individuais na forma de instâncias (*A-Box*). A linguagem, quando processada pelas ferramentas de inferência, é decidível e completa, mas pode exigir muito tempo para realização de inferências. Para diminuir este tempo, foram desenvolvidos perfis da linguagem, permitindo expressões que simplificam ou aumentam a complexidade das inferências (ex. Impedir axiomas de disjunção, ou negação).

Adicionalmente, foram avaliadas representações de lógica de primeira ordem, especialmente devido à dificuldade na definição de alguns termos médicos. A falta de decidibilidade e o tempo exigido para a realização das inferências são fortes argumentos contrários ao uso destas representações.

Em virtude da existência de ferramentas gratuitas e livres para criação de ontologias e análise das inferências (Protégé 4.2 e seus plugins), neste trabalho optou-se por representar a ontologia e os fragmentos em OWL2.

A pesquisa consistiu em 3 etapas sequenciais, descritas na seção 4.3:

1. Construção do arcabouço teórico
2. Análise ontológica do modelo OpenEHR
3. Representação ontológica da informação médica

Parte da presente pesquisa foi realizada durante um estágio sanduíche na Universidade Médica de Graz – Instituto de Informática Médica, Estatística e Documentação, coordenado pela Profa.Dra. Andrea Berhold, e sob a orientação do prof. Stefan Schulz.

As seguintes seções definem o objeto de pesquisa, os métodos de seleção da amostra estudada, os passos de construção da metodologia e seu teste.

4.2 OBJETOS DE PESQUISA E SELEÇÃO DA AMOSTRA

Para realização da metodologia, foi necessário selecionar três tipos de objetos:

1. uma especificação de modelo de informação
2. um conjunto de ontologias ortogonais
3. sentenças representativas de documentação médica

Durante a pesquisa bibliográfica (seção 3), foram identificados vários artefatos representativos dos paradigmas de modelagem avaliados. Alguns destes são utilizados como modelos de alto-nível para orientar a criação de artefatos mais próximos à prática. Os mais relevantes modelos de alto nível para a presente pesquisa são:

- Modelos de informação: HL7 *Clinical Document Architecture*, ISO/CEN 13606 *Health informatics - Electronic health record communication*, OpenEHR *standard*
- Ontologias: DOLCE, CYC e OpenCYC, BFO

Foram utilizados os seguintes critérios na seleção dos artefatos:

- existência de documentação acessível
- fundamentação de modelagem ou filosófica robusta

- disponibilidade de artefatos de uso prático ou corrente
- base de usuários

Após análise cuidadosa, decidiu-se trabalhar com o OpenEHR *standard* como modelo de informação e com a BFO como ontologia de alto-nível. Os motivos são expostos a seguir.

O *OpenEHR standard* possui documentação ampla e acessível. É um modelo baseado no paradigma de modelagem em dois níveis - nível informacional e nível clínico - e possui disponibilidade de uma ampla base de dados contendo diversos arquétipos clínicos (artefatos de modelagem clínica). A base de dados é aberta e disponível em (OPENEHR FOUNDATION, 2011). Além disso, possui uma ampla comunidade ativa distribuída em vários continentes, dentre estes a América do Sul. O modelo é mencionado em documentos governamentais brasileiros (PADILHA, 2011) e possui membros em projetos internacionais orientados à definição de padrões de interoperabilidade (SEMANTIC HEALTH NET CONSORTIUM, 2012).

Assim como a BFO, a DOLCE possui ampla documentação publicada em conferências e periódicos especializados e disponibilidade de artefatos de uso prático (ontologias aplicacionais). Portanto, o principal fator para seleção da BFO como ontologia de alto nível é sua fundamentação filosófica. Soma-se a este critério uma base de usuários crescente na comunidade biomédica, o que favoreceu a realização de diversas reuniões informais com usuários para esclarecimento e discussão de diversos assuntos relacionados a modelagem e uso dos artefatos.

A BFO passou por uma grande revisão e em julho de 2012 foi lançada uma versão preliminar da BFO2 (SMITH *et al.*, 2012). Porém, como a versão atual é instável e provisoriamente incompatível com outras ontologias realistas, optou-se por utilizar a versão anterior da ontologia. Quando pertinente, relações e classes da BFO2 foram importadas para ampliar o escopo de representação das ontologias realistas. Quando citadas neste documento, a origem das classes e relações será explicitada.

Afim de fornecer base empírica para embasar a transformação entre modelos de informação e ontologias, foram extraídos 169 fragmentos sentenciais de prontuários previamente anonimizados, coletados durante a execução do Blood Project. O Blood Project é um projeto colaborativo entre o Ontology Research Group da State University of New York

at Buffalo, a Fundação Hemominas e a Escola de Ciência da Informação da Universidade Federal de Minas Gerais.

Os prontuários médicos, conforme definição do Conselho Federal de Medicina, são compostos pelas seguintes seções (CONSELHO FEDERAL DE MEDICINA, 2002):

a. Identificação do paciente – nome completo, data de nascimento (dia, mês e ano com quatro dígitos), sexo, nome da mãe, naturalidade (indicando o município e o estado de nascimento), endereço completo (nome da via pública, número, complemento, bairro/distrito, município, estado e CEP);

b. Anamnese, exame físico, exames complementares solicitados e seus respectivos resultados, hipóteses diagnósticas, diagnóstico definitivo e tratamento efetuado;

c. Evolução diária do paciente, com data e hora, discriminação de todos os procedimentos aos quais o mesmo foi submetido e identificação dos profissionais que os realizaram, assinados eletronicamente quando elaborados e/ou armazenados em meio eletrônico;

d. Nos prontuários em suporte de papel é obrigatória a legibilidade da letra do profissional que atendeu o paciente, bem como a identificação dos profissionais prestadores do atendimento. São também obrigatórias a assinatura e o respectivo número do CRM;

e. Nos casos emergenciais, nos quais seja impossível a colheita de história clínica do paciente, deverá constar relato médico completo de todos os procedimentos realizados e que tenham possibilitado o diagnóstico e/ou a remoção para outra unidade.

Os fragmentos sentenciais para análise nesta pesquisa foram aleatoriamente selecionados das seguintes seções: Anamnese (subdividido em História da Doença Atual, Revisão de Sistemas e História Social); Exame Físico; e Hipóteses Diagnósticas, Diagnóstico Definitivo. Além destas seções, foram extraídas as frases que descrevem medicação prescrita e administrada. Estas seções foram selecionadas por sua cobertura de partes normalmente não-estruturadas do prontuário médico e pela ausência de

padronização consensual⁵⁴. Alguns exemplos podem ser vistos no quadro 2 e a origem dos fragmentos pode ser vista na tabela 2.

Dispneia e dor em mmii aos esforços moderados de cerca de 1 mês de evolução, com piora nos últimos dias.
Nega episódio de fezes desde ontem.
Mantenho anti-coagulação com clexane 60mg BID.

Quadro 2 – Exemplos de fragmentos sentenciais de prontuários anonimizados

Seção do prontuário a que se refere a sentença	Fragmentos sentenciais
História da Doença Atual	33
Revisão de Sistemas	18
Exame Físico	39
História Social	7
Eventos de Medicação	14
Avaliação	58
TOTAL	169

Tabela 2 – Seções originais dos fragmentos sentenciais selecionados

Uma vez que o processamento dos fragmentos – análise, representação no modelo de informação e representação ontológica – foi manual, o número de fragmentos foi

⁵⁴ Em oposição, existem padrões *de facto* para imagem (DICOM) e padrões locais para dados demográficos e laboratoriais. Estas seções foram, portanto, intencionalmente ignoradas.

determinado pelas limitações temporais inerentes a produção de uma tese e não reflete cálculo de amostra.

4.3 DETALHAMENTO DAS ETAPAS DA PESQUISA

Alguns pressupostos discutidos na revisão de literatura foram considerados para o desenvolvimento da metodologia de pesquisa:

- O prontuário médico é composto por um conjunto de afirmativas sobre o estado do mundo, sobre o modelo mental do profissional de saúde autor do documento e sobre planos para o futuro;
- Modelos de informação permitem a livre criação de conceitos e, portanto, permitem a expressão de quaisquer conjuntos de afirmações;
- Modelos de informação clínicos, devido ao seu histórico, são compostos por conceitos comuns em RES. Serão aqui considerados como fonte de melhores práticas na estruturação de RES;
- Ontologias construídas segundo a realismo ontológico são modelos robustos e escaláveis. Porém, o realismo ontológico impõe limites rígidos à representação de entidades sem existência na realidade física/material/referencial.

Considerando os pressupostos, pretendeu-se, neste pesquisa, utilizar o conhecimento médico embutido em modelos de informação para reconhecer e classificar as limitações de representação dos prontuários médicos por ontologias realistas. Podemos dividir a pesquisa em 3 etapas:

1. Construção do arcabouço teórico
2. Análise ontológica do modelo OpenEHR
3. Representação ontológica da informação médica

Cada uma das etapas será apresentada em uma seção distinta. Em seguida, a metodologia utilizada em cada uma das etapas é detalhada.

Os requisitos criados nesta etapa serão utilizados como base para avaliação de cada um dos passos subsequentes. Todos os testes empíricos descritos nas próximas seções foram realizados no software Protégé 4.1, e as tarefas de classificação foram realizadas no Hermit 1.3.6 (SHEARER *et al.*, 2008).

4.3.1 Construção do arcabouço teórico

O desenvolvimento recente da Ontologia Aplicada como disciplina independente é baseado na utilização de diversas fontes interdisciplinares, incluindo a inteligência artificial e filosofia. Esta última tem grande proeminência devido ao grande volume de contribuições ao longo dos séculos. Tomando como pressuposto a contribuição da filosofia ao estudo ontológico, a construção do arcabouço teórico foi baseada eminentemente em escolas filosóficas. A pesquisa foi conduzida por ampla revisão de literatura, com foco em autores ditos "realistas", em particular o filósofo austríaco Karl Popper.

O objetivo deste arcabouço foi fornecer base teórica para reconhecer afirmativas ontológicas e não-ontológicas contidas em prontuário médico. A validação do arcabouço teórico foi realizado através da análise de um pequeno extrato de um caso clínico disponível em (CONNORS; BRITTON, 2009). O resultado do arcabouço teórico é apresentado na seção 5.

4.3.2 Análise ontológica do Modelo OpenEHR

Esta etapa teve como intuito determinar as semelhanças e diferenças entre o modelo OpenEHR e ontologias realistas. Foi realizada em duas sub-etapas:

- Comparação do modelo de alto-nível
- Mapeamento dos conceitos contidos nos arquétipos clínicos

4.3.2.1 Comparação do modelo de alto-nível

Para a comparação dos modelos de alto-nível, foi analisada a documentação sobre o padrão OpenEHR, disponível em (OPENEHR FOUNDATION, 2007) e a ontologia CIR

(seção 3.3.2.3). Uma vez que a análise é focada no modelo de entrada do OpenEHR, que trata da informação clínica, serão omitidos dados relativos ao modelo de referência e a informações demográficas do paciente.

As ontologias realistas consideradas aqui serão a *Ontology for General Medical Science* (OGMS) (SCHEUERMANN *et al.*, 2009), a *Information Artifact Ontology* (IAO), ambas baseadas na *Basic Formal Ontology* (BFO).

A comparação entre os modelos será feita de forma manual, através da análise das definições em linguagem natural das principais classes da CIR e das ontologias realistas. As classes da CIR serão mapeadas às ontologias realistas, quando possível atribuindo uma definição lógica em lógica descritiva (axiomas).

O resultado final será apresentado como uma ontologia representada em OWL2-DL. Os resultados iniciais foram apresentados em um workshop internacional (ALMEIDA; ANDRADE, 2011; ANDRADE; ALMEIDA, 2011a; ANDRADE *et al.*, 2012a) e as modificações resultantes foram efetuadas na ontologia.

4.3.2.2 Mapeamento dos conceitos contidos nos arquétipos clínicos

Utilizando a ontologia criada pelo mapeamento entre a CIR e ontologias realistas como o modelo de alto-nível, vários arquétipos foram analisados para avaliar a possibilidade de representar o modelo de entrada do OpenEHR utilizando uma linguagem formal de acordo com os preceitos do realismo ontológico. Foram selecionados 13 arquétipos disponíveis no OpenEHR *Clinical Knowledge Manager* (OPENEHR FOUNDATION, 2011) seguindo os seguintes critérios:

1. Status de publicação – arquétipos publicados foram preferidos em relação a arquétipos ainda sob avaliação e revisão (*draft*). Porém, como o número de arquétipos disponíveis é ainda pequeno, a maioria dos arquétipos analisados ainda está em status *draft*.
2. Abrangência na representação dos fragmentos sentenciais selecionados (seção 4.2) – o arquétipo deve fornecer o contexto para interpretação que permita o uso de terminologias apropriadas, ou conjuntos de termos, para representar os fragmentos sentenciais obtidos dos prontuários.
3. Generalidade – dentre os arquétipos capazes de representar os fragmentos, foram selecionados aqueles de maior generalidade. Ex. O arquétipo Sintoma

foi selecionado para representar "dor", ainda que exista uma especialização deste arquétipo criado especificamente para este fim (arquétipo Dor/Sintoma)

Cada conceito do arquétipo selecionado – itens que podem receber algum valor, como Intensidade, Descrição clínica e Localização no corpo no arquétipo Sintoma (Figura 6, pg 88) – foi transferido para uma planilha juntamente com sua descrição em linguagem natural. O processo foi realizado mantendo a língua inglesa para evitar erros de tradução e possibilitar discussão com parceiros internacionais durante o estágio sanduíche. Para cada conceito foi atribuída uma expressão lógica baseada nas ontologias realistas com o objetivo de tornar explícito o significado do conceito. Um exemplo do resultado é apresentado na Tabela 3 a seguir.

Concept	Features:			
Definition	The features of the symptom			
Concept	Locations	Duration	Number occurrences	of Character
Definition	Locations in the body where the symptom is experienced.	The duration of the symptom since onset	The number of times this symptom has occurred	The character of the symptom
Concept	Location			
Definition	Single location in the body where the symptom is experienced.			
Concept	Location in body: openEHR-EHR-CLUSTER.anatomical_location.v1			
Definition	Details of the location in the body.			
Logical expression	(object or object_aggregate) AND partOf SOME patient AND inverse 'occurs in' some symptom	temporal_interval AND inverse hasDuration some symptom	RULE: Cardinality	ProcessProfile AND inverse (hasProcessProfile) SOME Symptom

Comment	Most likely refer to body regions, such as chest, legs, but can also refer to organs - the realist stance here can be problematic, as pain referred to a body region doesn't mean that the process is actually contained there. Could be simplified by a more permissive located_in relation accounting for occurs_in located in the spatial location of continuants during their full duration	Relation introduced in BFO2.0. In the case of a continuant, refers to the existence process	SymptomAggregate AND hasNumberOfGrains value X. Currently summarized under the SymptomAggregate to use the data property for number of occurrences, due to problems with cardinality reasoning	For continuant symptoms, hasQuality or anything else related to character. Classified as Partial mapping as it can receive subjective evaluations (ex. Bizarre symptom evolution)
----------------	---	---	--	---

Tabela 3- Extrato da planilha, demonstrando o processo de criação de definições lógicas para os conceitos dos arquétipos (em inglês)

As expressões lógicas que descrevem os conceitos dos arquétipos foram representadas em OWL2 utilizando o software de edição de ontologias Protégé 4.1. Os conceitos que não puderam ser mapeados foram identificados e classificados de acordo com o arcabouço teórico (seção 5). Os resultados desta etapa são apresentados na seção 6.

4.3.3 Representação ontológica da informação médica

Na última etapa, foram reunidos os achados das etapas anteriores para a definição de um modelo para a representação da informação médica. A partir do arcabouço teórico e dos conceitos não-mapeados provenientes dos arquétipos, foram propostas adições à representação realista.

Após a criação do modelo, foram criadas instâncias na ontologia, correspondentes às asserções contidas nos fragmentos sentenciais previamente selecionados. O processo foi realizado manualmente pelo autor, através dos seguintes passos:

1. Representação dos fragmentos sentenciais através de arquétipos
2. Criação das instâncias na ontologia de acordo com a definição lógica do conceito correspondente
3. Criação de relações entre instâncias conforme ditado pelo modelo

4.3.3.1 Representação dos fragmentos sentenciais por arquétipos

Neste passo, foi realizada a representação dos fragmentos sentenciais selecionados dos prontuários através dos conceitos dos arquétipos estudados. Para cada um dos fragmentos, foram preenchidos um ou mais pares conceito-valor. O Quadro 3 exemplifica o processo.

Fragmento sentencial: <i>Dispneia aos esforços moderados de cerca de 1 mês de evolução</i>	
Arquétipo: <i>openEHR-EHR-CLUSTER.symptom.v1</i>	
Conceito	Valor
Sintoma	Dispneia
Duração	1 mês
Fatores precipitantes	Esforços moderados

Quadro 3 – Representação dos fragmentos sentenciais em arquétipos

4.3.3.2 Criação das instâncias na ontologia de acordo com a definição lógica do conceito correspondente

Para a criação das instâncias na ontologia, foi utilizada a definição lógica dos sete conceitos (seção 4.3.3). O Quadro 4 demonstra o processo.

Conceito	Valor
Duração	1 mês
Definição lógica	Instância
<i>temporal_interval</i>	1_mês_001 ⁵⁵ Type <i>temporal_interval</i>
AND (BFO2:During_which_exists SOME Symptom)	1_mês_001 Type BFO2:During_which_exists SOME Symptom

Quadro 4 - Criação das instâncias na ontologia de acordo com a definição lógica do conceito correspondente

⁵⁵ O sufixo "001" foi adicionado para ressaltar que a entidade representada é uma instância, ou particular.

É importante notar que não é claro, a priori, se os valores preenchidos (ex. 1 mês) se referem a classes ou a instâncias. A interpretação padrão utilizada neste trabalho prediz que o valor refere-se a uma classe instanciada pela situação que o paciente relata. Por exemplo, o valor "1 mês" refere-se à duração da dispnéia particular do paciente X, duração esta que é do tipo "1_mês_de_duração", ou seja, intervalos temporais que tenham a extensão de 1 mês. Questões específicas às modelagens destes tipos não foram cobertos pela presente pesquisa, e devem ser objeto de pesquisas mais profundas. Por exemplo, 1 mês pode possuir de 28 a 31 dias e, portanto, não é um universal – esta questão não foi abordada.

4.3.3.3 Criação de relações entre instâncias conforme ditado pelo modelo

Finalmente, foram criadas relações entre as instâncias para manter a consistência entre as instâncias criadas. No exemplo do Quadro 4, é necessário relacionar a instância "1_mês_001" e a instância "sintoma_001", além de relacionar esta última ao "paciente_001". A criação das relações foi realizada manualmente para refletir o significado das asserções realizadas nos fragmentos sentenciais. Relacionamentos comuns foram reconhecidos e representados para facilitar e padronizar esta etapa.

A Figura 8 demonstra esquematicamente o resultado das 4 sub-etapas.

"Dispneia aos esforços moderados de cerca de 1 mês de evolução"

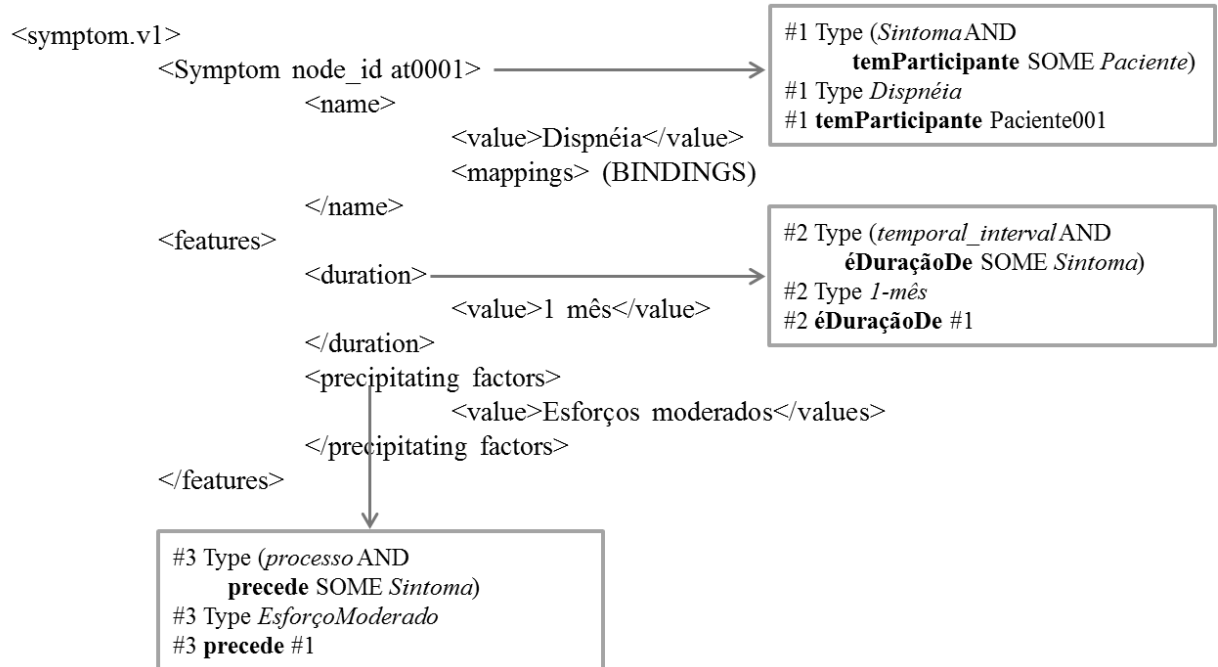


Figura 8 – Representação esquemática da representação ontológica dos fragmentos sentenciais

4.3.4 Critérios de avaliação da modelagem ontológica

A metodologia deve responder às necessidades atuais de organização da informação médica. Os critérios de avaliação podem ser divididos em duas categorias:

1) requisitos do sistema que usa a metodologia proposta para representação de prontuários médicos – ex. Capacidade de realizar inferências em lógica descritiva, uso de linguagens de marcação na representação da informação, etc.;

2) questões que devem ser respondidas através de manipulação simbólica do modelo de representação (questões de competência).

Para determinação dos requisitos, foi realizada revisão bibliográfica, somada à experiência do autor, que é profissional de saúde, no desenvolvimento e avaliação de sistemas de representação de prontuários médicos. Para objetivar a avaliação do modelo

criado, os critérios foram baseados na formulação de questões de competência (GRUNINGER; FOX, 1995). Essas questões devem ser adequadamente respondidas pelo sistema modelo de informação em conjunto com a ontologia, de forma que fique comprovado (pelo menos casuisticamente) que existem melhorias proporcionadas pelo modelo proposto. Originalmente, questões de competência devem ser respondidas através de inferências válidas em uma ontologia que modele um determinado domínio. Porém, como o objetivo não é o desenvolvimento de uma ontologia, mas o desenvolvimento de uma metodologia em que são utilizadas ontologias, as questões de competência foram utilizadas de forma mais abrangente.

A metodologia proposta, apesar de baseada nos princípios de desenvolvimento de ontologias, também deve promover o principal objetivo de modelos de informação em saúde: interoperabilidade. Considerando as definições da seção 3.1.1.1, a metodologia deve ser capaz de construir o significado proposicional de uma representação (EUZENAT J., 2001), deve atribuir a cada unidade de conhecimento importado a interpretação correta ou conjunto de modelos (PARK; RAM, 2004), e deve permitir trocar, compreender e agir em resposta a informação relacionada à saúde (STROETMANN *et al.*, 2009).

Considerando o descrito acima, foram criados os seguintes critérios:

1. Todo o conteúdo presente em prontuários médicos deve ser descrito através do mecanismo proposto, ou seja, a metodologia deve proporcionar possibilidade de mapeamento sistemático entre as duas propostas;
2. O conteúdo que não puder ser expresso em ontologias formais deve ser explicitamente descrito em detalhes para avaliação de sua importância em contextos práticos;
3. Deve permitir a integração entre aplicações diversas trabalhando sobre repositórios de dados heterogêneos;
4. A recuperação deve permitir inferências consistentes por operações em lógica descritiva.

A partir destes critérios, surgiram hipóteses que serão colocadas à prova durante o processo de avaliação:

- H1 - Ontologias podem expressar o mesmo conteúdo de modelos de informação

- H2 - A representação ontológica reduz ambiguidade conceitual e permite inferências robustas
- H3 - É possível recuperar informação de saúde representada ontologicamente através de operações em lógica descritiva
- H4 - O processo de representação da linguagem médica por ontologias é custo-efetivo no desenvolvimento de sistemas de informação em saúde

Para avaliar as hipóteses, serão utilizadas as seguintes questões para orientar a análise qualitativa, complementando com dados quantitativos sempre que possível:

- Quantos e quais conceitos presentes em arquétipos podem ser representados diretamente em ontologias realistas? Dentre os conceitos não representados, quantos e quais se referem a operações e sinais lógicos e quais se referem a limitações de representação da modelagem realista?
- A ontologia resultante fornece inferências corretas quando classificada? Qual é a natureza dos axiomas atribuídos aos conceitos presentes nos arquétipos?
- É possível recuperar corretamente as instâncias representadas na ontologia?

Por fim, o artefato resultante foi avaliado de acordo com os critérios identificados nos requisitos para avaliação da metodologia, contendo aspectos qualitativos e quantitativos. A avaliação segue práticas de avaliação de ontologias conforme descrito em (HOEHNDORF *et al.*, 2012), mas é importante ressaltar que o processo avaliativo é manual e a amostra estudada é notadamente insuficiente para análise estatisticamente significativa.

Com o intuito de avaliar de forma pormenorizada uma situação clínica, foi selecionado o domínio da trombose venosa, um problema comum em ambientes hospitalares. O caso de uso consiste na representação de um caso fictício criado por um expert de domínio (que não é um dos autores do presente trabalho), imitando a representação real de um caso em um prontuário médico. O expert foi instruído a reproduzir o caso da forma mais próxima à real, incluindo o uso de terminologia local, estrutura e conteúdo do prontuário. O registro completo pode ser encontrado no apêndice 10.1.

Além da transcrição do caso, foram avaliadas perguntas de competência baseadas no guia de auditoria clínica "Doenças tromboembólicas venosas: diagnóstico de trombose

venosa profunda"⁵⁶, criado pelo *National Institute for Health and Clinical Excellence* (NICE) e. As questões foram formuladas como questões OWL-DL ou SPARQL. Uma vez que o objetivo com estas questões não é a avaliação de aspectos sintáticos de OWL e RDF, existe flexibilidade na representação destas questões para avaliar se a informação necessária é disponibilizada e é recuperável por alguma das formas.

Para evitar repetição desnecessária, somente as questões aplicáveis ao caso de uso serão reproduzidas. A coleta dos dados originais e as questões geradas podem ser vistas na Tabela 4. A avaliação destas questões é discutida na seção 7. O documento original se encontra na língua inglesa, e as questões não foram traduzidas para não comprometer a fiabilidade.

	Original question	DL-query or SPARQL expression
	Did the patient have:	
1	<ul style="list-style-type: none"> an assessment of their medical history 	'clinical history'
2	<ul style="list-style-type: none"> a physical examination? 	'physical examination finding'
	If other causes were excluded and DVT was suspected, go to question 2. If not, end audit here.	
3	Was the patient's probability of DVT assessed using the two-level DVT Wells score?	'data item' AND WellsScore
	If 'yes', go to question 3. If 'no', end audit here.	
4	Record score:	<pre>SELECT Y WHERE ?x rdf:type WellsScore. ?x part_of PatientRecord001.</pre>

⁵⁶ Venous thromboembolic diseases: diagnosis of deep vein thrombosis – audit guide disponível em <http://guidance.nice.org.uk/CG144/ClinicalAudit/DVTDiagnosis/doc/English>

		?x 'is about' Patient001. ?x hasValue y?.
	If DVT was likely (2 points or more) go to question 4. If it was unlikely (1 point or less) go to question 18.	
	Question	
	Patients with a likely two-level DVT Wells score	
5	Was the patient offered a proximal leg vein ultrasound scan?	'plan specification' and ('is about' only (plan and (is_realized_by some (UltrasoundImagingTest AND has_participant SOME (leg AND part_of value patient010))))))
	If 'yes', go to question 5. If 'no', end audit here.	
6	Did the patient have a proximal leg vein ultrasound scan: Within 4 hours of being requested 0 Within 24 hours of being requested 0 More than 24 hours after being requested	US Request - Instance1: 'plan specification' and ('is about' only (plan and (is_realized_by some (Venous_duplex_Scan and (has_participant some (leg and (part_of value patient010)))))) -----#----- Instance2: temporal_interval AND hasBeginning SOME (temporal_instant AND inverse (hasBeginning) value Patient010USRequest) AND hasEnd

		<p>SOME (temporal_instant AND inverse (hasBeginning) SOME UltrasoundImagingTest)</p> <p>-----#-----</p> <p>SELECT x</p> <p>WHERE</p> <p>?x 'is about' instance2</p> <p>?x hasValue ?y</p>
<p>7</p>	<p>Was the patient offered a D-dimer test before the result of the proximal leg vein ultrasound scan was available?</p>	<p>'plan specification' AND ('is about' SOME (plan AND (is_realized_by SOME D-dimerTest AND has_participant VALUE patient010))) AND is_output_of SOME (OpenEHR:Investigation_request AND precedes SOME ('data item' AND is_output_of SOME (Venous_duplex_Scan AND has_participant SOME (leg and part_of value patient010))))</p>
<p>8</p>	<p>Was the proximal leg vein ultrasound scan positive?</p>	<p>'data item' AND is_output_of SOME (Venous_duplex_Scan AND has_participant SOME (leg and part_of value patient010)) AND 'is about' SOME (BloodClot AND located_in SOME (leg AND part_of value patient010))</p> <p>-----#-----</p> <p>('data item' AND is_output_of SOME (Venous_duplex_Scan AND has_participant SOME (leg and part_of value patient010))) AND (AbnormalValue)</p>

	If 'yes', go to question 15. If 'no', go to question 8.	
9	Was the patient offered a D-dimer test?	'plan specification' AND ('is about' SOME (plan AND (is_realized_by SOME (D-dimerTest AND has_participant VALUE patient010))))
	If 'yes', go to question 12. If 'no', end audit here.	
10	While waiting for a proximal leg vein ultrasound scan, did the patient have an interim 24-hour dose of parenteral anticoagulant ⁵⁷ ?	(medication_administration AND has_part SOME NeedleInjection) AND has_participant SOME (Medication AND has_part SOME (PortionOfChemicalSubstance AND bearer_of SOME AnticoagulantRole) AND precedes SOME (UltrasoundImagingTest AND has_part SOME (leg AND part_of VALUE patient010)))
11	While waiting for a proximal leg vein ultrasound scan, did the patient have a D-dimer test?	D-dimerTest AND has_participant VALUE patient010 AND precedes SOME (UltrasoundImagingTest AND has_participant SOME (leg AND part_of VALUE patient010))
	If 'yes', go to question 11. If 'no', end audit here.	
12	Was the proximal leg vein ultrasound scan positive? (repeated question)	'data item' AND is_output_of SOME (Venous_duplex_Scan AND has_participant SOME (leg and part_of value patient010)) AND 'is about' SOME (BloodClot AND located_in SOME (leg AND part_of

⁵⁷ Medication is here defined as a Processed material that has part some chemical substance that has some pharmacologically active role

		<p>value patient010))</p> <p>-----#-----</p> <p>('data item' AND is_output_of SOME (Venous_duplex_Scan AND has_participant SOME (leg and part_of value patient010))) AND (AbnormalValue)</p>
	If 'yes', go to question 15. If 'no', go to question 12.	
13	Was the D-dimer test positive?	<p>('data item' and is_output_of SOME (D-dimerTest AND has_participant VALUE patient010)) AND (AbnormalValue)</p>
	If 'yes', go to question 13. If 'no', go to question 16.	
14	Was the patient offered a repeat proximal leg vein ultrasound scan 6–8 days later?	<p>'plan specification' AND (('is about' some (plan and (is_realized_by only (UltrasoundImagingTest AND has_participant SOME (leg and part_of value patient010)))))) AND is_output_of SOME OrderRequestProcess AND inverse (precedes) SOME ((temporal_interval AND inverse ('is about') SOME ('measurement datum' AND hasValue value 6 and 'has measurement unit label' value day) AND inverse (precedes) SOME (UltrasoundImagingTest AND has_participant SOME (leg and part_of value patient010)))) and not ((temporal_interval AND inverse</p>

		<p>('is about') SOME ('measurement datum' AND hasValue value 8 and 'has measurement unit label' value day) AND inverse (precedes) SOME (UltrasoundImagingTest AND has_participant SOME (leg and part_of value patient010))))</p>
	If 'yes', go to question 14. If 'no', end audit here.	
15	Was the repeat proximal leg vein ultrasound scan positive?	<p>'data item' AND is_output_of SOME (UltrasoundImagingTest AND has_participant SOME (leg and part_of value patient010)) AND ('is about' SOME (BloodClot AND located_in SOME (leg AND part_of value patient010))) AND inverse (precedes) SOME (UltrasoundImagingTest AND has_participant SOME (leg and part_of value patient010))</p>
	If 'yes', go to question 15. If 'no', go to question 16.	
16	Was DVT diagnosed and treatment started?	<p>diagnosis AND 'is about' SOME (DeepVenousThrombosis AND 'inheres in' VALUE patient010) #_____ medication_administration AND has_participant SOME (Medication AND has_part SOME (PortionOfChemicalSubstance AND bearer_of SOME AnticoagulantRole))</p>
	End audit here.	

17	Were alternative diagnoses considered?	diagnosis AND 'is about' SOME (disease AND 'inheres in' VALUE patient010) AND NOT ('is about' SOME DeepVenousThrombosis)
	Was the patient advised that it was not likely they had PE?	NOT REPRESENTABLE USING CURRENT METHODS (requires free-text analysis)

Tabela 4 – Questões de competência originárias da trilha de auditoria para Trombose Venosa Profunda

A avaliação da metodologia consistiu na descrição quantitativa do processo de aplicação empírica da metodologia e avaliação da consistência interna dos dados armazenados. A avaliação foi realizada pelo próprio autor em conjunto com os supervisores. Os resultados desta etapa são apresentados na seção 7.

Os resultados e discussão da metodologia serão apresentados de forma conjunta, correspondendo à divisão em 3 etapas da metodologia. A Figura 9, a seguir, representa esquematicamente os passos realizados na metodologia.

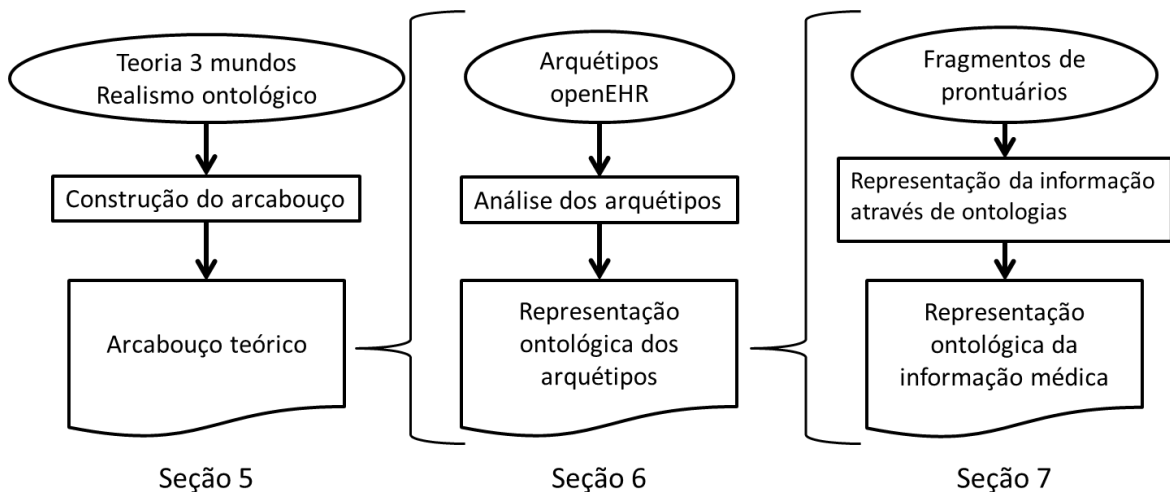


Figura 9 – Representação esquemática da metodologia utilizada na pesquisa

Esta metodologia permitirá explorar a natureza ontológica de dados do prontuário representados por modelos de informação e avaliar a capacidade de representar estes

dados por ontologias realistas. Caso possível, as ontologias poderão ser usadas para diminuir a ambiguidade dos conceitos usados nos modelos de informação e para permitir manipulação lógica das informações do prontuário médico.

5 CONSTRUÇÃO DO ARCABOUÇO TEÓRICO

A representação da linguagem médica utilizada em prontuários por ontologias requer vários passos intermediários afim de esclarecer o significado e permitir uma análise correta. O primeiro passo consistiu em desenvolver uma abordagem adequada onde poderíamos identificar o referente de um termo, se ele existir. O resultado foi publicado em algumas conferências (ALMEIDA; ANDRADE, 2011; ANDRADE; ALMEIDA, 2011b; ALMEIDA *et al.*, 2012) e é descrito nesta seção.

5.1 ABORDAGEM PARA ANÁLISE

Propôs-se um arcabouço representado na Figura 9, que foi criada para organizar a informação presente em prontuários médicos de acordo com a teoria dos três mundos.

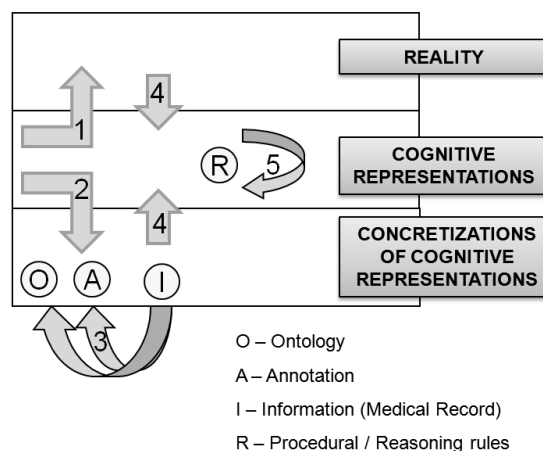


Figura 10 – Abordagem usada na análise

Fonte: (ALMEIDA; ANDRADE, 2011)

Nesta abordagem, tudo começa no nível de representações cognitivas, quando um médico observa a realidade do lado do paciente (seta 1). Cada uma destas entidades é filtrada pela cognição e representada por artefatos (seta 2). Entidades ontológicas (entidades O) são analisadas de acordo com a análise científica rigorosa, e são baseadas na realidade em si e não em representações mentais do médico. Exemplos de entidades

ontológicas são as células, as características anatômicas e substâncias químicas. Estas entidades denotam entidades do mundo 1 através de teorias que podem ser objetivamente falsificadas.

Entidades epistemológicas são registradas através de conceitos (entidades A). São representações cognitivas da realidade, e podem incluir entidades sem um referente na realidade. Estas entidades não são descritas através de medidas independentes e objetivas, e compreendem analogias, análises heurísticas e avaliações subjetivas. É importante destacar que um dos objetivos da ciência é criar novas construções e medições indiretas, a fim de teorizar e manipular entidades reais descritas por estas entidades conceituais. E sustenta-se aqui que, em um mundo ideal, cada entidade pode ser descrita adequadamente em ontologias. No entanto, duas questões motivam esta distinção: 1) o domínio médico é muito complexo e os profissionais de saúde constantemente usam analogias, heurísticas e práticas padronizadas para orientar o diagnóstico e terapia; 2) pretende-se que o resultado final deste trabalho melhore a descrição formal de prontuários médicos e comunicação médica, que muitas vezes contém termos ambíguos que só podem ser interpretados de uma forma não-científica. Exemplos destes incluem "gravidade" de uma dor e uma sensação de "sentir-se bem", que pode ser medido ou, em última análise descrita por termos objetivos, mas que não iria transmitir o significado geral pretendido pelo autor (respectivamente, a gradação ordinal da alteração objetiva no comportamento causado pela dor, e o conjunto de comportamentos compartilhados por cada pessoa que se sente bem).

Em seguida, o médico cria um registro (entidade I) para registrar as representações de acordo com seu conhecimento prático e teórico (seta 3). A clara distinção entre as entidades que denotam e informações sobre estas entidades é suposto representar a distinção entre a entidade e sua medição por algum padrão.

Outros médicos podem constantemente interpretar registros e realidade (setas 4), resultando em novas representações cognitivas. Finalmente, os médicos envolvidos nos cuidados de saúde fazem julgamentos e processam informações atuais e anteriores. Parte desse processamento de informações (seta 5) segue regras de formação acadêmica do médico, que determinam a probabilidade de um diagnóstico ou interpretação correta de um resultado de exame, para mencionar apenas alguns. A representação deste processo de raciocínio também é necessária para a continuação de cuidados, que é parte complementar do registro (entidade R). Exemplos disso incluem regras para a interpretação de dados

laboratoriais, como hemoglobina <12 g / dl significa "baixo nível de hemoglobina", e informação relevante negativa, como "ausência de alteração intestinal durante os episódios".

É possível reconhecer quatro tipos de informação:

- i. Informação que representa aspectos da realidade;
- ii. Informação que representa construtos úteis para a prática médica que não são empiricamente verificáveis;
- iii. Informação que representa observações sobre a realidade, não a realidade em si;
- iv. Informação que representa observações sobre a compreensão do médico ou paciente a respeito da situação clínica, não a realidade em si.

Segundo este modelo, nem representações da realidade, nem representações de processos de pensamento são interpretadas da mesma forma por duas pessoas. No entanto, permitir a manipulação de entidades do mundo 3 é fundamental para o desenvolvimento de novas funcionalidades nos sistemas de saúde, como apoio à decisão, inferências e classificação de informação.

O arcabouço irá fornecer os fundamentos para classificar as entidades em diferentes categorias. O primeiro passo é olhar para as entidades que denotam coisas no mundo 1. De acordo com a abordagem acima, apenas informações que representam aspectos da realidade podem ser adequadamente representadas por universais - o que significa que são ou continuantes ou ocorrentes, que exigem um referente na realidade. O segundo passo é a busca de entidades que não se referem à realidade, mas sim a uma interpretação da realidade. É evidente que (ii) e (iii) estão intimamente relacionadas com a realidade, sendo que (ii) é um substituto para um estado de coisas do lado do paciente, e (iii) é uma representação objetiva das suas medidas. Relações que permitem a interpretação adequada dessas declarações são específicas para cada domínio. Por exemplo, o exame da cor da esclerótica pode indicar icterícia, exibindo cor amarelada, substituto para problemas do fígado, ou anemia, exibindo cor azulada, substituto de anemia por deficiência de ferro. A interpretação sobre o significado destes sinais depende de treinamento, práticas culturais e características subjetivas. Similarmente, asserções como "nível de bilirrubina total no sangue do paciente X é alto" requerem o conhecimento sobre o método de amostragem

e análise, da distribuição probabilística da concentração de bilirrubina na população normal, a consideração de erros de medição e fatores de confusão e compreensão do significado de unidades de medida. A última categoria (iv) requer mais atenção, uma vez que as práticas de raciocínio médicos incluem tanto as relações ontológicas e regras heurísticas ad hoc que podem não corresponder à verdade no mundo real. Considera-se que a informação (iv) eventualmente será registrada no prontuário para descrever o processo de raciocínio.

Um exemplo preliminar do arcabouço foi criado através da análise de entidades informacionais contidas nos prontuários médicos. Como um exemplo, o quadro 5 demonstra um pequeno extrato de um caso clínico disponível em (CONNORS; BRITTON, 2009), escolhido devido à clareza e abrangência da descrição do caso, e devido à apresentação explícita do processo de raciocínio.

Após a obtenção de fragmentos sentenciais através da avaliação de um especialista de domínio, as entidades foram classificadas em um dos quatro tipos, através dos seguintes critérios:

- a) Informações que representam aspectos da realidade foram mapeados pelos processos 1) e 2) para entidades (O) (Quadro 6);
- b) Informações que representam construtos úteis para a prática médica foram mapeados pelos processos 1) e 2) para entidades (A) (Quadro 7);
- c) Informações que representam observações sobre a realidade foram mapeadas pelo processo 3) para entidades (I) (Quadro 8);
- d) Informações que representam observações sobre a compreensão dos médicos foram mapeadas pelo processo 4) para entidades (R) (Quadro 9).

A 62-year-old woman presented to the urgent care clinic with gingival bleeding after periodontal scaling of her lower-right second molar. She had undergone the procedure 5 hours before presentation, and the bleeding has persisted despite the application of pressure and ice.

The patient recalled a similar episode that had occurred 6 months earlier, also after a periodontal procedure, in which bleeding had stopped only after firm pressure had been applied and held for 6 hours.

She was otherwise in her usual state of good health. She reported no easy bruising, epistaxis, rectal bleeding, hematuria, weakness, fatigue, arthralgia, dyspnea, jaundice, abdominal pain, back pain, rash, or confusion.

Quadro 5 - exemplos de um extrato de um prontuário médico

Physician
 Woman
 62 years-old
 Patient report
 Time of bleeding
 Aspirin
 Aspirin taken daily (rule)
 Thiazide diuretic
 Physical exam finding of that encounter

Quadro 6 - Informações que representam aspectos da realidade

State of good health
 Uncomplicated spontaneous vaginal delivery
 No prior episodes of unpredictable bleeding
 No allergies
 Cosmetic blepharoplasty, without bleeding complications
 ...

Quadro 7 - Informações que representam construtos úteis para a prática médica

Heart rate: 80 bpm
 Blood pressure: 128/76 mmHg
 White-cell count = 6,200
 Lymphocytes = 37
 Platelet-count = 352,000
 Creatinine = 1.4
 Albumin = 3.9
 Prothrombin time = 13 sec
 ...

Quadro 8 - Informações que representam observações sobre a realidade

Patient class: "Emergency patient"
 Bleeding had persisted despite the application of pressure and ice
 Bleeding had stopped only after firm pressure had been applied and held for 6 hours
 Childbirth and orthopedic surgeries are "significant hemostatic challenges"
 Bleeding that do not stop after ice and pressure is a "bleeding complication"
 Patient underwent hemostatic challenges without bleeding complications, which suggests that this is an acquired bleeding diathesis
 Epistaxis, menorrhagia and gingival bleeding are mucocutaneous bleedings
 Mucocutaneous bleedings are typical or disorders of primary hemostasis
 Unprovoked hemarthroses and deep soft-tissue hematomas are characteristic of disorders of coagulation factors
 Factor VIII inhibitor can manifest as mucocutaneous bleeding, and is an exception to the previous rule
 The timing of bleeding after vascular trauma is different
 The patients presentation suggests platelet disorder
 Von Willebrand's disease cause acquired platelet disorder
 Factor VIII inhibitor cause acquired platelet disorder like syndrome
 Liver disease cause acquired platelet disorder
 Thrombocytopenia cause acquired platelet disorder
 Uremia cause acquired platelet disorder

Quadro 9 - Informações que representam observações sobre a compreensão dos médicos

5.2 CONCLUSÃO E IMPLICAÇÕES PARA A PESQUISA

Até aqui, mostrou-se que o arcabouço fornece orientações para distinguir entidades em diferentes níveis. Até agora, as teorias apresentadas não são falsificáveis, mas a motivação para esta tese é a formalização adequada, robusta, coerente e escalável dos prontuários médicos. Assim, a pergunta importante a ser respondida neste ponto é: "o arcabouço de análise baseada nas teorias de Karl Popper proporciona um melhor modelo para representar formalmente informações em registros de saúde?" Embora tal resposta, em última análise, requeira testes empíricos adicionais fora do âmbito do presente trabalho, nos próximos parágrafos serão apresentadas algumas conclusões surgidas durante a pesquisa e a arquitetura prevista para o sistema. Podemos identificar as principais vantagens desta abordagem, a saber: 1) Separação dos universais e classes conceituais; 2) Distinção de entidades reais e observações objetivas de entidades reais, 3) Ligação entre o conhecimento representado em lógica e outros tipos de operações de inferência.

1) A separação entre universais e classes conceituais: A partir de uma perspectiva de modelagem, este arcabouço oferece algumas vantagens. A primeira é a clara distinção entre a realidade científica e a abordagem médica à realidade, exemplificado pela diferença entre 1) ver um paciente tossir, desenvolver febre e mostrar uma condensação na base dos pulmões e 2) ver um paciente com pneumonia. Até agora, as ontologias têm tentado capturar as construções básicas do mundo real e compreender a natureza de entidades materiais, qualidades, processos e assim por diante. No entanto, um prontuário médico é um dispositivo de comunicação entre as pessoas, que pode analisar complexos "jogos de linguagem" que requerem conhecimento prático específico. No nosso exemplo, podemos ver a classe "complicações hemorrágicas". Enquanto muito útil para o raciocínio médico que aparece em *guidelines* científicos (HIRSH *et al.*, 2003), tal construção só pode ser compreendida no contexto de um determinado estado. Por exemplo, hemorragia intestinal não é uma complicação de cirurgia ortopédica, mas é uma complicação da terapia anticoagulante que pode ser necessária antes da cirurgia. Em análise posterior desta afirmação, pode-se identificar um claro processo ontológico, "sangramento", um processo que envolve a presença de sangue visível em um lugar inesperado como a pele ou intestino. A declaração "complicação" implica que este sangramento é inesperado e causado pela condição anterior. Uma vez que algum sangramento é esperado em qualquer a cirurgia, complicação hemorrágica é o conjunto de todos os sangramentos, mas a sua definição depende de processos anteriores e do conhecimento médico atual. E como o conhecimento médico é probabilístico, a resposta para a pergunta "um sangramento intestinal é uma complicação sangramento de uma cirurgia?" não pode ser ontologicamente determinado. O tipo não-ontológico do conhecimento é, por razões mencionadas acima, impróprio para ontologias realistas. No entanto, este termo pode ser tanto intensionalmente - hemorragia inesperada - e extensionalmente - lista de cada complicação para cada procedimento - definido, e pode ser representado de forma declarativa em formulações lógicas. Isto leva a uma discussão mais pragmática sobre limitação de diferentes formulações lógicas para proporcionar a semântica pretendida da linguagem natural, que está fora do escopo deste documento.

2) Distinção de entidades reais e observações objetivas de entidades reais: Outra vantagem da visão popperiana é a clara separação da informação e realidade. Isso tem sido evidente nos vários esforços de desenvolvimento de ontologias, em particular o consórcio OBI e o desenvolvimento da Information Artifact Ontology (IAO). Popper argumenta que as relações entre teorias distintas de mundo 3 será diferente das relações

entre seus referentes no mundo. A semelhança das duas medidas em unidades de medida diferentes (por exemplo, 120 mmHg e 12 cmHg) é uma teoria aritmética simples, que ainda acrescenta semântica (a manipulação de símbolos em sistemas de linguagem diferentes). A separação entre informações e realidade é a ligação entre a realidade e as diferentes perspectivas da realidade que serão encontradas em um registro, para que possamos relacionar as entidades "pressão arterial sistólica elevada" (entidade real) ", medição de pressão arterial sistólica" (processo), "medição da pressão arterial sistólica elevada" (resultado da medição, como interpretado por um conjunto de regras) e "pressão arterial sistólica: 150 mmHg" (resultado da medição, gravado em alguma unidade de medida padrão).

3) A conexão entre o conhecimento representado através de lógica e outros tipos de operações de inferência: O registro explícito do raciocínio e regras heurísticas em prontuários médicos requer uma nova distinção, que é fornecida pela estrutura proposta. Em nossos exemplos podemos distinguir três níveis, refletidos pelas perguntas: 1) O paciente tem um distúrbio de sangramento? 2) O paciente tem um diagnóstico de distúrbio hemorrágico? e 3) o médico suspeitou de distúrbios hemorrágicos dado os sinais e sintomas do paciente? Um exemplo é a comparação entre essas três afirmações, que diferem apenas de acordo com o processo de pensamento que lhes deu origem (Quadro 10). É importante observar que alguns deles resultam em probabilidades, mas outros resultam em algum valor de verdade.

Suspeita clínica	Suspeita clínica	Suspeita clínica
Probabilidade >70%	Probabilidade >80%	Probabilidade >90%
Provisório	Possível	Provável
Diagnóstico	Diagnóstico	Diagnóstico
Doença	Doença	Doença

Quadro 10 - Comparação entre afirmações em diferentes níveis ontológicos

Várias regras de inferência podem ser feitas explícitas e representadas em diferentes níveis do sistema. Um exemplo simples é o índice de massa corporal - IMC (peso dividido pelo quadrado da altura). A interpretação da medida determina, por exemplo, que uma

pessoa com IMC maior que 25 possui sobrepeso. Portanto, sobrepeso é um tipo da realidade ou o resultado de um processo de raciocínio?

Diferentes representações podem ser vistas na Tabela 5. Note que obesidade está no lado da realidade, enquanto avaliações clínicas são construtos médicos. Tal é exemplificado por achados de (CROSS *et al.*, 1995), que demonstram a falta de acordo sobre critérios clínicos para avaliação da obesidade infantil. Ainda assim, enquanto medidas objetivas permitem consenso, não refletem o verdadeiro status da obesidade (POIRIER, 2007). De toda forma, parece razoável que exista uma qualidade "obesidade" no mundo real, que através de uma combinação de diferentes avaliações subjetivas e objetivas possa ser inferido pelo clínico.

	Obesidade	Complicação hemorrágica da cirurgia	Diagnóstico suspeito de Hemofilia
Entidade O	Obesidade (qualidade)	Sangramento (processo) Cirurgia (processo)	Hemofilia (disordem)
Entidade A	Avaliação Clínica Da Obesidade = Obese	Sangramento001 (instanceOf Sangramento) isA Complicação Hemorrágica Da Cirurgia	Avaliação Possibilidade De Hemofilia = Possível
Entidade I	Medida De Peso (item de dados)		Diagnóstico (item de dados) é Sobre only Hemofilia
Entidade R	SE (Medida De Peso / Medida De Altura ²) > 35): Obesidade (qualidade) == verdade	SE (hemorragia ou hematoma complicando um procedimento OR intervenção, incluindo retorno para o bloco cirúrgico OR transfusões de produtos de sangue): Complicação Hemorrágica Da Cirurgia == verdade (STOKES <i>et al.</i> , 2011)	SE (Petéquias AND Dor Articular AND (Sangramento Nasal AND (NOT precedido Por Trauma))): Diagnóstico Hemofilia == suspeito

Tabela 5 – Comparação de perspectivas ontológicas e não ontológicas sobre a mesma entidade

6 ANÁLISE ONTOLÓGICA DO MODELO OPENEHR

6.1 COMPARAÇÃO DO MODELO DE ALTO-NÍVEL

A comparação entre os modelos OpenEHR e ontologias realistas foi feita a partir da distinção entre o Clinical Investigator Record (CIR) (Figura 5) e a ontologia OGMS. Segue-se uma breve descrição de cada entidade da CIR (BEALE; HEARD, 2007b; BEALE; HEARD, 2007a)⁵⁸:

- Informação de cuidado: as informações registradas em um registro de saúde constituem uma escolha seletiva de notas sobre eventos e outras situações reais, que serão interpretadas por outros profissionais;
- Evento Administrativo: registro de um evento de negócio ocorrendo em um contexto administrativo; ex. admissão, marcação, encaminhamento, alta, etc.:
- Observação: informação criada pelo ato de observação, mensuração, questionamento ou teste do paciente ou substância relacionada;
- Opinião: inferências do investigador sobre o significado das observações; inclui diagnósticos, avaliações, planos e objetivos;
- Instrução: instruções baseadas em opinião de forma suficientemente detalhadas para que possam ser executadas e cumpram uma determinada intervenção;
- Ação: registro de ações intervencionistas que já ocorreram, seja por instruções ou outros;
- Proposta: opinião sobre o que deve ser feito pelo investigador;
- Avaliação: opinião sobre o que está acontecendo no organismo do paciente;

⁵⁸ Repetem-se aqui trechos da seção 3.3.2, para facilitar a compreensão da análise.

- Diagnóstico: conjunto de sinais e sintomas observados, que os designam no entendimento do investigador como um fenômeno conhecido;
- Risco: uma avaliação quantificada que fornece a base para uma investigação (ex. História familiar);
- Prognóstico: avaliação de um resultado relacionado ao diagnóstico corrente;
- Cenário: pensamento clínico baseado em possibilidades de resultado (perguntas do tipo "what-if");
- Objetivo: pensamento clínico que consiste em afirmações sobre um estado desejado para o paciente;
- Plano: pensamento clínico que consiste em afirmações sobre como os objetivos são alcançados;
- Solicitação de investigação: conjunto de instruções com o propósito de gerar novas observações;
- Solicitação de intervenção: conjunto de instruções com o propósito de alterar o estado do sistema do paciente;

A CIR é graficamente estruturada como uma taxonomia. As setas parecem representar ligações do tipo *É-Um*, porém não é claro exatamente a qual tipo se referem (SCHWARZ; SMITH, 2008). Uma vez que a CIR não trata de universais, mas sim de classes definidas (ver seção 3.2.2), optou-se aqui por interpretar as relações taxonômicas entre A e B como uma relação formalmente definida como: $A \text{ É-Um } B = \text{para todo } x, t, \text{ se } x \text{ É-Instância-de } A, \text{ então } x \text{ É-Instância-de } B \text{ no tempo } t$. Esta definição inclui um aspecto temporal, mas é capaz de fornecer apenas condições necessárias, não suficientes. Por exemplo, permite representar *Leucócitos-Em-Uma-Bolsa É-Um Leucócito*, ainda que *Leucócitos-Em-Uma-Bolsa* não seja um tipo especial de leucócito.

Fica claro, à primeira inspeção, que todas as classes da CIR referem-se a artefatos de informação, com propósitos e conteúdos distintos. Exibe, portanto, o mesmo princípio

classificatório da hierarquia "Information Content Entity", presente na IAO (figura 11). Esta classe é definida como "uma entidade que é genericamente dependente em algum artefato e possui relação do tipo é-sobre com alguma entidade"⁵⁹. Mais especificamente, as classes da CIR correspondem à classe "Data Item", definida como "um entidade de conteúdo informacional que pretende ser uma afirmativa verdadeira sobre algo (...) e é construída/adquirida por um método que confiavelmente tende a produzir afirmativas (aproximadamente) verdadeiras"⁶⁰.



Figura 11 – Extrato da união entre OGMS e IAO, demonstrando subclasses da classe "information content entity"

⁵⁹ an information content entity is an entity that is generically dependent on some artifact and stands in relation of aboutness to some entity

⁶⁰ a data item is an information content entity that is intended to be a truthful statement about something (modulo, e.g., measurement precision or other systematic errors) and is constructed/acquired by a method which reliably tends to produce (approximately) truthful statements.

Análise mais pormenorizada da OGMS demonstra classes criadas com intuito similar à CIR, como por exemplo a classe "*Clinical history*", definida como "séries de afirmativas representando qualidades relevantes à saúde de um paciente ou à família do paciente" ⁶¹. A IAO também fornece classes semelhantes, como "*Plan Specification*", definido como "uma entidade de informação diretiva que, quando concretizada, é realizada em um processo no qual o ator tenta atingir os objetivos, em parte fazendo as ações especificada" ⁶².

Após a comparação das definições, foi feito um mapeamento *ad hoc* das classes de ambos os modelos, que pode ser visto na tabela 6. As classes oriundas da CIR foram identificadas com o prefixo "OpenEHR:".

⁶¹ series of statements representing health-relevant qualities of a patient and of a patient's family.

⁶² a directive information entity that when concretized it is realized in a process in which the bearer tries to achieve the objectives, in part by taking the actions specified. Plan specifications includes parts such as objective specification, action specifications and conditional specifications.

CIR		OGMS/IAO
Informação de Cuidado	É-Um	Data item
História	É-Um	Clinical history
Observação	É-Um	Image finding OR lab. Finding OR Clinical finding OR physical examination
Ação	É-Um	Report AND is_about ONLY process
Opinião	É-Um	Clinical finding
Avaliação	É-Um	Clinical picture
Diagnóstico	É-Um	Diagnosis
Prognóstico	É-Um	Prognosis
Proposta	É-Um	Objective specification
Cenário	É-Um	Objective specification
Objetivo	É-Um	Objective specification
Recomendação	É-Um	Objective specification
Instrução	É-Um	Plan specification
Solicitação de intervenção	É-Um	Plan specification AND (is_about some Treatment)
Solicitação de investigação	É-Um	Plan specification AND (is_about some (Laboratory_test))

Tabela 6 – Correspondências ad hoc entre a CIR e IAO/OGMS

Enquanto a motivação para a construção de cada estrutura é diferente, as ontologias IAO e OGMS provaram ser capazes de representar adequadamente o significado de cada tipo de informação na CIR. Uma observação é um item de dados que resulta de um encontro médico, sendo a descrição de uma entidade, normalmente, o paciente. Ao classificar as outras classes de acordo com o resultado desejado, as sub-classes de "Proposta" foram reunidas sob "*Objective Specification*", e as sub-classes de "Instrução" sob "*Plan Specification*". Finalmente, "Ação" foi representada como um tipo especial de documento, uma vez que necessariamente descreve um processo. A única classe não representada é a classe "Risco", por sua incompatibilidade com ontologias realistas. As Figuras 12 e 13 apresentam o resultado final do alinhamento entre a CIR e as ontologias IAO e OGMS, representados no software Protégé.

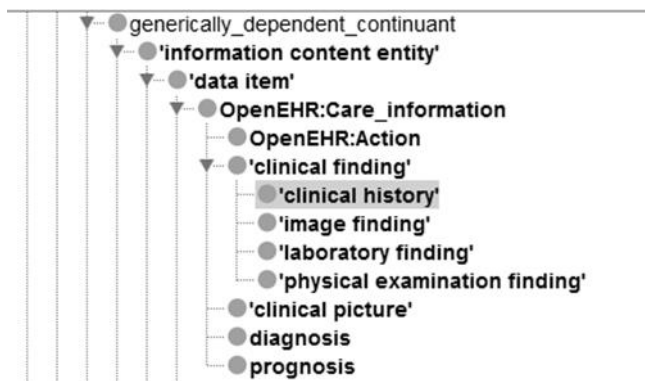


Figura 12 – Hierarquia OGMS/IAO/CIR – Care information

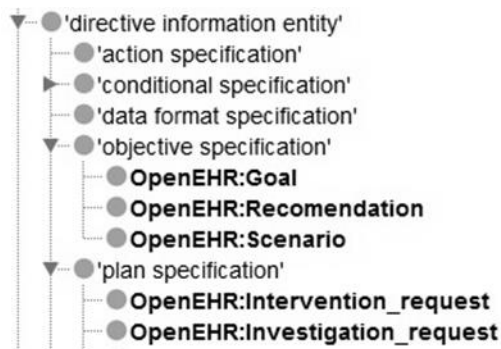


Figura 13 - Hierarquia OGMS/IAO/CIR – Directive information entity

Para interpretar a relação entre observações e informação, é necessário entender o que é uma "Entrada" (no original, "*Entry*") no OpenEHR. Uma "Entrada" é definida como "uma afirmação clínica logicamente única, e que pode ser uma frase narrativa pequena, mas que pode também conter um volume significativo de dados. Por exemplo um resultado microbiológico completo, uma nota de exame psiquiátrico, ou uma complexa solicitação de medicação." ⁶³ (BEALE; HEARD, 2007a, p.29-30). A semelhança entre uma "Entrada" e a classe da IAO "*Data Item*" permite reconhecer que parte das distinções entre as classes da CIR reflete a distinção entre o referente da informação. Por exemplo, uma "Observação" é uma informação sobre qualquer entidade relatada pelo paciente, enquanto uma "Ação" refere-se exclusivamente a processos executados para o cuidado do paciente. Ao mesmo tempo, evidencia-se neste fato a frágil distinção ontológica que serve de base para o

⁶³ a logically single 'clinical statement', and may be a single short narrative phrase, but may also contain a significant amount of data, e.g. an entire microbiology result, a psychiatric examination note, a complex medication order

OpenEHR. Uma pequena narrativa contendo a história de um paciente pode conter referências a medicações passadas, cirurgias realizadas e diagnósticos anteriores – é difícil determinar ontologicamente se a narrativa corresponde a uma entrada ou a 3 entradas diferentes oriundas de hierarquias diferentes.

A favor dos princípios de modelagem do OpenEHR, é possível argumentar que a CIR representa adequadamente as convenções correntes sobre a forma como a documentação médica deve ser dividida. Porém, a falta de rigor na modelagem gera ambiguidade. Uma análise mais detalhada da classe "Observação" sugere que esta classe representa informação sobre a realidade do paciente. Entretanto, a definição torna confuso se a informação se refere a alguma entidade do lado do paciente (ex. pressão arterial de 120/85 mmHg), ou se se refere ao processo de observação ou mensuração (ex. mensuração da pressão arterial). Um resultado microbiológico contém não apenas os micro-organismos encontrados na amostra do paciente, mas também os métodos utilizados para esta investigação. Este aspecto tornar-se-á mais claro na próxima seção.

6.2 MAPEAMENTO DOS CONCEITOS CONTIDOS NOS ARQUÉTIPOS CLÍNICOS

Foram selecionados, a partir dos critérios pré-estabelecidos, treze arquétipos para análise (Tabela 7).

Nome de arquétipo	Número de conceitos
openEHR-EHR-CLUSTER.symptom.v1	29
openEHR-EHR-OBSERVATION.bodily_output-defaecation.v1	5

openEHR-EHR-OBSERVATION.urine_output.v1	5
openEHR-EHR-CLUSTER.fluid.v1	20
openEHR-EHR-EVALUATION.reason_for_encounter.v1	2
openEHR-EHR-OBSERVATION.global.v1	20
openEHR-EHR-CLUSTER.inspection.v1	12
openEHR-EHR-CLUSTER.palpation.v1	15
openEHR-EHR-CLUSTER.percussion.v1	5
openEHR-EHR-EVALUATION.substance_use_summary.v1	10
openEHR-EHR-OBSERVATION.substance_use.v1	7
openEHR-EHR-ITEM_TREE.medication.v1	33
openEHR-EHR-EVALUATION.problem-diagnosis.v1	25
TOTAL	188

Tabela 7 – Arquétipos selecionados

Após a avaliação destes 188 conceitos, foram geradas definições em OWL. A Tabela 8 mostra exemplos do mapeamento entre conceitos OpenEHR e ontologias realistas.

Conceito do arquétipo	Definição realista (OWL)
Onset description	processual_entity AND occupies SOME (temporal_interval AND has_part SOME (temporal_instant AND inverse hasBeginning SOME symptom))
Period since bowels last	temporal_interval AND hasBeginning SOME (temporal_instant

moved	AND inverse (hasEnd) SOME BowelMovement AND hasEnd VALUE (Current_temporal_instant))
Amount (Substance Use)	(VolumeQuality OR QuantityQuality) AND inheres_in SOME (Substance AND participantOf SOME (Process AND partOfOccurent SOME SubstanceConsumptionProcess AND occursIn SOME temporal_interval))

Tabela 8 – Definições lógicas realistas para conceitos de arquétipos OpenEHR

Os arquétipos são organizados como árvore, com conceitos do tipo "cluster" que agrupam outros conceitos. Por exemplo, "*Previous episodes*" agrupa os conceitos "*Date/time of previous episodes*", "*Details*", "*Comparison*". Estes conceitos "cluster" não foram traduzidos para uma formulação ontológica. Estes conceitos foram considerados para desambiguação dos conceitos hierarquicamente inferiores que recebem os valores introduzidos pelos usuários. Por exemplo, o conceito "*Body Site*", agrupado sob "*Localized Palpation*", refere-se especificamente à parte do corpo que esta sendo examinada por uma palpação.

Um extrato da ontologia que representa as entidades do openEHR segundo o realismo ontológico pode ser visto nas Figuras 13, 14 e 15. Foi adicionado o sufixo "Episteme" para as classes criadas especificamente para esta representação.

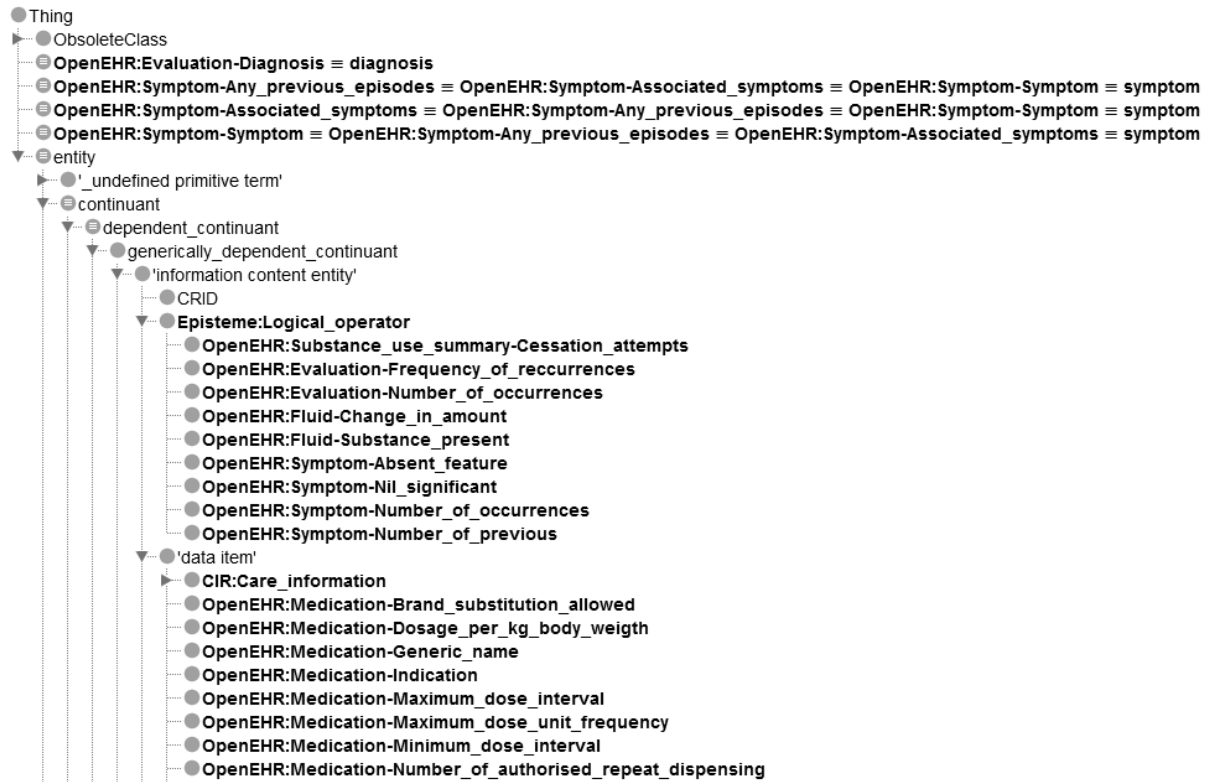


Figura 14 – Extrato da ontologia criada a partir dos conceitos avaliados dos arquétipos OpenEHR, baseado nas ontologias OGMS e IAO (parte 1)

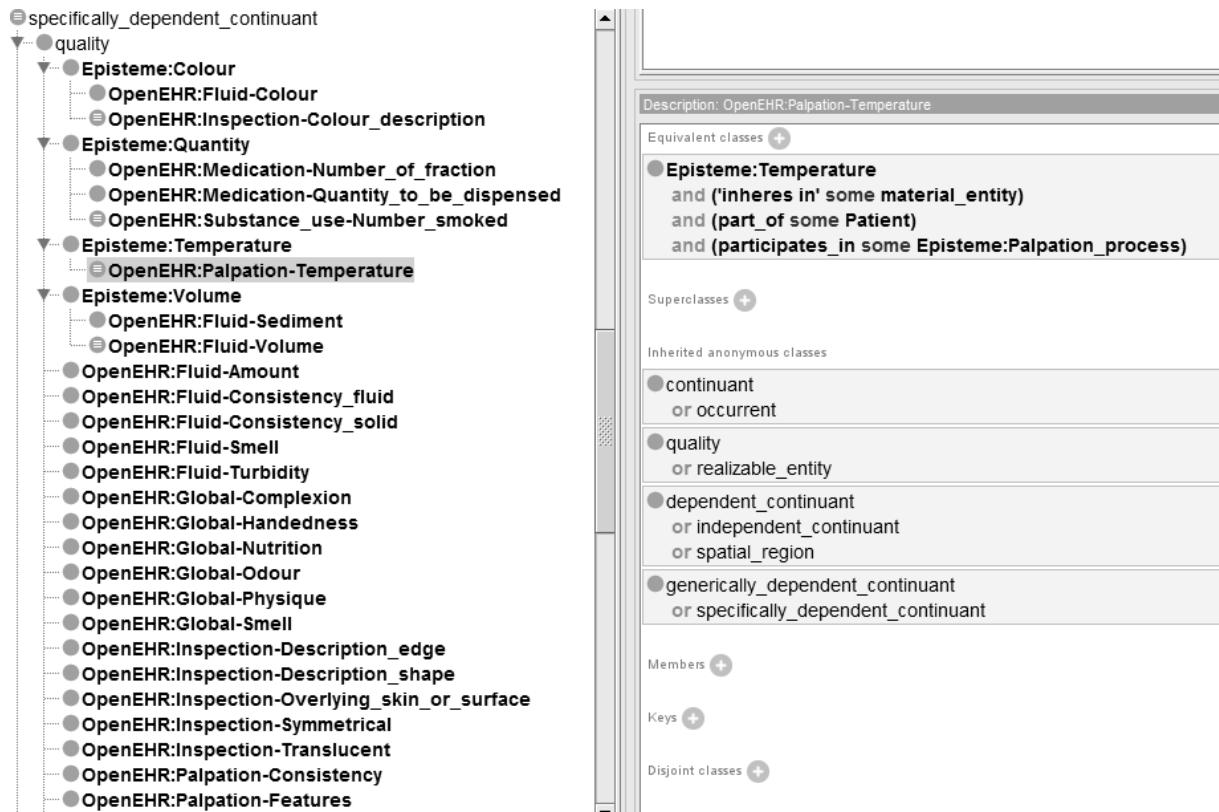


Figura 15 - Extrato da ontologia criada a partir dos conceitos avaliados dos arquétipos OpenEHR, baseado nas ontologias OGMS e IAO (parte 2)

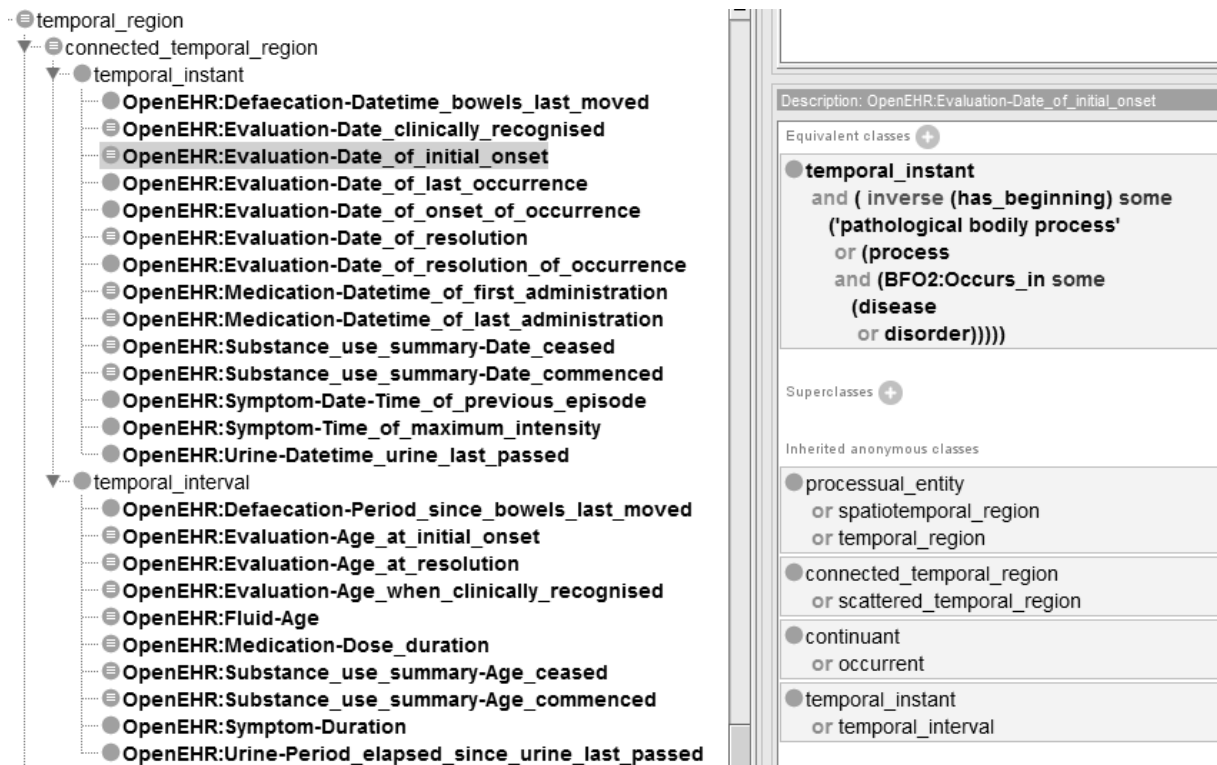


Figura 16 - Extrato da ontologia criada a partir dos conceitos avaliados dos arquétipos OpenEHR, baseado nas ontologias OGMS e IAO (parte 3)

Após a transcrição das expressões para o Protégé, os conceitos foram classificados segundo o arcabouço teórico (seção 5), que distingue quatro tipos de entidades: O (ontologia), A (anotações), I (informação) e R (raciocínio). Durante o mapeamento, apresentado na tabela 9, foram identificados ainda dois tipos especiais de entidades contidas nos arquétipos. A **primeira categoria** inclui operadores lógicos e matemáticos, cuja representação é feita a partir do uso da gramática da linguagem de representação. Nesta categoria se encaixam os operadores de negação e quantificação, e operadores não suportados na lógica descritiva, tais como afirmações sobre modalidade, definições que requerem acompanhar instâncias individuais (aumento da velocidade, em que a velocidade a cada momento é maior que a anterior e menor que a velocidade que se segue) e afirmações epistêmicas (ex. Eu acredito que X). Estas limitações estão fora do escopo da presente pesquisa, apesar de consistir em um problema relevante que deve ser investigado.

Nome do arquétipo	Entidades O	Entidades A	Entidades I	Entidades R	Operadores lógicos	Conceitos ambíguos
openEHR-EHR-CLUSTER.symptom.v1	9	7	0	5	4	4
openEHR-EHR-OBSERVATION.bodily_output-defaecation.v1	3	0	0	0	0	2
openEHR-EHR-OBSERVATION.urine_output.v1	3	0	0	0	0	2
openEHR-EHR-CLUSTER.fluid.v1	11	1	1	0	2	5
openEHR-EHR-EVALUATION.reason_for_encounter.v1	0	0	0	0	0	2
openEHR-EHR-OBSERVATION.global.v1	12	4	0	1	0	3
openEHR-EHR-CLUSTER.inspection.v1	8	1	0	0	0	3
openEHR-EHR-CLUSTER.palpation.v1	11	0	0	2	0	2
openEHR-EHR-CLUSTER.percussion.v1	4	0	0	0	0	1
openEHR-EHR-EVALUATION.substance_use_summary.v1	7	0	0	0	1	2
openEHR-EHR-OBSERVATION.substance_use.v1	4	1	0	1	0	1
openEHR-EHR-ITEM_TREE.medication.v1	14	2	10	5	0	2
openEHR-EHR-EVALUATION.problem-diagnosis.v1	8	6	0	3	2	6
TOTAL	94	22	11	17	9	35
Porcentagem do total	50%	12%	6%	9%	5%	19%

Tabela 9 – Classificação dos conceitos segundo as categorias do arcabouço teórico

A **segunda categoria** refere-se a conceitos ambíguos ou com definição parcial que impossibilita a categorização da entidade. Tais conceitos são muito comuns em formulários e sistemas classificatórios em medicina, e são melhor exemplificados pelas expressões "outros" ou "detalhes". Estas duas categorias representam 24% dos conceitos avaliados. A distribuição dos conceitos em cada categoria pode ser vista na Tabela 9.

Dentre os conceitos avaliados, um total de 50% foram adequadamente representados por ontologias realistas, ou seja, possuem um referente direto do lado do paciente. São representados, dentre outros, por:

- Entidades materiais: "*Anatomical location*" (*Symptom archetype*), "*Substance*" (*Substance Use summary archetype*)
- Qualidades: "*Physique*" (*Global Examination archetype*), "*Volume*" (*Bodily Defecation archetype*)
- Processos: "*Onset description*" (*Symptom archetype*), "*Mood description*" (*Global Examination archetype*)
- Regiões temporais: "*Date commenced*" (*Substance Use summary archetype*), "*Dose duration*" (*Medication archetype*)

Entidades R, ou seja, que se referem ao processo de raciocínio e avaliação do médico ou paciente, representam apenas 9% do total de conceitos. Para estes conceitos, não foi possível reconhecer nenhum referente do lado do paciente. Por exemplo, dois episódios de dor no peito são entidades realistas, mas a comparação entre os episódios é um processo mental que ocorre na mente do paciente ou médico. Da mesma forma, conceitos como indicação de uma medicação, ou razão para seu início são processos exclusivamente mentais.

Um total de 12% dos conceitos foram classificados como A, ou seja, possuem um referente indireto na realidade, mas são representados como construtos médicos. Esta categoria representa, portanto, entidades que são igualmente processos mentais e reais. Por exemplo, o conceito "Fatores precipitantes" refere-se a algum processo que inicia e causa o sintoma. Porém, para que o processo seja um fator precipitante, é necessário atribuir causalidade ou associação de acordo com o modelo mental do médico. Desta forma, um mesmo processo pode ser fator precipitante de acordo com um médico, e uma

associação espúria de acordo com outro médico, o que justifica a classificação como entidade A.

As entidades I estão eminentemente concentradas no arquétipo "Medicação", o único que se refere a um plano. Porém, é importante ressaltar que as mensurações objetivas não são o principal objetivo desta tese e devem ser investigadas em pesquisas futuras. As entidades informacionais em prontuários, como sinais vitais, dados laboratoriais e de imagem, foram intencionalmente ignorados nesta pesquisa devido à profusão de padrões atualmente disponíveis para sua representação.

A proporção de conceitos com diferentes classificações nos arquétipos encontrou variação considerável, já que algumas seções, e portanto alguns arquétipos, são consideravelmente mais epistemicamente orientadas que outras. Notavelmente, as seções "História da Doença Atual" e "Avaliação" tiveram a maior proporção de conceitos epistêmicos, classificados como entidades A ou R.

A classificação da ontologia pelo Hermit sugere que vários conceitos de arquétipos distintos referem-se ao mesmo tipo na realidade. O principal exemplo são: 8 conceitos que se referem a parte do corpo/localização corporal; 31 conceitos que se referem a qualidade com múltiplas referências a cor, cheiro, volume e massa; 24 conceitos que se referem a região temporal.

Ainda como consequência desta etapa, foi realizado um estudo sobre qualidades de processos, visto o grande número de entidades que se encaixam nesta categoria. O estudo foi realizado como um estudo especial durante o estágio sanduíche na Universidade Médica de Graz e contou com a colaboração de autores de diversas universidades. O artigo gerado ao fim do estudo (ANDRADE *et al.*, 2012b) pode ser encontrado no anexo 10.2.

6.3 IMPLICAÇÕES PARA A PESQUISA

Ontologias formais são bons instrumentos para tornar explícitos os comprometimentos ontológicos de modelos de representação. Na presente pesquisa, a análise ontológica dos conceitos representados nos arquétipos do OpenEHR, sob a perspectiva do realismo ontológico, revelou várias possíveis fontes de ambiguidade

normalmente despercebidas pelo especialista de domínio. Dentre estas, são particularmente pertinentes no escopo desta tese:

- Separação entre informação e realidade
- Separação entre o objeto e a interpretação sobre o objeto

Apesar da crescente tendência em separar informação e realidade na construção de modelos de informação, os achados da pesquisa sugerem que esta separação não é inteiramente respeitada durante o desenvolvimento dos arquétipos. Trabalhos anteriores sugerem que esta distinção é desconsiderada em algumas terminologias (SCHULZ *et al.*, 2010a) e modelos de informação (SMITH; CEUSTERS, 2006). A confusão pode ser vista já nas definições dos conceitos. Por exemplo, o conceito "Modifying Factor" é definido como "informação sobre um fator específico que altera o sintoma" ⁶⁴. Já sua subpropriedade "Factor" é definida como "um evento ou atividade que modifica o sintoma" ⁶⁵.

Seis por cento dos conceitos analisados referem-se a entidades informacionais. Estas entidades são esperadas em arquétipos que representam exames laboratoriais e de imagem, que contêm os resultados de mensurações químicas e das análises das imagens. Considerando que estes arquétipos não foram analisados, a proporção destas entidades é acima da esperada. Além disso, 34% dos conceitos são classes extremamente genéricas, sugerindo a ausência de separação explícita informação/realidade durante a construção. Por exemplo, o conceito "Clinical Description" permite a entrada de qualquer valor que descreva o sintoma. Em termos realistas, isto significa que a única coisa que se sabe sobre a entidade é que existe uma relação de intensionalidade, ou seja, que a entrada é a respeito do sintoma ("**is about**"). Tal constatação fica evidente durante a criação das definições lógicas dos conceitos, conforme pode ser visto no uso da relação "**is about**" no quadro 11, relativo ao arquétipo Sintoma.

⁶⁴ Information about a specific factor that changes the symptom

⁶⁵ An event or activity that modifies the symptom

Conceito	Definição em linguagem natural	Definição lógica
Clinical Description	Description of the symptom	'data item' AND 'is about' value Symptom001
Previous Episodes / Previous Episode / Details	Details of previous symptoms and comparison to this episode	'data item' AND 'is about' SOME (Symptom AND precedes VALUE Symptom001)

Quadro 11 – Relação entre conceitos classificados como ambíguos e definições em linguagem natural e lógica

Percebeu-se também ausência de separação clara entre entidades reais e a interpretação do médico sobre estas entidades. Apesar de avanços significativos da Medicina Baseada em Evidências, a prática médica é fortemente baseada na avaliação de sinais e sintomas, descritos em sete características (BICKLEY; SZILAGYI, 2009): Localização; Qualidade; Quantidade ou gravidade; Temporalidade; Contexto em que ocorre; Fatores exacerbantes ou aliviantes; Manifestações associadas. Informação contextual depende de condições que o médico considera relevante para confirmação ou exclusão de hipóteses diagnósticas. Causalidade é, portanto, de grande relevância para o discurso médico. Relações de causalidade e justificativa foram particularmente frequentes nos conceitos classificados como entidades A. É possível representar ontologicamente o processo que precede um sintoma, não o fato que acredita-se que este processo causa o sintoma. Comparações também demonstram dificuldade na separação da entidade e sua interpretação. Comparações são entidades puramente mentais, pois não adicionam nenhum ingrediente extra de existência às entidades comparadas (SMITH *et al.*, 2012). Por exemplo, saber que a dor no peito dura menos que a sensação de falta de ar não altera a duração destes dois processos.

Retornando à hipótese 1 da metodologia (seção 4.3.4), é possível perceber que existem limites para a representação exclusivamente realista do prontuário médico, demonstrado pelo número de conceitos classificados como entidade O.

A influência da linguagem natural nos modelos de informação gera, em parte, a dificuldade na representação ontológica dos conceitos. Como resultado, alguns pares conceito-valor dos arquétipos não podem ser representados como entidades, mas como operadores lógicos. Dos nove conceitos lógicos, foram encontrados cinco que refletem a cardinalidade de alguma entidade e três que refletem a existência ou inexistência da entidade – que poderia ser traduzido como o operador verdadeiro/falso, ou pela relação *lacks* (CEUSTERS *et al.*, 2007). O último conceito, *change in amount*, reflete a diferença entre duas quantidades. É importante notar que cardinalidade e negação são operadores suportados pela linguagem OWL2 e OWL-DL e, portanto, é possível representar o conceito através de regras.

A ausência de relações de causalidade, justificativa (descrição do raciocínio) e comparação foram também limitações significativas na representação da informação médica por ontologias realistas. É indiscutível que o estado ontológico destas relações é controverso (SCHAFFER, 2007; SMITH *et al.*, 2012), e um dos objetivos das ontologias realistas é fornecer raciocínio lógico robusto e coerente, o que não é garantido para entidades de natureza indefinida ou epistêmica. Ao mesmo tempo, o uso das relações em textos médicos é regular quando representam relações entre instâncias particulares. Afirmações de classe do tipo "todo X é causado por algum Y" e "todo Y causa algum X" não foram encontradas nos prontuários analisados, e provavelmente são incomuns em prontuários médicos. Tais afirmações não seriam, portanto, utilizadas em inferência lógica mas essencialmente para registrar fatos individuais.

Tal limitação para representação da informação epistemológica já havia sido reconhecida anteriormente (BODENREIDER *et al.*, 2004), o que justifica a existência de classes informacionais na OGMS, como diagnóstico e prognóstico. Propõe-se, aqui, estratégia semelhante para representação dos conceitos epistemológicos encontrados nos arquétipos. Desta forma, torna-se possível a representação de todas as entidades do prontuário. Porém, a definição lógica de artefatos de informação é necessariamente indireta (SCHULZ; KARLSSON, 2011). Por exemplo, o conceito "Symptom Severity" (gravidade do sintoma) pode ser representado como um tipo de "OGMS:Clinical Finding" (achado clínico), que por sua vez é uma entidade informacional. Esta representação traz consequências para a realização de inferências: a representação de uma classe de sintomas graves significa considerar todas os sintomas que, quando avaliados por um ser humano e representados como informação, esta informação é do tipo "Symptom Severity - Severe"

Considerando a hipótese 2 da seção 6, os achados demonstram que a análise ontológica dos conceitos reduz significativamente a ambiguidade.

As relações entre as entidades registradas no modelo de informação são esclarecidas durante a formulação de definições lógicas. Por exemplo, a localização de um sintoma relaciona um processo (sintoma) e a entidade material onde o processo ocorre, enquanto a localização de um paciente relaciona um paciente (entidade material) e a sala (região espacial) onde o paciente está localizado. A polissemia da palavra "localização" torna-se, assim, explícita.

A transcrição da ontologia no Protégé evidenciou diversas semelhanças entre conceitos de diferentes arquétipos, particularmente após a classificação automática pelo Hermit. O principal exemplo foi a categorização comum dos diversos conceitos relativos a partes corporais, como "Palpation/Body site localized palpation", "Symptom/Location in body" e "Medication/Site of administration". O poder inferencial pode ser usado para três fins principais:

1. Restringir as hierarquias que podem ser usadas como valor para os conceitos.
Ex. Um processo não pode ser uma localização
2. Alinhar diferentes modelos de informação e arquétipos. Ex. O conceito "Object" do arquétipo "Temperature" e o conceito "Body site" do arquétipo "Palpation", quando referenciando a temperatura do corpo.
3. Adquirir informação preliminar sobre uma classe de uma terminologia desconhecida. Ex. Qualquer valor após o conceito "Color description" é do tipo "Cor" (Qualidade).

Estas conclusões, apesar de eminentemente qualitativas, reforçam as características das ontologias realistas. Enquanto existem restrições para a criação de classes definidas exclusivamente pelo homem e inexistentes na realidade, a representação é robusta e independente da terminologia e conceitualização utilizada.

Os conceitos utilizados no OpenEHR exemplificam a consequência da utilização sem rigor da linguagem natural na criação de artefatos de organização da informação. Até mesmo as definições dos conceitos demonstram esta falta de rigor. Por exemplo, o conceito

"Form" do arquétipo "Medication" é definido como "a forma da medicação" ⁶⁶. A definição é circular, uma vez que não esclarece o que é forma. Além disso, depende do contexto para interpretação. "Forma" refere-se, usualmente, à configuração espacial ou aparência visual de um objeto. Neste caso, forma refere-se à formulação farmacêutica que traz, dentre seus componentes, a substância referida – ex. comprimido, solução, suspensão. O mesmo termo ("Form") reaparece no arquétipo "Substance Use", sendo definido como "identifica a forma de uma substância"⁶⁷, o que torna uma definição precisa impossível.

Por outro lado, algumas definições lógicas se mostraram arbitrárias. Um único evento médico possui vários componentes mas a distinção entre eles é desconsiderada na linguagem natural. Por exemplo, melena pode ser compreendido como a presença de células sanguíneas nas fezes (uma entidade continuante) ou como um processo de sangramento no intestino (uma entidade ocorrente). Argumento similar pode ser feito a respeito de terminologias médicas, como a descrição do câncer (continuante) e da história de câncer (ocorrente).

Justifica-se, assim, o grande número de conceitos classificados como ambíguos. Para estes casos, deve-se investigar futuramente se o uso de ontologia como ferramenta de modelagem permitiria maior clareza durante a construção dos arquétipos.

De forma similar, a falta de rigor na criação dos conceitos leva à existência de conceitos de nomes diferentes que se referem ao mesmo tipo de entidade na realidade, como *Smell* e *Odour* do arquétipo Global examination e as diversas referências a regiões corporais. É importante notar que estes exemplos foram encontrados apesar do pequeno número de arquétipos analisados e do critério de seleção destes arquétipos. Segundo o critério, foram selecionados arquétipos que representam partes independentes do prontuário médico. Em uma situação real, serão encontrados arquétipos com diferentes graus de sobreposição (ex. Um arquétipo para descrição de sintomas cardiovasculares e outro para descrição de sintomas pulmonares). Neste caso, espera-se um número ainda maior de conceitos semelhantes.

⁶⁶ The form of the medication

⁶⁷ Identify the form of the substance

7 REPRESENTAÇÃO ONTOLÓGICA DA INFORMAÇÃO MÉDICA

Como resultado dos passos descritos na seção 4.3.3, os fragmentos sentenciais foram representados de acordo com os conceitos respectivos dos arquétipos, em seguida foram representados diretamente na ontologia na forma de instâncias. Finalmente, relacionamentos restantes foram criados na ontologia. A Tabela 10 apresenta alguns exemplos do processo.

Fragmentos sentenciais	Conceito do arquétipo	Definição em OWL	Valor	Instância OWL
Edema +++/4+, leve calor, dor a palpação de panturrilha,	Clinical Description	Thing	Edema +++/4+	Patient1EdemaQuality Type Thing Patient1EdemaQuality Type EdemaQuality
	Findings/ Tenderness/ Category	ProcessProfile AND isProcessProfileOf SOME PalpationExamination	Superficial tenderness	Patient1Tenderness Type ProcessProfile Patient1Tenderness isProcessProfileOf Patient1PhysicalExamination
	Findings/ Tenderness/ Body site	(MaterialEntity) AND (partOf SOME Patient) AND (participantOf SOME PalpationExamination)	Left leg	Patient1LeftLeg Type MaterialEntity Patient1LeftLeg partOf Patient1
	Findings / Temperature	TemperatureQuality AND (inheresIn SOME (MaterialEntity) AND (partOf SOME Patient)AND participantOf SOME PalpationExamination)	warm	Patient1Warmth Type TemperatureQuality Patient1Warmth inheresIn Patient1LeftLeg

Tabela 10 – exemplo do processo de tradução dos prontuários em instâncias OWL a partir dos pares conceito-valor dos arquétipos

A maior parte das instâncias oriundas dos fragmentos sentenciais foi representada por ontologias clínicas (317 em 366, ou 86%). Dentre as instâncias não representadas, 51% expressam incerteza (diagnóstico não confirmado), 26% são avaliações subjetivas e 22% são representações causais ou justificativa do raciocínio médico.

O resultado obtido nas instanciação dos fragmentos revelou propriedades importantes do método, assim com as várias dificuldades em sua implementação. Para contemplar as nuances do processo, as dificuldades enfrentadas e as decisões tomadas, a

discussão está organizada em forma de questões, seguida pela exploração de possibilidades alternativas de solução.

7.1 DISCUSSÃO E QUESTÕES

No restante desta seção, são apresentados os diversos casos e questões encontrados durante o processo de representação ontológica do prontuário médico, seguidos da decisão tomada durante a pesquisa. Algumas questões são interdependentes mas, para fins didáticos, são apresentadas separadamente.

7.1.1 Questão 1 – Como lidar com afirmativas que excedem as capacidades da lógica descritiva sob a ótica realista?

Como explorado nas seções anteriores, a lógica descritiva possui várias vantagens para a representação e definição de conceitos. Seguindo a metodologia realista, ontologias representadas em lógica descritiva são robustas e a inferência é, do ponto de vista lógico, completa e decidível. Porém, durante a representação dos fragmentos na ontologia, limitações inerentes à linguagem foram encontradas.

Dentre os fragmentos, misturam-se expressões de modalidade, probabilidade, negação e operações matemáticas. Expressões de modalidade são modificadores de afirmações que expressam as condições nas quais a afirmação é verdadeira, como expressões de possibilidade e necessidade. Expressões de necessidade não oferecem obstáculo significativo para representação ontológica – quando inferidas pelo contexto do fragmento, podem ser representadas com axiomas em lógica descritiva. Ex. É necessário que o paciente seja um ser humano.

Diferentemente, expressões de possibilidade não podem ser diretamente representadas de forma fidedigna ao registro. Por exemplo, a expressão "ITU?" deve ser interpretada como "é possível que o paciente X tenha a doença *infecção do trato urinário* em atividade". Se representarmos esta afirmação com um axioma, corremos o risco de uma afirmação contrária no futuro, caso o diagnóstico seja excluído, o que cria uma

inconsistência no banco. De fato, esta questão é eminentemente epistêmica: o médico raciocina considerando que o paciente possui a doença e pedirá testes para excluir esta possibilidade.

Outra limitação encontrada na representação das instâncias é a dificuldade na expressão de operações matemáticas. A interpretação de diversos exames diagnósticos é baseado na comparação de resultados do paciente com padrões populacionais. Estas comparações são refletidas nos prontuários, com expressões como "anúria há mais de 24 horas", "queda de potássio (3,4)" e "usar medicamento de 15 em 15 dias".

Foram exploradas algumas soluções para estas limitações. A primeira opção explorada foi o uso de lógicas mais expressivas, como a lógica modal e a lógica epistêmica. Apesar do aumento de expressividade, estas opções foram descartadas por serem indecidíveis e requererem tempo excessivo para processamento.

A segunda opção estudada é a representação de hipóteses e operações como artefatos de informação da IAO, ou seja, a hipótese de infecção do trato urinário é um registro sobre a condição do paciente – sobre esta condição, porém, nada é afirmado. Uma vez que esta opção segue os preceitos do realismo ontológico, as inferências obtidas são corretas. Porém, esta saída claramente é um subterfúgio que não representa completamente a expressão. No exemplo mencionado, a busca por pacientes com infecção do trato urinário não retorna o paciente em questão, pois nenhuma patologia foi instanciada, apenas a hipótese diagnóstica. Portanto, existem casos em que essa representação não responde aos objetivos pragmáticos da busca.

Uma adaptação da representação de artefatos de informação é a criação de um axioma com a propriedade "is_about" e o quantificador de lógica descritiva ONLY. Neste caso, à entidade "hipótese de infecção do trato urinário" é adicionada a relação "is_about ONLY InfecçãoDoTratoUrinário". Esta adaptação garante que inferências incorretas não serão realizadas. O axioma afirma que a hipótese, caso diga respeito a alguma instância, diz respeito a uma instância do grupo infecção do trato urinário. Desta forma, não há a exigência que esta instância exista e afirmações contrárias não causam inconsistência no banco. Apesar disto, buscas por pacientes com a doença não recuperam estes casos, e a busca deve ser alterada para encontrar hipóteses que digam respeito ("is_about") apenas ("ONLY") a infecção do trato urinário. Esta foi a opção escolhida no presente trabalho.

Porém, as tentativas de recuperação dos dados representados demonstraram que esta representação exige que o pesquisador saiba exatamente o que está buscando.

Outra questão que se mostrou não trivial foi a representação e recuperação de afirmações negativas, considerando o *Open World Assumption*. As questões que visaram recuperar asserções negativas – ex. Diagnósticos alternativos, ou diagnósticos que não são de um tipo específico, no caso, Trombose Venosa Profunda – não recuperaram as instâncias corretas. Para inferir que o diagnóstico 001 de Fratura da Perna não é um diagnóstico de Trombose Venosa Profunda, é necessário um axioma de disjunção de classe ou instância:

Diagnóstico Equivalent_To Diagnóstico_de_Fratura OR
Diagnóstico_de_Trombose_Venosa_Profunda

Ou:

Diagnóstico001 Type NOT(Diagnóstico_de_Trombose_Venosa_Profunda)

É importante apontar que, ao fim da tese, uma terceira opção foi avaliada: o controle externo da interpretação. Apesar de todas as vantagens da representação lógica e, particularmente, da lógica descritiva, a linguagem não é adequada para a representação de regras (como as operações matemáticas). Estas regras são mais bem tratadas por linguagens procedurais, como a maioria das linguagens atuais de desenvolvimento de software. Certas linguagens desenvolvidas para a internet, como o SWRL (HORROCKS *et al.*, 2004) e o SPIN (KNUBLAUCH *et al.*, 2011) também são capazes de representar as regras, contando com ferramentas robustas e velozes capazes de realizar o processamento.

A dificuldade na representação e busca de registros com afirmações de possibilidade também sugeriu que controles externos permitem mais maleabilidade ao modelo, ainda que adicionem dificuldades de interoperabilidade ao ignorar as definições lógicas estritas. Por exemplo, ao marcar todas os axiomas que representam hipóteses, é possível filtrar estes axiomas se a robustez da inferência é necessária, ou permitir caso exista necessidade de uma busca mais abrangente no banco de dados. Voltando ao caso em questão, o axioma "CondiçãoDoPacienteX Type InfecçãoDoTratoUrinarío" pode ser considerado para inferência, ou retirado da base de conhecimento. O mesmo vale para afirmações negativas. Neste

caso, por exemplo, é simples recuperar todas as instâncias e criar regras específicas para filtrar as instâncias desejadas.

7.1.2 Questão 2 – As afirmativas devem ser representadas como classe ou como instância?

A obtenção de inferências é um dos principais objetivos da representação do conhecimento médico através de ontologias, particularmente quando representadas em linguagens lógicas. É desejável, por exemplo, que um registro de "Dor Torácica" seja igualado a um registro de "Dor localizado_em Tórax". Um dos principais problemas encontrados é a complexidade do domínio médico, que faz com que esta tarefa seja impraticável devido ao número de classes e relacionamentos (BODENREIDER *et al.*, 2007; SCHULZ *et al.*, 2008; SCHULZ *et al.*, 2011). Esta responsabilidade é normalmente atribuída a terminologias externas, particularmente àquelas representadas logicamente, como o SNOMED CT. Desta forma, as terminologias forneceriam as classes e suas definições lógicas (relacionamento entre as várias classes) que seriam instanciadas na representação do prontuário.

Porém, a pesquisa mostrou que a separação entre classe e instância não é trivial. Ao transformar as afirmativas em pares conceito-valor dos arquétipos em instâncias OWL, percebeu-se ausência de distinção entre instâncias e classes. De forma geral, a relação entre terminologias e modelos de informação é, de fato, uma ligação entre classes reais e classes informacionais, respectivamente. Em outras palavras, quando o paciente relata dispnéia, existe uma informação registrada do tipo *Informação sobre Sintoma* que se refere a algum elemento da classe *Dispnéia*. Os conceitos remanescentes, porém, podem se referir ao episódio específico (ex. *Momento de intensidade máxima*), ao conjunto de ocorrências individuais (ex. *Número de ocorrências*) ou ao padrão ou tipo do sintoma típico (ex. *Fatores precipitantes*). Na prática, isto significa que não é possível saber a priori se cada afirmativa deve ser representada por:

- uma instância: Paciente001 contains_process Sintoma001
- uma regra: Paciente001 Type contains_process EXACTLY 4 Symptom
- uma restrição de classe: Sintoma001 Type preceded_by SOME Exercício

Uma vez que nessa pesquisa toda representação ontológica foi realizada manualmente, duas soluções foram adotadas para representar fidedignamente os fragmentos sentenciais dos prontuários:

1. Resolução ad hoc das ambiguidades utilizando o conhecimento de domínio do autor da tese;
2. Representação das entidades como um tipo de restrição dos artefatos informacionais: ao definir que o registro deve ser sobre um tipo específico de entidade, é possível inferir o significado real da afirmativa sem criar novas instâncias. Por exemplo, uma prescrição neste formato é representada como:

```
'is about' some (plan and
(is_realized_by only (medication_administration and
(has_participant some ('Processed Material' and
(has_part some (paracetamol and
(bearer_of some (mass and
( denotedBy some ('mass measurement datum' and
(has_part value miligram) and ('has measurement value' value 500)))))))))
and
(has_participant value patient001))))
```

Estas soluções possuem impacto direto na forma de representação e recuperação da informação. Por um lado, a resolução ad hoc deve ser feita no momento de registro da informação e, portanto, a modelagem dos registros eletrônicos em saúde deve considerar que não há uma única forma de representar pares conceito-valor. Em outras palavras, a modelagem dos prontuários eletrônicos deve permitir as diferentes relações entre conceitos e valores, como relações de instanciação, regra ou restrição.

Como impacto na busca, deve-se notar que ao representara informação como artefato de informação (versus a instanciação direta da entidade), a busca deve procurar especificamente por artefatos de informação. Por exemplo, uma dor torácica pode ser representada por, pelo menos, quatro instâncias distintas:

1. DorTorácica1 TYPE DorTorácica
2. DorTorácica2 TYPE InformationArtifact AND isAbout SOME DorTorácica
3. DorTorácica3 TYPE InformationArtifact AND isAbout ONLY DorTorácica
4. DorTorácica4 TYPE InformationArtifact; DorTorácica4 isAbout DorTorácica1

A representação da informação como nova camada de referência impacta nas inferências, como demonstrado na Tabela 11.

Expressões em lógica descritiva	Instâncias recuperadas?
DorTorácica	1
Dor AND locatedIn SOME Torax	1
InformationArtifact AND isAbout SOME DorTorácica	2, 4
InformationArtifact AND isAbout SOME (Dor AND locatedIn SOME Torax)	2, 4
InformationArtifact AND isAbout ONLY DorTorácica	3
InformationArtifact AND isAbout ONLY (Dor AND locatedIn SOME Torax)	3

Tabela 11 – Relação entre a forma de representação das afirmativas (instâncias) e as expressões que recuperam estas instâncias

7.1.3 Questão 3: Como representar relações temporais?

Como apresentado na seção 3.1.4, lógica descritiva restringe a representação de relações com aridade maior que 2 (relações ternárias, quaternárias, etc.). Por exemplo, não é possível descrever aceleração da frequência cardíaca através de entidades temporais e processos, já que aceleração é uma comparação entre duas relações (batimentos por minuto) sequenciais no tempo. Porém, para a representação das informações do paciente, a questão mais importante refere-se à representação temporal das instâncias representadas, ou seja, a relação entre a dor torácica e o exercício moderado que a precede e a causa. No relato médico, todas as afirmações são indexadas pela data do evento médico, complementadas por afirmações no texto do prontuário (quadro 12).

Paciente: José da Silva Data:01/01/2011

Na última semana, o paciente relata dor no peito precedido por exercício moderado.

Quadro 12 – Indexação da afirmação no tempo

Ao aderir aos princípios rigorosos e básicos da abordagem BFO, é necessário representar relações complexas e fundamentais, muitas vezes estranhas à primeira vista. Por exemplo, a relação entre um processo e o momento de sua ocorrência requer a distinção entre os espaços temporais ocupados pelo processo, a instanciação do instante que precede todos os outros deste espaço e a informação que descreve este espaço. Então a chegada de um paciente (processo) ocupa um espaço temporal, cujo primeiro instante é denotado pelo número "11h30 do dia 27/01/2013, horário de Brasília". Ao distinguir diferentes tipos de dados como diferentes tipos de resultados de mensuração, torna-se difícil fazer a conexão direta entre o valor e seu lugar na ontologia. Ainda com o exemplo de afirmações temporais, a duração de um sintoma é definida como:

SymptomDuration **EquivalentTo** *temporal_interval* and (**isDurationOf** some symptom)

Porém, ao invés de receber a classe Duração_Um_Mês, que seria uma subclasse de Duração_do_Sintoma, o valor encontrado é um dado pré-codificado segundo algum padrão internacional que é equivalente a 1 mês. A representação ontológica perde a capacidade de operações matemáticas simples, não apoiadas por classificadores OWL. Enquanto OWL possui definições embutidas para lidar com tais valores (propriedade de dados), ontologias descendentes da BFO precisam conectar a classe ao seu resultado, o que leva ao seguinte padrão de conversão:

'*time measurement datum*' **SubclassOf** 'has measurement value' min 1 Literal AND 'has measurement unit label' min 1 *Thing*

SymptomDurationMeasurement **EquivalentTo** 'time measurement datum' AND 'is about' SOME *SymptomDuration*

Symptom

Symptom: Dyspnea

Duration: 1 month

SymptomX **Type** *Dyspnea*

SymptomX **Type** **hasDuration** ONLY (*temporal_interval* AND **inverse** ('is about') ONLY (*SymptomDurationMeasurement* AND 'has measurement value' value 1 AND 'has measurement unit label' value month))

A avaliação do caso de uso Trombose Venosa Profunda reforçou algumas vantagens e limitações para a representação ontológica da informação médica. As questões que não exigiram relacionamentos temporais complexos entre sintomas ou solicitações de exames foram recuperadas adequadamente, sugerindo o potencial do uso de ontologias para melhoria do registro de entidades clínicas complexas, em situações nas quais a simples anotação por terminologias é insuficiente.

Tempo é um aspecto essencial na avaliação do cuidado clínico – não apenas é necessário saber o que aconteceu, mas quando aconteceu e como o evento se desenrolou. Este problema pode ser desdobrado em duas partes. A primeira é uma questão de modelagem. Existe algum caso de uso em que a separação entre um instante e sua mensuração é relevante? Em outras palavras, faz sentido diferenciar as expressões `instante001` e `2005-10-30T10:45UTC`? A segunda parte é uma questão tecnológica. Optou-se neste pequeno ensaio em utilizar as tecnologias homologadas e recomendadas pela W3C para representação dos dados. Neste caso, seria necessário recorrer a regras definindo precedência (`IF time(x) < time(y) THEN x precedes y`) em linguagens de regras, como SWRL, ou delegar a inferência temporal ao software que usa a ontologia. Porém, as instâncias podem ser representadas de várias formas alternativas, como bancos de dados relacionais, utilizando a ontologia como esquema do banco de dados. Tal distinção seria, então, simplesmente desconsiderada sem prejuízos relevantes para o sistema.

Durante o processo de modelagem e busca dos fragmentos, a indexação por tempo mostrou-se um desafio para modelagem. Poucos fragmentos foram explicitamente indexados e, quando havia menção explícita temporal, a afirmativa era vaga ou inespecífica ("por volta de 1 mês", "história de cirurgia"). Esta dificuldade é evidenciada na análise do caso da Trombose Venosa Profunda. Das 5 (cinco) respostas incorretas, 4 (quatro) foram relacionadas à falta de marcadores temporais. Na prática, o médico compreende a questão temporal utilizando operações matemáticas com as datas e assumindo que a data do registro (data do evento médico) pode ser considerada como referência para todos os acontecimentos não indexados.

Três soluções surgiram durante a realização da tese. A primeira é o uso de reificação de relações ternárias, dentre eles os processos temporais (GREWE, 2010). Considerando o exemplo do paciente João, que tinha a perna quebrada no dia 01/05/2013 ("PernaQuebrada locatedIn João at tempo01/05/2013"), a expressão em lógica descritiva é transformada em:

João participantIn FraturaDaPernaDoJoão(Processo)

PernaQuebrada participantIn FraturaDaPernaDoJoão(Processo)

Em seguida, bastam axiomas que afirmam que o espaço temporal ocupado pela fratura(processo) contém o intervalo de 01/05/2013. Esta solução resolve a questão da representação ternária, mas sugere a criação de múltiplos processos que eventualmente levariam ao descontrole da ontologia. Afinal, o médico pode se referir ao tumor ou ao processo de formação do tumor, assim como é possível referir uma pessoa ou a vida desta pessoa. Além disso, a principal dificuldade seria o uso das operações temporais, como ordenar os eventos que se sucedem ou que possuem interseção.

Outra possibilidade é a utilização de uma ontologia de relações temporais como o OWL-Time. Os critérios para inferência temporal em lógica de primeira ordem são bem descritos pelas relações temporais de Allen (ALLEN; HAYES, 1990). O uso destas relações é, no entanto, impossível neste caso visto que seria necessário determinar a interrelação temporal entre os processos descritos no prontuário (o esforço precede a dispnéia, que precede a chegada do paciente ao hospital, que precede a solicitação do exame), o que não é factível. Esta solução requer ordenação precisa dos eventos através de axiomas os que relacionem temporalmente.

O insucesso obtido na representação temporal dos fragmentos dos prontuários sugere que aqui também o controle externo é uma opção plausível, pelo menos para a recuperação de instâncias. No outro extremo de complexidade de representação temporal, os sistemas que implementam modelos de informação como o openEHR indexam temporalmente cada instância com o momento de sua criação. Adicionalmente, permitem que o autor adicione propriedades que definem precisamente o momento em que a condição é verdadeira. Por exemplo, a fratura do paciente João é representada (simplificadamente) como:

Documento 1 – Data: 02/05/2013

- Paciente João
- Diagnóstico: Fratura da perna
 - Quando: Ontem

Sugere-se aqui adotar futuramente uma abordagem intermediária que permita a indexação temporal de ocorrentes a partir do início até seu fim, e a indexação de continuantes do início de sua existência até o fim da existência. A reificação de relações entre continuantes e qualidades garante que toda expressão pode ser desta forma indexada. Por exemplo, substituir a expressão "A pele ficou vermelha durante 1 hora" por "a relação entre a pele e a cor vermelha teve início às 14h00 e fim às 15h00.

7.1.4 Questão 4: Como representar entidades que podem ser descritas, pelo médico, de várias formas iguais para o raciocínio, mas diferentes ontologicamente

Existem vários exemplos de entidades que podem ser descritas como processo, como continuante e como uma observação. Esta questão, mencionada durante a discussão de relações temporais, foi um achado frequente durante a representação dos fragmentos. Por exemplo, eu posso descrever um tumor, o processo de formação do tumor, uma radiografia evidenciando um tumor ou uma anamnese descrevendo o adoecimento do paciente por um tumor. Não se trata, portanto, de sinonímia, mas de aspectos ontologicamente diferentes de um mesmo estado de coisas. Nos fragmentos, esta questão foi representada por:

"Paciente iniciou quadro de anúria (...) após medicamentos, voltou a urinar" (disposição / processo)

"Presença de dois picos febris -38,7º e 39,6º" (processo / informação)

"Episódio de fezes escurecidas" (processo / continuante)

"Hepatomegalia⁶⁸ – fígado palpável a 4 cm do rebordo costal direito" (continuante / processo)

Esta situação, apesar de frequente, normalmente é controlada com o uso de uma terminologia médica, como o CID-10, SNOMED CT, dentre outras. Faz parte do processo de

⁶⁸ Hepatomegalia: Fígado aumentado acima do tamanho esperado

modelagem do sistema padronizar os termos utilizados e, certamente, esta questão surgiu devido à liberdade proporcionada pelo registro em linguagem natural.

Além da padronização, existem outras possibilidades para contornar o problema como o registro de múltiplas entidades registrar todas as entidades de um estado de coisas, sejam processos, coisas ou itens de informação. Desta forma, as inferências e a busca por instâncias são garantidas, mantendo-se a robustez, já que todas as afirmações são verdadeiras.

Para o caso de terminologias em que a padronização absoluta mostra-se de difícil implementação, sugere-se o uso da abordagem por situações (SCHULZ; KARLSSON, 2011). Uma avaliação mais abrangente sobre o assunto foi desenvolvida pelos autores desta tese e colaboradores (trabalho sob avaliação editorial) e será disponibilizado futuramente.

7.1.5 Questão 5: Como representar atributos (qualidades) de processos

Outra questão abordada na tese é a representação de atributos de processos. Atributos de processo são entidades que descrevem modificadores para determinados processos, como por exemplo, afirmar que a gravidez (processo) é de alto risco, ou que o batimento cardíaco (processo) é acelerado. Este assunto gerou um artigo (ANDRADE *et al.*, 2012b), adicionado no anexo 2, que discute amplamente o assunto.

7.1.6 Questão 6: Como representar ontologicamente entidades mentais?

Durante o desenvolvimento da pesquisa, foram encontrados termos utilizados no prontuário médico que não possuem representação clara na abordagem realista, ou seja, não podem ser definidos como continuante ou ocorrente (ANDRADE; ALMEIDA, 2011b). Durante a representação dos fragmentos dos prontuários médicos, estas entidades mentais e de raciocínio (entidades A e R, segundo a tipologia da seção 5) foram novamente percebidas. Os tipos recorrentes encontrados são:

Ato linguístico: ordens e recomendações, como prescrições médicas. Ex. "Tomar 1 comprimido pela manhã".

Causalidade: comum tanto na descrição ontológica das classes (seção 6.2) quanto na descrição de caso do paciente. Ex. "Paciente relata dor no peito causada por exercício moderado".

Julgamento: o prontuário médico tem, como um de seus fins, o objetivo de comunicação entre os médicos do paciente. Portanto, foram encontrados termos que agregam e resumem uma avaliação heurística do médico, ainda que não tenham uma correlação direta com alguma entidade específica da realidade. Por exemplo, no fragmento "Paciente hemodinamicamente instável, apresentando alto risco para cirurgia", a primeira afirmação tem um referente na realidade mas a segunda afirmação representa uma interpretação do médico sobre o paciente. Frequentemente, estes termos possuem uma definição consensual na literatura médica, apoiando-se em *scores* ou critérios definidos por *experts*. Um exemplo pode ser visto no score de Borg para avaliação de dispnéia (falta de ar):

SCALE	SEVERITY
0	<i>No Breathlessness* At All</i>
0.5	<i>Very Very Slight (Just Noticeable)</i>
1	<i>Very Slight</i>
2	<i>Slight Breathlessness</i>
3	<i>Moderate</i>
4	<i>Some What Severe</i>
5	<i>Severe Breathlessness</i>
6	
7	<i>Very Severe Breathlessness</i>
8	
9	<i>Very Very Severe (Almost Maximum)</i>
10	<i>Maximum</i>

Tabela 12 – Escala de Borg para avaliação de dispnéia (falta de ar) pelo paciente

Relação entre sinal e significante: conforme já identificado na seção 5, a prática médica se apoia na semiologia, ou seja, no estudo de sinais e sintomas para identificar um problema de saúde subjacente. Enquanto equivalentes para os médicos, não existe relação

direta, na realidade, entre sinal e significante. Por exemplo, o fragmento "fígado palpável a 3 cm do rebordo costal direito" pode ser compreendido como equivalente de "hepatomegalia à palpação".

Entidades epistemicamente objetivas: devido à importância da semiologia, são consideradas, para o raciocínio clínico, várias entidades que não podem ser empiricamente demonstradas. O exemplo clássico é a descrição da dor sentida pelo paciente, que é descrita como "dor em queimação" ou "dor em aperto".

Nos fragmentos analisados, foram encontradas afirmativas que atribuíam status epistêmico a outras afirmativas. Esta questão é intimamente relacionada à questão 2, não por limitações da linguagem lógica, mas pela existência de um conhecedor referente. Por exemplo, o fragmento "angina instável?" é composto por uma afirmação "paciente apresenta angina instável" e outra que diz que o autor desta afirmação suspeita que ela seja verdade. Outro exemplo é o fragmento "paciente não sabe informar medicação prescrita no hospital", composto por "a medicação prescrita ao paciente é X" e "paciente desconhece X".

A questão da intrusão epistêmica em prontuários médicos foi um dos grandes motivadores deste trabalho. Ao longo do desenvolvimento do trabalho, foram realizadas várias reuniões para discussão, com menção especial para as reuniões ocorridas durante o projeto Semantic Health Net. Nestas reuniões, diversas alternativas foram sugeridas, com suas vantagens e desvantagens. Porém, uma comparação pragmática mostrou-se particularmente difícil na ausência de um caso de uso evidente. Discutem-se aqui algumas das soluções com suas vantagens e desvantagens.

A primeira solução parte da estratégia de representar todas as afirmações como artefatos de informação sobre o paciente. Neste caso, o status epistêmico é atribuído à entidade de informação (asserção). Ou seja, toda entidade de informação é uma entidade epistêmica – algumas possuem referente na realidade, outros são apenas entidades mentais. Esta forma de representação mantém a fidelidade à realidade e não gera inferências incorretas. Ao mesmo tempo, nenhum axioma relaciona a entidade de informação à situação na realidade que está sendo descrita. Por exemplo, ao representar "AltoRiscoParaCirurgia" como entidade informacional, nenhum axioma relaciona esta entidade à entidade "Cirurgia".

Outra solução, já mencionada na questão 2, a restrição através da relação "only" é defendida por Schulz e Karlsson (2011) como uma forma de representar fidedignamente o conteúdo da informação sem a necessidade de instanciação de algo que pode não existir na realidade. Utilizando esta estratégia, não foram obtidas inferências incorretas nem a necessidade de invalidar axiomas anteriores (ex. hipóteses diagnósticas que se demonstram posteriormente incorretas). Assim como demonstrado na questão 2, a recuperação destas instâncias requer adaptação da expressão de busca. O relacionamento entre a informação e a entidade descrita é necessariamente a relação "isAbout". Desta forma, tal solução requer a criação de uma nova ontologia de entidades mentais, como julgamentos, entidades de interpretação, entre outras.

Considerando a ontologia de entidades mentais, a criação de uma ontologia não realista permitiria a representação plena destas entidades. A questão que se coloca é a ausência de um critério claro para orientar a construção de tal ontologia. A multiplicação de ontologias disponíveis na internet demonstra que esta falta de critério é um grande risco de criar silos, impedindo a interoperabilidade de sistemas. De toda forma, a constância de alguns termos encontrados nos fragmentos dos prontuários e nos arquétipos sugere que a padronização de alguns termos é uma opção viável que não compromete significativamente a robustez da ontologia. Alguns exemplos são avaliações de gravidade da doença, julgamentos de risco e relações de causalidade. Por exemplo, o aumento da gravidade da febre não tem nada em comum com o aumento de gravidade da dor torácica, mas a existência de um atributo "gravidade" como atributo de um sintoma ou doença é percebida em quase todas as terminologias médicas, o que facilita seu uso pelo médico e não prejudica a recuperação da informação.

Finalmente, existe a opção de simplesmente não representar entidades puramente epistêmicas na ontologia, ficando a representação destas em separado. De fato, uma das principais lições aprendidas durante as reuniões e testes de representação de instâncias é que reconhecer as limitações das ontologias pode ser mais importante que encontrar soluções na mesma metodologia. Os casos de uso de interoperabilidade em medicina são heterogêneos e parece improvável, no atual estado da ciência, a existência de uma única metodologia que responda a todas as necessidades. Assim como outras técnicas de engenharia, os usuários da técnica devem compreender seus benefícios e limitações e definir, caso a caso, para qual situação e objetivo esta se aplica.

7.2 CONCLUSÃO

A grande maioria dos fragmentos sentenciais foi representada adequadamente na ontologia. Foram encontradas diversas dificuldades durante a conversão dos dados do prontuário em uma representação ontológica. Os problemas mais imediatos se referem ao tratamento complexo de diferentes tipos de dados por ontologias realistas. Foram, também, encontradas limitações inerentes à tecnologia utilizada (implementações baseadas em lógica descritiva) e mesmo inerentes à maleabilidade de interpretação da linguagem natural por seres humanos, que ainda está longe de ser imitada por sistemas automatizados.

O objetivo da tese de *avaliar a representação ontológica de dados do prontuário, a partir da estruturação do conhecimento contido em modelos de informação médicos*, foi cumprido. Durante a pesquisa, foram aventadas quatro hipóteses principais:

- H1 - Ontologias podem expressar o mesmo conteúdo de modelos de informação
- H2 - A representação ontológica reduz ambiguidade conceitual e permite inferências robustas
- H3 - É possível recuperar informação de saúde representada ontologicamente através de operações em lógica descritiva
- H4 – O processo de representação da linguagem médica por ontologias é custo-efetivo no desenvolvimento de sistemas de informação em saúde

A pesquisa demonstrou que a resposta a estas hipóteses não é binária e que a respostas se sobrepõem em alguma medida. Sobre a hipótese 1, a grande maioria dos conceitos foram representados adequadamente em ontologias realistas. Para os conceitos cuja representação foi questionada, foram encontradas soluções para representar cada um dos conceitos dos modelos de informação em ontologias. A hipótese 2 sofre impacto direto destas soluções. Vários conceitos, em especial aqueles que lidavam com incerteza, processos de raciocínio e operações matemáticas, não puderam ser utilizados em representações lógicas. Ou seja, é possível representar uma "Complicação de cirurgia" como um artefato de informação, mas não é possível afirmar que toda complicação de cirurgia é **precedida** por uma cirurgia e **causada** por ela. Para que a hipótese 2 seja

verdadeira, o escopo da representação ontológica deve ser limitado conforme sugerido no arcabouço teórico, discutido na seção 5.

A avaliação da hipótese 3 evidenciou outras limitações importantes. Novamente, a recuperação foi bem sucedida para a maior parte das informações de saúde estudadas. Foi possível demonstrar, adicionalmente, que entidades comumente descritas como atributos de processos podem ser representadas adequadamente em ontologias. Porém, foram encontradas limitações na recuperação de informações temporais e mentais. Enquanto a primeira é um reflexo da dificuldade de representar logicamente as operações temporais, a segunda é uma consequência da refutação da hipótese 2.

Finalmente, os achados sugerem que a hipótese 4 é apenas parcialmente verdadeira. Além do limite da eficácia do método na recuperação de informação de saúde, o esforço necessário para a definição precisa das entidades descritas no prontuário é significativo. Ontologias formais de alto nível definem axiomas fundamentais, como a definição de espaço temporal. Do ponto de vista prático, estes axiomas exigem esforço inferencial desnecessário – é pouco relevante se uma dor torácica possui interseção espacial com uma dor da coluna torácica, uma vez que a coluna está localizada no tórax.

No entanto, é importante ressaltar que as ontologias desenvolvidas de forma robusta têm grande potencial de re-uso e modularização. Desta forma, a construção de sistemas de informação baseados em ontologias se beneficia do conhecimento a priori encontrado nas ontologias já existentes. Além disso, o uso de ontologias revelou ambiguidades na modelagem clínica que só seriam percebidos quando o sistema apresentasse as inevitáveis inconsistências. A análise dos arquétipos mostrou informações duplicadas em vários artefatos diferentes – o custo deste erro, por si só, justifica o trabalho de modelagem ontológica, independentemente da construção dos artefatos para inferência automatizada.

Considerando a hipótese 4, a melhor resposta encontrada no decorrer do trabalho é: depende do caso de uso. Os desenvolvedores devem considerar, principalmente, o tamanho e complexidade do sistema. Quanto maior e mais complexo, maior o ganho no uso de ontologias, seja para modelagem do sistema ou para uso dos artefatos distribuídos pelo sistema.

8 CONSIDERAÇÕES FINAIS E CONTRIBUIÇÕES DA TESE

O objeto desta tese é bastante amplo – a representação da linguagem de um prontuário médico utilizando ontologias. Enquanto é óbvio que a solução perfeita para a estrutura e organização da informação médica ainda está longe, espera-se que a presente tese tenha contribuído para o mapeamento e melhoria do estado da arte.

Foi realizada uma revisão de literatura sobre padrões de interoperabilidade em medicina, ontologias como artefatos de sistemas informatizados e como disciplina filosófica, e linguagem e comunicação. A parte empírica considerou modelos de da informação médica (arquétipos) e fragmentos sentenciais de prontuários, utilizando o realismo ontológico como arcabouço para criar uma representação ontológica de modelos e afirmações sobre pacientes.

Durante a construção desta tese, ficou claro que a utilização de ontologias para representação da linguagem médica utilizada em prontuários tem vantagens e limitações de acordo com os usos pretendidos desta informação. Desta forma, sugere-se aqui que as ontologias não devem ser encaradas como dogma, mas sim como uma ferramenta que, em condições corretas, é extremamente útil. Para auxiliar o leitor nesta tarefa, pretendeu-se nesta tese deixar algumas contribuições que esclarecem as situações nas quais as ontologias podem ser usadas.

A primeira contribuição do estudo (seção 5) é a compreensão dos termos médicos do ponto de vista empírico e filosófico. A linguagem técnica utilizada por profissionais de saúde é adquirida durante um período educacional longo e intensivo, durante o qual eles aprendem como pensar e comunicar seu pensamento. Como consequência, não é surpreendente que vários termos refiram o processo mental, seja através de asserções puramente epistêmicas (eu acredito que X), relações causais (eu acredito que X é causado por Y), descrição da justificativa ou raciocínio (eu fiz X porque Y) e interpretação subjetiva de eventos ou objetos (minha impressão de X é Y). Além disso, como processadores de informação, profissionais de saúde raramente distinguem entre a realidade per se e informação sobre a realidade, e falam sobre "concentração de hemoglobina baixa" e "Hb: 8mg/dl" como sinônimos. Para entender estes termos e como eles devem ser representados na ontologia, foi apresentado um arcabouço baseado na teoria dos três mundos de Karl Popper, que pretendeu:

- Reconhecer o que deve e o que não deve ser representado em uma ontologia robusta para representação de um domínio
- Explicitar quais entidades referem algo na realidade e quais referem informação sobre a realidade
- Mostrar que entidades em diferentes níveis da realidade são definidas de formas diferentes.

A segunda contribuição da tese (seção 6) foi o alinhamento entre modelos de informação em saúde e ontologias realistas. Levando em conta a abordagem de Popper, foi possível identificar as entidades representadas nos modelos de informação. Foi possível identificar, também, a ambiguidade relativa à distinção realidade informação e a ausência de definição adequada de vários termos. Por exemplo, ao entender que "Severidade" refere-se a uma avaliação e não a um tipo, fica claro que este termo não pode ser definido utilizando classes ontológicas. Não se trata aqui de uma avaliação sobre a frequência e importância destes termos – o que se argumenta é que estes termos não devem ter representação ontológica, mas devem ser representados de alguma forma.

A terceira contribuição (seção 6 e anexo 10.2) é uma tentativa de expandir o escopo de ontologias realistas. Análises anteriores mostraram que a forma com que processos são descritos em prontuário médicos é incompatível com a descrição de processos em ontologias. Portanto, foi proposta uma nova forma de criar classes definidas que descrevem processos sem alterar as premissas básicas da BFO, que foi chamada de Process Attributes (atributos de processo). Tal empreitada foi motivada por anúncios sobre o lançamento de uma nova versão da BFO – a BFO2.0 (SMITH *et al.*, 2012) que contempla tais classes, chamadas nesta ontologia de Process Profiles (SMITH, 2012).

A quarta contribuição (seção 6) foi o mapeamento entre diversos arquétipos de um modelo de informação e ontologias realistas. Ainda que tenha existido a restrição do mapeamento a termos clínicos, foi demonstrado que a maior parte dos termos possui algum tipo de representação em ontologias realistas, seja através de uma definição completa (condições necessárias e suficientes) ou parcial (condições necessárias). As definições pretendem ser úteis durante o alinhamento de diferentes modelos de informação ou modelos clínicos (arquétipos). A avaliação da utilidade por meios quantitativos está fora do escopo da pesquisa, mas deve ser avaliada futuramente.

Finalmente, a quinta contribuição (seção 7) foi a representação de trechos de prontuários médicos como um conjunto de asserções A-Box (instâncias) utilizando ontologias realistas. Ainda que se restrinja ao estudo de uma série de casos (apenas 188 fragmentos sentenciais e um prontuário resumindo uma história clínica), os achados sugerem a viabilidade da abordagem para a recuperação de informação a partir de questões clínicas reais. Revelam também as limitações das ontologias para representação das afirmativas contidas nos prontuários médicos.

Durante a realização do doutorado, vários artigos completos foram publicados em periódicos ou em anais de conferências nacionais e internacionais - (ALMEIDA; ANDRADE, 2011), (ANDRADE; ALMEIDA, 2011b), (ANDRADE; ALMEIDA, 2011a), (ANDRADE *et al.*, 2012a), (ALMEIDA *et al.*, 2012), (ANDRADE *et al.*, 2012b), (ANDRADE *et al.*, 2012c), (MARTÍNEZ-COSTA *et al.*, 2012). Estes artigos foram essenciais, devido ao retorno que geraram, e contribuíram significativamente para a qualidade da pesquisa.

Como próximos passos, a pesquisa deve caminhar para a aplicação das lições aprendidas em situações reais, no sentido de obter maior validação das vantagens e desvantagens e sua relação com diferentes casos de uso. Este relacionamento entre o uso e a técnica certamente promoverá a maior eficácia e eficiência de projetos que requerem interoperabilidade de dados de saúde, levando ao final é a melhoria do cuidado ao paciente.

9 BIBLIOGRAFIA

ALLEN, J. F.; HAYES, P. J. Moments and points in an interval-based temporal logic. **Comput. Intell.**, v. 5, n. 4, p. 225-238, 1990. ISSN 0824-7935.

ALMEIDA, M. B. **Um modelo baseado em ontologias para representação da memória organizacional**. 2006. 321 (PhD). Escola de Ciência da Informação, UFMG, Belo Horizonte.

ALMEIDA, M. B. A proposal to evaluate ontology content. **Applied Ontology**, v. 4, n. 3-4, p. 245-265, 2009. ISSN 1570-5838.

ALMEIDA, M. B.; ANDRADE, A. Q. **Information, Reality and Epistemology: An Ontological Take**. 3rd Workshop Ontologies in Biomedicine and Life Sciences (OBML), 2011, Berlim, Germany. Universitaet Leipzig.

ALMEIDA, M. B.; ANDRADE, A. Q.; MENDONÇA, F. M. **Epistemology and medical records: an applied evaluation**. ONTOBRAS-MOST 2012. Recife, Brazil 2012.

ALMEIDA, M. B. et al. **The Blood Ontology: an ontology in the domain of hematology**. ICBO 2011, 2011a, Buffalo.

ALMEIDA, M. B.; SOUZA, R. R.; COELHO, K. C. **Estudo exploratório sobre ontologias aplicadas a modelos de sistemas de informação: perspectivas de pesquisa em Ciência da Informação**. IV Encontro Ibérico EDIBCIC 2009. Coimbra, Portugal 2009.

ALMEIDA, M. B.; SOUZA, R. R.; FONSECA, F. Semantics in the Semantic Web: a critical evaluation. **Knowledge Organization Journal**, 2011b.

ANDRADE, A. Q.; ALMEIDA, M. B. **A realism-based analysis of the OpenEHR entry model**. Proceedings of the MIXHS'11, 2011a, Glasgow, Scotland, UK. ACM. p.51-58.

ANDRADE, A. Q.; ALMEIDA, M. B. **Realist representation of the medical practice: an ontological and epistemological analysis**. Proceedings of the 4th Ontobras, 2011b, Gramado, Brazil.

ANDRADE, A. Q.; ALMEIDA, M. B.; SCHULZ, S. **Revisiting ontological foundations of the OpenEHR Entry Model**. Proceedings of the 3rd International Conference on Biomedical Ontology (ICBO 2012), 2012a, Graz, Austria.

ANDRADE, A. Q. et al. Process attributes in bio-ontologies. **Bmc Bioinformatics**, v. 13, n. 1, p. 217, Aug 28 2012b. ISSN 1471-2105 (Electronic) 1471-2105 (Linking).

ANDRADE, A. Q. et al. Requirements for semantic biobanks. **Stud Health Technol Inform**, v. 180, p. 569-73, 2012c. ISSN 0926-9630 (Print)

0926-9630 (Linking).

ASHBURNER, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. **Nature Genetics**, v. 25, n. 1, p. 25-9, May 2000. ISSN 1061-4036 (Print)
1061-4036 (Linking).

BAADER, F. et al., Eds. **The Description Logic Handbook: Theory, Implementation and Applications**. New York: Cambridge University Press, p.555ed. 2003.

BARD, J.; RHEE, S. Y.; ASHBURNER, M. An ontology for cell types. **Genome Biol**, v. 6, n. 2, p. R21, 2005. ISSN 1465-6914 (Electronic)
1465-6906 (Linking).

BATCHELOR, C.; HASTINGS, J.; STEINBECK, C. **Processes and properties**. In: SHAH, N.; SANSONE, S.-A., *et al*, Proceedings of the BioOntologies SIG at ISMB/ECCB 2011, 2011, Vienna, Austria. 15-16th July.

BAWDEN, D. The three worlds of health information. **Journal of Information Science**, v. 28, n. 1, p. 51-62, 2002. ISSN 0165-5515. Disponível em: < <Go to ISI>://000174166300006 >.

BEALE, T.; HEARD, S. **Architecture Overview**. OpenEHR Foundation, p.87. 2007a

BEALE, T.; HEARD, S. An ontology-based model of clinical information. **Stud Health Technol Inform**, v. 129, n. Pt 1, p. 760-4, 2007b. ISSN 0926-9630 (Print)
0926-9630 (Linking).

BEIßWANGER, E. et al. BioTop: An Upper Domain Ontology for the Life Sciences - A Description of its Current Structure, Contents, and Interfaces to OBO Ontologies. **Applied Ontology**, v. 3, n. 4, p. 205-212, December 2008.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, v. 284, n. 5, p. 34-38, May 2001. ISSN 0036-8733. Disponível em: < <Go to ISI>://000168217200023 >.

BHOGAL, J.; MACFARLANE, A.; SMITH, P. A review of ontology based query expansion. **Information Processing & Management**, v. 43, n. 4, p. 866-886, 2007. ISSN 0306-4573. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0306457306001476> >.

BICKLEY, L. S.; SZILAGYI, P. G. **Bates' Guide to Physical Examination and History Taking**. 10th Edition. Lippincott Williams & Wilkins, 2009.

BLOBEL, B.; ENGEL, K.; PHAROW, P. Semantic interoperability - HL7 version 3 compared to advanced architecture standards. **Methods of Information in Medicine**, v. 45, n. 4, p. 343-353, 2006. ISSN 0026-1270. Disponível em: < <Go to ISI>://WOS:000239824800003 >.

BODENREIDER, O.; OLKEN, F. Ontology Summit 2007 Communiqué. 2007. Disponível em: < http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2007_Communique >. Acesso em: 4th January.

BODENREIDER, O.; SMITH, B.; BURGUN, A. **The Ontology-Epistemology Divide: A Case Study in Medical Terminology**. In: VARZI, A. e VIEU, L., 3rd Conference on Formal Ontology in Information Systems (FOIS 2004), 2004, Turin.

BODENREIDER, O. et al. Investigating subsumption in SNOMED CT: An exploration into large description logic-based biomedical terminologies. **Artificial Intelligence in Medicine**, v. 39, n. 3, p. 183-195, 2007. ISSN 0933-3657. Disponível em: < <Go to ISI>://WOS:000246657200002 >.

BODENREIDER, O.; STEVENS, R. Bio-ontologies: current trends and future directions. **Briefings in Bioinformatics**, v. 7, n. 3, p. 256-274, 2006. ISSN 1467-5463. Disponível em: < <Go to ISI>://WOS:000240964500005 >.

BREWSTER, C.; O'HARA, K. Knowledge representation with ontologies: Present challenge - Future possibilities. **International Journal of Human-Computer Studies**, v. 65, n. 7, p. 563-568, Jul 2007. ISSN 1071-5819. Disponível em: < <Go to ISI>://000247427700001 >.

CAMPBELL, J. R. et al. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. CPRI Work Group on Codes and Structures. **J Am Med Inform Assoc**, v. 4, n. 3, p. 238-51, May-Jun 1997. ISSN 1067-5027 (Print) 1067-5027 (Linking).

CEN/ISO EN13606 INFORMATION SITE. The CEN/ISO EN13606 Standard. 2009. Disponível em: < <http://www.en13606.org/> >. Acesso em: 6th March.

CEUSTERS, W. et al. An Evolutionary Approach to Realism-based Adverse Event Representations. **Methods of Information in Medicine**, v. 50, n. 1, p. 62-73, 2011. ISSN 0026-1270. Disponível em: < <Go to ISI>://WOS:000286799500008 >.

CEUSTERS, W.; ELKIN, P.; SMITH, B. Negative findings in electronic health records and biomedical ontologies: a realist approach. **Int J Med Inform**, v. 76 Suppl 3, p. S326-33, Dec 2007. ISSN 1872-8243 (Electronic) 1386-5056 (Linking).

CEUSTERS, W.; SMITH, B. Referent tracking for treatment optimisation in schizophrenic patients: A case study in applying philosophical ontology to diagnostic algorithms. **Journal of Web Semantics**, v. 4, n. 3, p. 229-236, 2006. ISSN 1570-8268. Disponível em: < <Go to ISI>://WOS:000247054000009 >.

CEUSTERS, W.; SMITH, B.; GOLDBERG, L. A terminological and ontological analysis of the NCI Thesaurus. **Methods Inf Med**, v. 44, n. 4, p. 498-507, 2005. ISSN 0026-1270 (Print) 0026-1270 (Linking).

CHANDRASEKARAN, B.; JOSEPHSON, J. R.; BENJAMINS, V. R. What are ontologies, and why do we need them? **Ieee Intelligent Systems & Their Applications**, v. 14, n. 1, p. 20-26, Jan-Feb 1999. ISSN 1094-7167. Disponível em: < <Go to ISI>://000078503400006 >.

CIMINO, J. J. Desiderata for controlled medical vocabularies in the twenty-first century. **Methods Inf Med**, v. 37, n. 4-5, p. 394-403, Nov 1998. ISSN 0026-1270 (Print) 0026-1270 (Linking).

CIMINO, J. J. In defense of the Desiderata. **J Biomed Inform**, v. 39, n. 3, p. 299-306, Jun 2006. ISSN 1532-0480 (Electronic) 1532-0464 (Linking).

CIMINO, J. J.; ZHU, X. The practical impact of ontologies on biomedical informatics. **Methods of Information in Medicine**, v. 45, p. 124-135, 2006. ISSN 0026-1270. Disponível em: < <Go to ISI>://WOS:000242775200022 > .

CONNORS, J. M.; BRITTON, K. A. A Bloody Mystery. October 15, 2009 2009. Disponível em: < <http://www.nejm.org/doi/full/10.1056/NEJMimc0902429> >. Acesso em: December 28th.

CONSELHO FEDERAL DE MEDICINA. **Resolução 1638/2002**. CONSELHO FEDERAL DE MEDICINA 2002.

COSTA, C. M.; MENARGUEZ-TORTOSA, M.; FERNANDEZ-BREIS, J. T. Clinical data interoperability based on archetype transformation. **J Biomed Inform**, v. 44, n. 5, p. 869-80, Oct 2011. ISSN 1532-0480 (Electronic) 1532-0464 (Linking).

COURTOT, M. et al. MIREOT: The minimum information to reference an external ontology term. **Applied Ontology**, v. 6, n. 1, p. 23-33, 2011.

CREATH, R.; FRIEDMAN, M. **The Cambridge companion to Carnap**. Cambridge University Press, 2007. 371.

CROSS, J. H. et al. Clinical examination compared with anthropometry in evaluating nutritional status. **Arch Dis Child**, v. 72, n. 1, p. 60-1, Jan 1995. ISSN 1468-2044 (Electronic) 0003-9888 (Linking).

DE ROOS, A. **A critical evaluation of HL7 version 3 as an interoperable EHR standard** 2007.

DOLIN, R. H. et al. HL7 Clinical Document Architecture, Release 2. **Journal of the American Medical Informatics Association**, v. 13, n. 1, p. 30-39, 2006. ISSN 1067-5027. Disponível em: < <Go to ISI>://WOS:000234724000006 > .

DRUMMOND, N. et al. **Putting OWL in Order: Patterns for sequences in OWL**. In: GRAU, B. C.; HITZLER, P., *et al*, OWL Experiences and Directions (OWLEd 2006), 2006, Athens Georgia (USA).

EUZENAT J. **Towards a principled approach to semantic interoperability**. IJCAI-01 Workshop on Ontologies and Information Sharing, 2001, Seattle, USA. August 4-5. p.19-25.

FERANTI, J. M. et al. The clinical document architecture and the continuity of care record: A critical analysis. **Journal of the American Medical Informatics Association**, v. 13, n. 3, p. 245-252, 2006. ISSN 1067-5027. Disponível em: < <Go to ISI>://WOS:000238264200002 > .

FONSECA, F. The double role of ontologies in information science research. **Journal of the American Society for Information Science and Technology**, v. 58, n. 6, p. 786-793, 2007. ISSN 1532-2882. Disponível em: < <Go to ISI>://WOS:000246379800003 > .

FONSECA, F.; MARTIN, J. **Play as the Way Out of the Newspeak-Tower of Babel Dilemma in Data Modeling**. ICIS 2005, 2005a.

FONSECA, F.; MARTIN, J. Beyond Newspeak: three arguments for the persistence of the informal in the creation and use of computational ontologies. **Knowledge Management Research & Practice**, v. 7, n. 3, p. 196-205, 2009. ISSN 1477-8238. Disponível em: < <Go to ISI>://WOS:000282845600003 >.

FONSECA, F. T.; MARTIN, J. E. Toward an alternative notion of information systems ontologies: Information engineering as a hermeneutic enterprise. **Journal of the American Society for Information Science and Technology**, v. 56, n. 1, p. 46-57, Jan 2005b. ISSN 1532-2882. Disponível em: < <Go to ISI>://000225799300006 >.

FREITAS, F.; SCHULZ, S.; MORAES, E. Survey of current terminologies and ontologies in biology and medicine. **Electronic Journal of Communication Information and Innovation in Health**, v. 3, n. 1, p. 7-18, Mar, 2009 2009.

GALTON, A.; MIZOGUCHI, R. The water falls but the waterfall does not fall: New perspectives on objects, processes and events. **Applied Ontology**, v. 4, n. 2, p. 71-107, 2009 2009. ISSN 1570-5838. Disponível em: < <Go to ISI>://WOS:000207936600001 >.

GANGEMI, A. et al. Sweetening ontologies with DOLCE. **Knowledge Engineering and Knowledge Management, Proceedings**, v. 2473, p. 166-181, 2002. ISSN 0302-9743. Disponível em: < <Go to ISI>://WOS:000189412200018 >.

GARDE, S. et al. Expressing clinical data sets with openEHR archetypes: A solid basis for ubiquitous computing. **International Journal of Medical Informatics**, v. 76, p. S334-S341, 2007. ISSN 1386-5056. Disponível em: < <Go to ISI>://WOS:000207730900003 >.

GIL, A. C. **Como Elaborar Projetos de Pesquisa**. São Paulo: Atlas, 1991.

GOLDFAIN, A. et al. **Vital Sign Ontology**. In: SHAH, N.; SANSONE, S.-A., *et al*, Proceedings of the BioOntologies SIG at ISMB/ECCB 2011, 2011, Vienna, Austria. 15-16th July.

GÓMEZ-PÉREZ, A.; FERNÁNDEZ-LÓPEZ, M.; CORCHO, O. **Ontological engineering : with examples from the areas of knowledge management, e-commerce and the semantic web**. London: Springer-Verlag, 2004.

GRENON, P.; SMITH, B.; GOLDBERG, L. Biodynamic ontology: applying BFO in the biomedical domain. In: PISANELLI, D. M. (Ed.). **Ontologies in Medicine**. Amsterdam: IOS Press, 2004. p.20-38.

GREWE, N. **A Generic Reification Strategy for n-ary Relations in DL**. OBML 2010 - 2nd Workshop of Ontologies in Biomedicine and Life Sciences. Mannheim, Germany 2010.

GREWE, N. **Relating Processes and Events for Granularity-neutral Modeling**. In: SHAH, N.; SANSONE, S.-A., *et al*, Proceedings of the BioOntologies SIG at ISMB/ECCB 2011, 2011, Vienna, Austria. 15-16th July 2011.

GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing. **International Journal of Human-Computer Studies**, v. 43, n. 5-6, p. 907-928, Nov-Dec 1995. ISSN 1071-5819. Disponível em: < <Go to ISI>://A1995TY52900017 >.

GRUNINGER, M. et al. Ontology Summit 2007 - Ontology, taxonomy, folksonomy: Understanding the distinctions. **Applied Ontology**, v. 3, n. 3, p. 191-200, 2008.

GRUNINGER, M.; FOX, M. S. **Methodology for the Design and Evaluation of Ontologies**. Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95. Montreal 1995.

GUARINO, N. Formal ontology, conceptual analysis and knowledge representation. **International Journal of Human-Computer Studies**, v. 43, n. 5-6, p. 625-640, Nov-Dec 1995. ISSN 1071-5819. Disponível em: < <Go to ISI>://A1995TY52900002 >.

GUARINO, N. **Formal Ontology and Information Systems**. In: GUARINO, N., FOIS'98, 1998, Trento, Italy. IOS Press, november 20, 2007. p.3-15.

GUARINO, N.; POLI, R. The role of formal ontology in the information technology. **International Journal of Human-Computer Studies**, v. 43, n. 5-6, p. 623-624, 1995. ISSN 1071-5819. Disponível em: < <Go to ISI>://WOS:A1995TY52900001 >.

GUARINO, N.; WELTY, C. Evaluating ontological decisions with ontoclean. **Communications of the Acm**, v. 45, n. 2, p. 61-65, 2002. ISSN 0001-0782. Disponível em: < <Go to ISI>://WOS:000173489300022 >.

HARE, M. et al. Activating event knowledge. **Cognition**, v. 111, n. 2, p. 151-167, 2009. ISSN 0010-0277. Disponível em: < <Go to ISI>://WOS:000265571100001 >.

HILL, D. P. et al. Gene Ontology annotations: what they mean and where they come from. **Bmc Bioinformatics**, v. 9, 2008. ISSN 1471-2105. Disponível em: < <Go to ISI>://WOS:000259023600002 >.

HIRSH, J. et al. American Heart Association/American College of Cardiology Foundation guide to warfarin therapy. **Circulation**, v. 107, n. 12, p. 1692-711, Apr 1 2003. ISSN 1524-4539 (Electronic) 0009-7322 (Linking).

HITZLER, P. et al., Eds. **OWL 2 Web Ontology Language Primer** W3C Recommendation. 2009.

HJORLAND, B. Semantics and knowledge organization. **Annual Review of Information Science and Technology**, v. 41, p. 367-405, 2007. ISSN 0066-4200. Disponível em: < <Go to ISI>://000242813600009 >.

HL7. HL7 V3 Reference Information Model: RIM Version V 02-32 September 2010. Membership Normative Ballot. 08/22/2010 8:35 PM 2010. Disponível em: < <http://www.hl7.org/v3ballot2010SEP/html/welcome/environment/index.htm> >. Acesso em: 20 outubro.

HOEHNDORF, R.; DUMONTIER, M.; GKOUTOS, G. V. Evaluation of research in biomedical ontologies. **Briefings in Bioinformatics**, September 8, 2012 2012. Disponível em: < <http://bib.oxfordjournals.org/content/early/2012/09/07/bib.bbs053.abstract> >.

HORRIDGE, M. et al. **The Manchester OWL syntax**. *Proc. of the 2006 OWL Experiences and Directions Workshop (OWL-ED2006)*. Athens, Georgia: CEUR 2006.

HORROCKS, I. et al. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. 2004. Disponível em: < <http://www.w3.org/Submission/SWRL/> >. Acesso em: June 6th.

HOVENGA, E.; GARDE, S.; HEARD, S. Nursing constraint models for electronic health records: a vision for domain knowledge governance. **Int J Med Inform**, v. 74, n. 11-12, p. 886-98, Dec 2005. ISSN 1386-5056 (Print) 1386-5056 (Linking).

IHTSDO. 2012. Disponível em: < <http://www.ihtsdo.org/snomed-ct/> >. Acesso em: November 20th.

JANSEN, L.; SCHULZ, S. Grains, components and mixtures in biomedical ontologies. **J Biomed Semantics**, v. 9, n. 2, p. 9, 2011a. ISSN 2041-1480 (Electronic).

JANSEN, L.; SCHULZ, S. **The Ten Commandments of Ontological Engineering**. In: HERRE, H.; HOEHNDORF, R., *et al*, Proceedings of the 3rd Workshop Ontologies in Biomedicine and Life Sciences, 2011b, Berlim, Germany. Universitaet Leipzig.

JASPER, R.; USCHOLD, M. A framework for understanding and classifying ontology applications. 1999. Disponível em: < citeulike-article-id:7749652 <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.83.7541> >. Acesso em: 18th February.

JETTE, N. et al. The Development, Evolution, and Modifications of ICD-10 Challenges to the International Comparability of Morbidity Data. **Medical Care**, v. 48, n. 12, p. 1105-1110, 2010. ISSN 0025-7079. Disponível em: < <Go to ISI>://WOS:000284380200009 >.

KALRA, D. Electronic health record standards. **Methods of Information in Medicine**, v. 45, p. 136-144, 2006. ISSN 0026-1270. Disponível em: < <Go to ISI>://000242775200023 >.

KAMINKER, D.; CAMPOS, F. HL7 V3 Intro: Guide to the Standard. In: SHAFAMAN, M.; KAMINKER, D., *et al* (Ed.). **HL7 E-Learning Course**, 2008.

KLEIN, G. O.; SMITH, B. Concept Systems and Ontologies: Recommendations for Basic Terminology. **Information and Media Technologies**, v. 5, n. 2, p. 720-728, 2010.

KNUBLAUCH, H.; HENDLER, J. A.; IDEHEN, K. SPIN - Overview and Motivation. 2011. Disponível em: < <http://www.w3.org/Submission/spin-overview/> >. Acesso em: June 6th.

KÖHLER, J.; PHILIPPI, S.; LANGE, M. SEMEDA: ontology based semantic integration of biological databases. **Bioinformatics**, v. 19, n. 18, p. 2420-2427, December 12, 2003 2003. Disponível em: < <http://bioinformatics.oxfordjournals.org/content/19/18/2420.abstract> >.

KUTZ, O. et al. **Hyperontology for the Biomedical Ontologist: A sketch and some examples**. Proceedings of the Workshop on Working with Multiple Biomedical Ontologies (co-located with ICBO 2011), 2011, Buffalo, USA.

LENAT, D. B. Cyc - a Large-Scale Investment in Knowledge Infrastructure. **Communications of the Acm**, v. 38, n. 11, p. 33-38, 1995. ISSN 0001-0782. Disponível em: < <Go to ISI>://WOS:A1995TC17500013 >.

LORD, P.; STEVENS, R. Adding a little reality to building ontologies for biology. **PLoS One**, v. 5, n. 9, p. e12258, 2010. ISSN 1932-6203 (Electronic) 1932-6203 (Linking).

MARTÍNEZ-COSTA, C. et al. **Towards the harmonization of Clinical Information and Terminologies by Formal Representation (accepted for presentation)**. 13th International HL7 Interoperability Conference (IHIC), 2012, Viena, Austria. 26-27th September.

MARTINEZ-COSTA, C.; MENARGUEZ-TORTOSA, M.; FERNANDEZ-BREIS, J. T. An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes. **J Biomed Inform**, v. 43, n. 5, p. 736-46, Oct 2010. ISSN 1532-0480 (Electronic) 1532-0464 (Linking).

MCCARTHY, J. **Programs with Common Sense**. Teddington Conference on the Mechanization of Thought Processes, 1959, London. Her Majesty's Stationery Office.

MCCARTHY, J. Circumscription - a Form of Non-Monotonic Reasoning. **Artificial Intelligence**, v. 13, n. 1-2, p. 27-39, 1980. ISSN 0004-3702. Disponível em: < <Go to ISI>://WOS:A1980JW24000003 >.

MCCARTHY, J.; HAYES, P. Some Philosophical Problems from the Standpoint of Artificial Intelligence. **Machine Intelligence**, v. 4, p. 463-502, 1969. Disponível em: < <http://www-formal.stanford.edu/jmc/mcchay69.pdf> >.

MERRILL, G. H. Concepts and synonymy in the UMLS Metathesaurus. **J Biomed Discov Collab**, v. 4, p. 7, 2009. ISSN 1747-5333 (Electronic) 1747-5333 (Linking).

MERRILL, G. H. Ontological realism: Methodology or misdirection? **Applied Ontology**, v. 5, p. 79-108, 2010a. ISSN DOI 10.3233/AO-2010-0076.

MERRILL, G. H. Realism and reference ontologies: Considerations, reflections and problems. **Applied Ontology**, v. 5, p. 189-221, 2010b.

NEHTA. **Review of shared electronic health record standards**. National E-Health Transition Authority Ltd. 2006

NIINILUOTO, I. **Critical scientific realism**. New York: Oxford University Press, 1999. 355.

NIRENBURG, S.; WILKS, Y. What's in a symbol: ontology, representation and language. **Journal of Experimental & Theoretical Artificial Intelligence**, v. 13, n. 1, p. 9-23, 2001. ISSN 0952-813X. Disponível em: < <Go to ISI>://WOS:000168648600002 >.

NOY, N. F. et al. BioPortal: ontologies and integrated data resources at the click of a mouse. **Nucleic Acids Research**, v. 37, p. W170-W173, 2009. ISSN 0305-1048. Disponível em: < <Go to ISI>://WOS:000267889100031 >.

OBO PHENOTYPE GROUP. PATO - Phenotypic Quality Ontology. Disponível em: < http://www.bioontology.org/wiki/index.php/PATO:Main_Page >. Acesso em: 2nd January.

OPENEHR FOUNDATION. OpenEHR. 2007. Disponível em: < <http://www.openehr.org> >. Acesso em: September 18th.

OPENEHR FOUNDATION. Clinical Knowledge Manager. 2011. Disponível em: < <http://www.openehr.org/knowledge/> >. Acesso em: June 19th.

PADILHA, A. R. S. **PORTARIA Nº 2.073, DE 31 DE AGOSTO DE 2011**. MINISTÉRIO DA SAÚDE. Brasília 2011.

PARK, J.; RAM, S. Information systems interoperability: What lies beneath? **Acm Transactions on Information Systems**, v. 22, n. 4, p. 595-632, 2004. ISSN 1046-8188. Disponível em: < <Go to ISI>://WOS:000224839300004 >.

POIRIER, P. Adiposity and cardiovascular disease: are we using the right definition of obesity? **Eur Heart J**, v. 28, n. 17, p. 2047-8, Sep 2007. ISSN 0195-668X (Print) 0195-668X (Linking).

POLI, R. Ontology: The Categorical Stance. In: POLI, R. e SEIBT, J. (Ed.). **Theory and Applications of Ontology: Philosophical Perspectives**. 1st. Berlin: Springer, 2010. p.1-22.

POLI, R. et al. Formal and Descriptive Ontologists. 11th February 2011 2002. Disponível em: < <http://www.formalontology.com/essays/table-ontologists.pdf> >. Acesso em: 16th March.

POPPER, K. **Three Worlds**. The Tanner Lectures on Human Values. 2011 1978.

RAPAPORT, W. J. Holism, conceptual-role semantics, and syntactic semantics. **Minds and Machines**, v. 12, n. 1, p. 3-59, Feb 2002. ISSN 0924-6495. Disponível em: < <Go to ISI>://000173210100001 >.

RAPAPORT, W. J. What did you mean by that? Misunderstanding, negotiation, and syntactic semantics. **Minds and Machines**, v. 13, n. 3, p. 397-427, 2003. ISSN 0924-6495. Disponível em: < <Go to ISI>://WOS:000183475000003 >.

RAPAPORT, W. J. Intensionality vs. Intentionality. 15 April 2009 2009. Disponível em: < <http://www.cse.buffalo.edu/~rapaport/intensional.html> >. Acesso em: 2nd March.

RECTOR, A. **Knowledge Driven Software and "Fractal Tailoring": Ontologies in development environments for clinical systems**. Proceeding of the 2010 conference on Formal Ontology in Information Systems: Proceedings of the Sixth International Conference (FOIS 2010). Toronto, Canada: IOS Press 2010.

RECTOR, A.; ROGERS, J. Ontological and practical issues in using a description logic to represent medical concept systems: Experience from GALEN. **Reasoning Web**, v. 4126, p. 197-231, 2006. ISSN 0302-9743. Disponível em: < <Go to ISI>://WOS:000240539300009 >.

RECTOR, A.; ROGERS, J.; BITTNER, T. Granularity, scale and collectivity: When size does and does not matter. **Journal of Biomedical Informatics**, v. 39, n. 3, p. 333-349, Jun 2006. ISSN 1532-0464. Disponível em: < <Go to ISI>://000237980400010 >.

RECTOR, A. L.; QAMAR, R.; MARLEY, T. Binding ontologies and coding systems to electronic health records and messages. **Applied Ontology**, v. 4, n. 1, p. 51-69, 2009. ISSN 1570-5838. Disponível em: <<Go to ISI>://WOS:000207936500005 >.

ROBINSON, P. N.; MUNDLOS, S. The human phenotype ontology. **Clinical Genetics**, v. 77, n. 6, p. 525-34, Jun 2010. ISSN 1399-0004 (Electronic) 0009-9163 (Linking).

SANTOS, M. R. **Sistema de registro eletrônico de saúde baseado na norma ISO 13606: aplicações na Secretaria de Estado de Saúde de Minas Gerais**. 2011. 175 (Doctoral degree). Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte.

SANTOS, M. R.; BAX, M. P.; KALRA, D. Building a logical EHR architecture based on ISO 13606 standard and semantic web technologies. **Stud Health Technol Inform**, v. 160, n. Pt 1, p. 161-5, 2010. ISSN 0926-9630 (Print) 0926-9630 (Linking).

SANTOS, M. R.; BAX, M. P.; KALRA, D. Dealing with the Archetypes Development Process for a Regional EHR System. **Applied Clinical Informatics**, v. 3, n. 3, p. 258-275, 2012.

SCHADOW, G.; MEAD, C. N.; WALKER, D. M. The HL7 reference information model under scrutiny. **Stud Health Technol Inform**, v. 124, p. 151-6, 2006. ISSN 0926-9630 (Print) 0926-9630 (Linking).

SCHAFFER, J. The Metaphysics of Causation. 2007. Disponível em: <<http://plato.stanford.edu/entries/causation-metaphysics/>>. Acesso em: 15th of January.

SCHEUERMANN, R. H.; CEUSTERS, W.; SMITH, B. Toward an ontological treatment of disease and diagnosis. **Summit on Translat Bioinforma**, v. 2009, p. 116-20, 2009. ISSN 2153-6430 (Electronic) 2153-6430 (Linking).

SCHLOEFFEL, P. et al. **The relationship between CEN 13606, HL7, and openEHR**. Ocean Informatics Pty Ltd. Sydney, Australia. 2006

SCHULZ, S. et al. Alignment of the UMLS semantic network with BioTop: methodology and assessment. **Bioinformatics**, v. 25, n. 12, p. i69-76, Jun 15 2009a. ISSN 1367-4811 (Electronic) 1367-4803 (Linking).

SCHULZ, S.; CORNET, R. **SNOMED CT's Ontological Commitment**. International Conference on Biomedical Ontology. 2009

SCHULZ, S.; HAHN, U. Part-whole representation and reasoning in formal biomedical ontologies. **Artificial Intelligence in Medicine**, v. 34, n. 3, p. 179-200, Jul 2005. ISSN 0933-3657. Disponível em: <<Go to ISI>://000230983700001 >.

SCHULZ, S.; HAHN, U. Towards the ontological foundations of symbolic biological theories. **Artif Intell Med**, v. 39, n. 3, p. 237-50, Mar 2007. ISSN 0933-3657 (Print) 0933-3657 (Linking).

SCHULZ, S.; KARLSSON, D. **Records and situations. Integrating contextual aspects in clinical ontologies.** In: SHAH, N.; SANSONE, S., *et al*, Proceedings of the BioOntologies SIG at ISMB/ECCB 2011, 2011, Vienna, Austria. p.49-52.

SCHULZ, S.; KLEIN, G. O. SNOMED CT – advances in concept mapping, retrieval, and ontological foundations. Selected contributions to the Semantic Mining Conference on SNOMED CT (SMCS 2006). **BMC Medical Informatics and Decision Making**, v. 8, n. S1, 2008.

SCHULZ, S.; KUMAR, A.; BITTNER, T. Biomedical ontologies: What part-of is and isn't. **Journal of Biomedical Informatics**, v. 39, n. 3, p. 350-361, Jun 2006. ISSN 1532-0464. Disponível em: < <Go to ISI>://000237980400011 >.

SCHULZ, S.; MARKO, K.; HAHN, U. Spatial location and its relevance for terminological inferences in bio-ontologies. **Bmc Bioinformatics**, v. 8, p. 12, Apr 2007. ISSN 1471-2105. Disponível em: < <Go to ISI>://000246973600001 >.

SCHULZ, S.; MARKÓ, K.; SUNTISRIVARAPORN, B. Formal representation of complex SNOMED CT expressions. **BMC Medical Informatics and Decision Making**, v. 8(Suppl 1), n. S9, 2008.

SCHULZ, S. *et al*. Bridging the semantics gap between terminologies, ontologies, and information models. **Stud Health Technol Inform**, v. 160, n. Pt 2, p. 1000-4, 2010a. ISSN 0926-9630 (Print)
0926-9630 (Linking).

SCHULZ, S. *et al*. Scalable representations of diseases in biomedical ontologies. **J Biomed Semantics**, v. 2 Suppl 2, p. S6, 2011. ISSN 2041-1480 (Electronic).

SCHULZ, S. *et al*. **Clinical Ontologies Interfacing the Real World** 2010b.

SCHULZ, S. *et al*. Strengths and limitations of formal ontologies in the biomedical domain. **RECIIS Rev Electron Comun Inf Inov Saude**, v. 3, n. 1, p. 31-45, Mar 1 2009b. ISSN 1981-6278 (Electronic)
1981-6278 (Linking).

SCHWARZ, U.; SMITH, B. Ontological relations. In: MUNN, K. e SMITH, B. (Ed.). **Applied Ontology. An Introduction**. Frankfurt/Paris/Lancaster/New Brunswick: Ontos Verlag, 2008. p.220-234.

SEMANTIC HEALTH NET CONSORTIUM. Semantic Health Net. 2012. Disponível em: < <http://www.semantichealthnet.eu/> >. Acesso em: 6th January.

SHADBOLT, N.; HALL, W.; BERNERS-LEE, T. The Semantic Web revisited. **Ieee Intelligent Systems**, v. 21, n. 3, p. 96-101, 2006. ISSN 1541-1672. Disponível em: < <Go to ISI>://WOS:000238052700022 >.

SHEARER, R.; MOTIK, B.; HORROCKS, I. **HermiT: A Highly-Efficient OWL Reasoner**. In: RUTTENBERG, A.; SATTLER, U., *et al*, Proc. of the 5th Int. Workshop on OWL: Experiences and Directions (OWLED 2008 EU), 2008, Karlsruhe, Germany. October 26--27.

SHORTLIFFE, E. H.; BUCHANAN, B. G.; FEIGENBAUM, E. A. Knowledge Engineering for Medical Decision-Making - Review of Computer-Based Clinical Decision Aids. **Proceedings of the IEEE**, v. 67, n. 9, p. 1207-1224, 1979. ISSN 0018-9219. Disponível em: < <Go to ISI>://WOS:A1979HJ92900003 >.

SICILIA, J. J. et al. Knowledge representation issues in ontology-based clinical Knowledge Management systems. **International Journal of Technology Management**, v. 47, n. 1-3, p. 191-206, 2009. ISSN 0267-5730. Disponível em: < <Go to ISI>://WOS:000264910000013 >.

SIMON, J. et al. Formal ontology for natural language processing and the integration of biomedical databases. **International Journal of Medical Informatics**, v. 75, n. 3-4, p. 224-231, 2006. ISSN 1386-5056. Disponível em: < <Go to ISI>://WOS:000236360500007 >.

SIRIN, E. et al. Pellet: A practical OWL-DL reasoner. **Journal of Web Semantics**, v. 5, n. 2, p. 51-53, 2007. ISSN 1570-8268. Disponível em: < <Go to ISI>://WOS:000247054400001 >.

SMITH, B. Ontology. In: FLORIDI, L. (Ed.). **The Blackwell Guide to the Philosophy of Computing and Information** Malden, MA: Blackwell, 2003a. p.155-166.

SMITH, B. Ontology and information systems. 2003b. Disponível em: < http://ontology.buffalo.edu/ontology_long.pdf >. Acesso em: november 11.

SMITH, B. **Beyond concepts: Ontology as reality representation**. In: VARZI, A. e VIEU, L., FOIS 2004, 2004, Turin, . IOS Press, 4-6 November. p.73-84.

SMITH, B. From concepts to clinical reality: An essay on the benchmarking of biomedical terminologies. **Journal of Biomedical Informatics**, v. 39, n. 3, p. 288-298, 2006. ISSN 1532-0464. Disponível em: < <Go to ISI>://WOS:000237980400005 >.

SMITH, B. Classifying Processes: An Essay In Applied Ontology. **Ratio**, v. in press, 2012.

SMITH, B. et al. Basic Formal Ontology 2.0: Draft Specification and User's Guide. April 5th 2012 2012. Disponível em: < <http://ontology.buffalo.edu/bfo/Reference/> >. Acesso em: November 20th.

SMITH, B. et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. **Nature Biotechnology**, v. 25, n. 11, p. 1251-1255, 2007. ISSN 1087-0156. Disponível em: < <Go to ISI>://WOS:000251086500025 >.

SMITH, B.; BROCHHAUSEN, M. Putting biomedical ontologies to work. **Methods Inf Med**, v. 49, n. 2, p. 135-40, 2010. ISSN 0026-1270 (Print) 0026-1270 (Linking).

SMITH, B.; CEUSTERS, W. Towards industrial strength philosophy: how analytical ontology can help medical informatics. **Interdisciplinary Science Reviews**, v. 28, n. 2, p. 106-111, 2003. ISSN 0308-0188. Disponível em: < <Go to ISI>://WOS:000184444000006 >.

SMITH, B.; CEUSTERS, W. HL7 RIM: An Incoherent Standard. **Studies in Health Technology and Informatics**, v. 124 p. 133-138, 2006.

SMITH, B.; CEUSTERS, W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. **Applied Ontology**, v. 5, p. 139-188, 2010.

SMITH, B. et al. Relations in biomedical ontologies. **Genome Biology**, v. 6, n. 5, p. 15, 2005. ISSN 1465-6914. Disponível em: <<Go to ISI>://000228864300013 >.

SMITH, B. et al. **Towards a reference terminology for ontology research and development in the biomedical domain**. KR-MED 2006, 2006. p.57-66.

SMITH, B.; VARZI, A. C. Fiat and bona fide boundaries. **Philosophy and Phenomenological Research**, v. 60, n. 2, p. 401-420, 2000. ISSN 0031-8205. Disponível em: <<Go to ISI>://WOS:000085817700007 >.

SMITH, B.; WELTY, C. **Ontology: Towards a New Synthesis**. FOIS'01. Ogunquit, Maine, USA: ACM 2001.

SOWA, J. F. Ontology, Metadata, and Semiotics. 04/12/2001 2000. Disponível em: <<http://users.bestweb.net/~sowa/peirce/ontometa.htm> >. Acesso em: 28th february.

SPEAKS, J. Theories of meaning. Jun 1, 2010 2010. Disponível em: <<http://plato.stanford.edu/archives/win2010/entries/meaning/> >. Acesso em: 26th february.

STENZHORN, H. et al. Adapting Clinical Ontologies in Real-World Environments. **Journal of Universal Computer Science**, v. 14, n. 22, p. 3767-3780, 2008. ISSN 0948-695X. Disponível em: <<Go to ISI>://000264755000008 >.

STOKES, M. et al. Impact of bleeding-related complications and/or blood product transfusions on hospital costs in inpatient surgical patients. **Bmc Health Services Research**, v. 11, n. 1, p. 135, 2011. ISSN 1472-6963. Disponível em: <<http://www.biomedcentral.com/1472-6963/11/135> >.

STROETMANN, V. N. et al. **Semantic Interoperability for better health and safer healthcare - Deployment and Research Roadmap for Europe**. European Communities. 2009

SWEDISH ASSOCIATION OF LOCAL AUTHORITIES AND REGIONS, A. B. F. H. A national decision in Sweden on healthcare informatic standard. 2010. Disponível em: <http://www.cehis.se/images/uploads/dokumentarkiv/Strategy_archetypes_madrid_short_Presentation_20100624_version_10.pdf >. Acesso em: September 18th.

TASK FORCE OF THE EUROPEAN SOCIETY OF CARDIOLOGY THE NORTH AMERICAN SOCIETY OF PACING ELECTROPHYSIOLOGY. Heart Rate Variability : Standards of Measurement, Physiological Interpretation, and Clinical Use. **Circulation**, v. 93, n. 5, p. 1043-1065, March 1, 1996 1996. Disponível em: <<http://circ.ahajournals.org/content/93/5/1043.short> >.

TSARKOV, D.; HORROCKS, I. FaCT++ Description Logic reasoner: System description. **Automated Reasoning, Proceedings**, v. 4130, p. 292-297, 2006. ISSN 0302-9743. Disponível em: <<Go to ISI>://WOS:000240085600026 >.

VICKERY, B. C. Ontologies. **Journal of Information Science**, v. 23, n. 4, p. 277-286, 1997. ISSN 0165-5515. Disponível em: <<Go to ISI>://WOS:A1997XM46500002 >.

WAND, Y.; WEBER, R. An Ontological Model of an Information-System. **Ieee Transactions on Software Engineering**, v. 16, n. 11, p. 1282-1292, 1990. ISSN 0098-5589. Disponível em: <<Go to ISI>://WOS:A1990ED91000006 >.

WINOGRAD, T.; FLORES, F. **Understanding Computers and Cognition: A New Foundation for Design**. Addison-Wesley Professional, 1987. 207 ISBN 0201112973.

WITTGENSTEIN, L. **Philosophical Investigations**. 2nd. Malden, Massachusetts: Blackwell Publishers, 1997. 272.

WOODS, W. What's in a Link - Revisited (William Woods presentation at Cambridge Semantic Web community event March 9th 2010 -). 2010. Disponível em: <<http://vimeo.com/14459415>>. Acesso em: 4th March.

WOODS, W. A. Meaning and links. **Ai Magazine**, v. 28, n. 4, p. 71-92, 2007. ISSN 0738-4602. Disponível em: <<Go to ISI>://WOS:000251846800005 >.

YIN, R. K. **Case Study Research: Design and Methods**. 2. 1994. 181.

10 APÊNDICE

10.1 CASO DE USO – TROMBOSE VENOSA PROFUNDA

<CASO CLÍNICO

Paciente010

02/01

Paciente evoluindo há 4 dias da admissão com dor e edema de MIE, associado a edema em MMSS assimétrico, predominante à D

Peço D-Dimero

03/01

MIE: Edema +++/4+, leve calor, dor a palpação de panturrilha, Roams: negativo

HD: TVP em MIE ? (alta probabilidade pelo score de Wells (3))

Tromboflebite?

D-dimero:211 (VR:positivo se »230)

#CD:

- Mantenho anticoagulacao com clexane 60mg BID

- Solicito duplex scan venoso de MIE

05/01

US: Presença de trombo móvel em terço proximal da veia poplítea esquerda.>

	Pergunta original	DL-query	Resultado	Resposta correta?
	Did the patient have:			
1	an assessment of their medical history	'clinical history'	Patient10History02012011	y
2	a physical examination?	'physical examination finding'	Patient10PhysicalExamination	y
3	Was the patient's probability of DVT assessed using the two-level DVT Wells score?	'data item' AND WellsScore	Patient010WellsScore	y
4	Record score:	SELECT Y WHERE ?x rdf:type WellsScore. ?x part_of PatientRecord001. ?x 'is about' Patient001. ?x hasValue y?.		
5	Was the patient offered a proximal leg vein ultrasound scan?	'plan specification' and ('is about' only (plan and (is_realized_by SOME (UltrasoundImagingTest AND has_participant SOME (leg AND part_of VALUE patient010))))))	Patient010USRequest	y
6	Did the patient have a proximal leg vein ultrasound scan: Within 4 hours of being requested OR Within 24 hours of being requested OR More than 24 hours after being requested	US Request – Instance1: 'plan specification' and ('is about' only (plan and (is_realized_by some (Venous_duplex_Scan and (has_participant some (leg and (part_of value patient010))))))	Patient010USRequest	y
		Instance2: temporal_interval AND (hasBeginning SOME (temporal_instant AND inverse (hasBeginning) value Patient010USRequest)) AND (hasEnd SOME (temporal_instant AND inverse (hasBeginning) SOME UltrasoundImagingTest))	No answer	n
		SELECT x WHERE ?x 'is about' instance2 ?x hasValue ?y		

7	Was the patient offered a D-dimer test before the result of the proximal leg vein ultrasound scan was available?	plan specification' AND ('is about' SOME (plan AND (is_realized_by SOME D-dimerTest AND (has_participant VALUE patient010)))) AND (is_output_of SOME (OpenEHR:Investigation_request AND precedes SOME ('data item' AND is_output_of SOME (Venous_duplex_Scan AND has_participant SOME (leg and part_of value patient010))))))	No answer	n
8	Was the proximal leg vein ultrasound scan positive?	'data item' AND (outcomeOf SOME (UltraSoundExamination AND (hasPart SOME ProximalLegUltraSoundExamination) AND (hasParticipant value Patient001)))) AND ('is about' SOME (BloodClot AND locatedIn SOME (Leg AND partOf value Patient001)))	patient010USResult	y
		-----#----- ('data item' AND outcomeOf SOME (UltraSoundExamination AND (hasPart SOME ProximalLegUltraSoundExamination) AND (hasParticipant value Patient001))) AND (AbnormalValue)	patient010USResult	y
9	Was the patient offered a D-dimer test?	'plan specification' AND ('is about' SOME (plan AND (is_realized_by SOME D-DimerExam AND hasParticipant VALUE Patient001)))	patient010D-dimerRequest	y
10	While waiting for a proximal leg vein ultrasound scan, did the patient have an interim 24-hour dose of parenteral anticoagulant?	(medication_administration AND (has_part SOME NeedleInjection) AND (has_participant SOME (Medication AND has_part SOME (PortionOfChemicalSubstance AND bearer_of SOME AnticoagulantRole))) AND precedes SOME (UltrasoundImagingTest AND has_part SOME (leg AND part_of VALUE patient010)))	No answer	n
11	While waiting for a proximal leg vein ultrasound scan, did the patient have a	D-dimerTest AND (has_participant VALUE patient010) AND (precedes SOME (UltrasoundImagingTest AND has_participant SOME (leg	No answer	n

	D-dimer test?	AND part_of VALUE patient010)))		
⁶⁹ 13	Was the D-dimer test positive?	('data item' and is_output_of SOME (D-dimerTest AND has_participant VALUE patient010)) AND (AbnormalValue)	no answer	y
14	Was the patient offered a repeat proximal leg vein ultrasound scan 6-8 days later?	plan specification' and (not (temporal_interval and (inverse ('is about') some ('measurement datum' and ('has measurement unit label' value day) and (hasValue value 8))) and (inverse (precedes) some (UltrasoundImagingTest and (has_participant some (leg and (part_of value patient010)))))) and ('is about' some (plan and (is_realized_by only (UltrasoundImagingTest and (has_participant some (leg and (part_of value patient010)))))) and (is_output_of some OrderRequestProcess) and (inverse (precedes) some (temporal_interval and (inverse ('is about') some ('measurement datum' and ('has measurement unit label' value day) and (hasValue value 6))) and (inverse (precedes) some (UltrasoundImagingTest and (has_participant some (leg and (part_of value patient010))))))	No answer	y

⁶⁹ Question 12 was repeated and is omitted here.

15	Was the repeat proximal leg vein ultrasound scan positive?	PositiveResultDataItem AND (outcomeOf SOME (UltraSoundExamination AND (hasPart SOME ProximalLegUltraSoundExamination) AND (hasParticipant value Patient001) AND (inverse (precedes) SOME (UltraSoundExamination AND (hasPart SOME ProximalLegUltraSoundExamination) AND (hasParticipant value Patient001))))))	Not applicable	
16	Was DVT diagnosed and treatment started?	Diagnosis1: Diagnosis AND ('is about' SOME (DeepVeinThrombosis AND inheresIn VALUE Patient001))	Patient010Diagnosis1 Patient010Diagnosis3	y
		MedicationAdministration AND (hasParticipant SOME (Medication AND hasPart SOME Heparin)) AND (inverse (precedes) SOME ('diagnostic process' AND hasOutcome VALUE Diagnosis1))	Patient010MedicationAdministration	y
17	Were alternative diagnoses considered?	diagnosis AND 'is about' SOME (disease AND 'inheres in' VALUE patient010) AND (NOT ('is about' SOME DeepVenousThrombosis))	No answer	n
	Was the patient advised that it was not likely they had PE?	NOT REPRESENTABLE USING CURRENT METHODS (requires free-text analysis)		

10.2 PROCESS ATTRIBUTES IN BIOMEDICAL ONTOLOGIES – PREPRINT VERSION⁷⁰

While a static description of the structure of an organism can provide some information about the state of its health at a given point in time, the adequate description of its dynamic functioning conveys a wealth of additional information. Biological processes include the intracellular transformations that consume and metabolize nutrients in order to produce energy, the overarching developmental process of a growing organism, the cycles of sleep and waking that structure our daily lives, and the pathological processes of unrestrained cellular proliferation that characterize certain types of cancer. Many of these processes can be perturbed by underlying pathological conditions, and understanding their normal operation as well as being able to recognize and annotate various sorts of processual pathologies such as delays or irregularities is essential to the development of knowledge-based information systems to support modern data-driven biological research. The distinction between a static description of the structure of an organism and the dynamic description of organismal processes corresponds to the distinction between the Gene Ontology (ASHBURNER *et al.*, 2000) branches for *cellular components* and *biological processes*. Knowledge resources in medical contexts similarly distinguish anatomical descriptions, e.g. of the structure of a heart, from descriptions of processes such as the heart's beating. Such distinctions are required for proper and unambiguous definitions of pathological phenomena and clinical findings like regular heartbeats and arrhythmia, and for distinguishing between normal and raised heart rates automatically. Health-related processes and the attributes by which these phenomena are described and modified (e.g. heart beating, regular or arrhythmic), are common in health records and appear in bio-ontologies, such as the Gene Ontology and the Human Phenotype Ontology (ROBINSON; MUNDLOS, 2010), and medical terminologies such as SNOMED CT.

However, the adequate ontological representation of process attributes such as rates and regularities, which have been referred to as process “qualities”, has been a contentious topic in recent years. In particular, within the frameworks of the Basic Formal Ontology (BFO)

⁷⁰ Este apêndice reproduz a versão pré-impressão do artigo publicado em (ANDRADE *et al.*, 2012b), oriundo do estágio sanduíche na Universidade Médica de Graz.

(GRENON *et al.*, 2004) and the methodology of Ontological Realism (SMITH; CEUSTERS, 2010), process “qualities” have not been supported. Where ontologies that have been aligned with BFO have needed to include such attributes, they have heretofore resorted to using workarounds, including representing the attributes of processes as if they were attributes of the participants, or representing complicated process hierarchies directly without reference to such attributes, having the consequence that there is no mechanism for comparison of processes based on their attributes.

Here, we present a novel solution to the problem of representing process attributes. Our solution rests on two key tenets: firstly, that the sorts of process attributes that are of biomedical interest emerge from the ways that repeated parts of such processes combine to constitute the overall process; secondly, that full logical definition of process attributes that follow this pattern is in principle possible, so that they do not need to be treated as primitives within ontologies. Using the Web Ontology Language (OWL) version 2 (HITZLER *et al.*, 2009), we will show how full logical definitions can be created for representation of attributes relevant for modeling normal and abnormal heart rates and heart cycles. We will further discuss examples of process attributes for which the expressivity offered by OWL cannot fully formalize the attributes of processes, but for which the use of full first-order logic (FOL) does allow the relevant formalization. We will evaluate these representations against a set of domain questions and tasks, and against a set of ontology evaluation criteria.

10.2.1 Ontological Realism and their problem with process “qualities”

Ontological realism has been repeatedly advocated as a theory which is particularly useful for the development of scientific ontologies. It advocates adherence to a set of philosophically grounded tenets in order to improve the quality and interoperability of the resulting ontology artifacts (SMITH; CEUSTERS, 2010). One such tenet is strict alignment with upper-level ontologies. BFO (GRENON *et al.*, 2004), which most deeply implements the precepts of ontological realism, makes a fundamental distinction between types of entities based on the relationship to time: continuants are those entities that *continue* to exist through time, and exist in full at all times during which they exist, such as a human being, while occurrents are those entities that *unfold* or *happen* in time, and have temporal parts, such as the life of a human being. The biological and biomedical reality that is described by bio-

ontologies such as those being gathered by the OBO Foundry (SMITH *et al.*, 2007) is broadly divided between continuants such as cells, molecules, genes, and tissues, and the occurrents in which those continuants participate such as biological processes, chemical reactions, and so on. Continuants are further categorized in independent continuants, which do not depend on other entities for their existence, and dependent continuants, which require an independent continuant to inhere in and to be borne by. In BFO, qualities are special types of dependent continuants, for example, the *colour* of a fruit or the *weight* of a person; neither the colour nor the weight can exist without their respective bearers – the fruit and the person – existing.

Whereas other top level ontologies like DOLCE, GFO, or BioTop refrain from an upper level bipartition, in BFO, qualities, as being continuants, have no temporal parts, and thus cannot inhere in occurrents, which are, by definition, partitioned along time. If qualities could inhere in processes, then they would necessarily have temporal parts. For instance, assuming that *being intermittent* is a quality of a disease course process. It is not possible to make any statement on whether a process is intermittent by inspecting a snapshot of this process at a particular moment in time. In contrast, it is perfectly possible to ascribe a colour to an apple at a moment in time. Furthermore, within the BFO framework, there is not even the possibility that there could be other sorts of quality-like entities that inhere in occurrents, which would correspond to attributes of processes, since qualities are the sorts of things that can change in their bearers over time (as an apple changes colour as it ripens), while processes cannot change over time, since processes *are* changes (GALTON; MIZOGUCHI, 2009). Each process has at least one (continuant) process participant and exactly one duration (the extent of the time interval between inception and ending). Processes can generally be split into numerous sub-processes, each of which has a duration and some participant. Since many words used in natural language to qualify processes are actually time-related properties of the sub-events and their material participants, our working hypothesis is the following:

Given that processes are characterized by their duration, their parts (sub-processes), their participants, and the qualities of their participants, these parameters are sufficient for representing the meaning of the alleged process qualities.

However, practically all processes of interest within biomedical science, and therefore which are subject to descriptions and formalizations in biomedical ontologies, are highly

complex entities, composed by numerous sub-processes, often of different kinds. Scientists are accustomed to using natural language for the assertion of biological events, in which adverbs modify verbs following the same pattern in which adjectives modify nouns, and considering that verbs frequently denote processes, users expect that adverbs should denote process attributes. Indeed, the use of process modifiers is widespread in the scientific literature. To represent these terms, several biomedical ontologies contain terms to modify processes, closely related to their use in natural language. In SNOMED CT (IHTSDO, 2012), among the so-called qualifier values there are numerous that can be post-coordinated with disease concepts, such as *Deterioration of status*, *Improvement of status*, *Chronic persistent*, *Progressive*, *Precipitant* and many more. PATO (OBO PHENOTYPE GROUP), the ontology of phenotypic qualities, distinguishes many flavors of decreased and increased properties of processes, such as *occurrence*, *rate*, *frequency*, and *duration*. Other properties include *synchronicity*, *acceleration* (a property of change), *intensity*, and *regularity / normality* vs. *irregularity / abnormality* (regarding rhythm or sleep pattern), or having extra or missing sub-process parts. The Human Phenotype Ontology (ROBINSON; MUNDLOS, 2010) does not separate the properties from their bearers, but contains equally numerous process terms with modifiers such as *Growth retardation*, *Slowly progressive disorder*, *Asymmetric growth*, *Paroxysmal bursts of laughter*, or *Limited shoulder movement*. Irregular or abnormal patterns of e.g. growth, movements etc. are also frequent.

Other ontologies have a different approach. Instead of conceding that processes require modifiers, these ontologies ascribe the qualities to the participants of the described process. For instance, the Vital Sign Ontology (VSO) (GOLDFAIN *et al.*, 2011), an extension of the Ontology for General Medical Science (OGMS) (SCHEUERMANN *et al.*, 2009), also describes aspects commonly thought as process modifiers, such as rates drawn from PATO. In the VSO, a pulse rate is described as “The rate at which an artery pulses (i.e., participates in expansion-contraction cycles) as blood passes through it.”, and is represented as a quality of some artery. This is clearly a work-around of the core issue: while convenient and very intuitive for healthcare researchers and practitioners, it is difficult to bite the bullet and accept that a rate is the sort of thing “that exists in full at any time in which it exists at all, persists through time while maintaining its identity and has no temporal parts”, i.e. a continuant. It is also not clear whether a pulse rate of 72 beats/minute means that there is something about the artery itself, rather than something about what the artery is doing (processually), namely that it is beating and will beat 72 times in the next minute.

We will demonstrate in what follows how the alleged “qualities of processes” can be fully defined in terms of patterns of relationships between the sub processes and the overall containing process. This provides a solution to the dilemma of whether to include them in bio-ontologies, since entities that are fully defined in terms of other, “genuine” entities, preserve the ontological commitments of extant upper level ontologies.

10.2.2 Preliminary evaluation

10.2.2.1 Use case: the heart cycle

Definitions for quality-like entities of processes depend on the domain of discourse. As a use case, we will assess to what extent we can provide ontological descriptions for a basic aspect of the heart physiology: the heart cycle and associated rates. As was shown in (BATCHELOR *et al.*, 2011), representing heart rates is not trivial and requires further understanding of the rationale behind some medical statements. The main function of a heart is to pump blood to other organs. Since this function is mechanic, we can describe it by a series of muscle contraction and relaxation cycles. During the relaxation phase, called diastole, the blood fills the heart cavities, whereas in the contraction part, called systole, the blood is pumped from the heart ventricles in the peripheral circulatory system. A heart cycle is, therefore, composed of two distinct parts, each being a precondition for the next (that is, there can be no contraction in a fully contracted heart).

In clinical practice, cardiologic evaluation is comprised by several distinct measures, including the search for signs of cardiac dysfunction (dyspnoea or breathlessness) and the use of devices to evaluate blood flow and electrical characteristics. These measures may refer to individual heart cycles (such as in Echocardiographs), or series of cycles, named here heart beating process. The most commonly evaluated surrogates for the heart beating process are heart rate and pulse rhythm. “Heart rate” is commonly defined as the number of times a complete heart cycle event occurs within a given time, usually per minute. It describes the mean duration of each cycle on the specific measured period. It is called fast if the mean duration of the cycles is shorter than normal, and slow if the mean duration is longer than normal. “Pulse Rhythm” is commonly defined as the regularity between the time intervals in a set of three or more subsequent cycles, which is called rhythmic if the intervals between sequential cycles are similar; and arrhythmic if the intervals show great variation.

Alterations of heart behaviour are seen in many diseases, and there are several names for most commonly observed patterns, like higher than normal (tachycardia) and slower than normal heart rate (bradycardia).

Current ontological models, that lack any attributes for processes, would treat heart rate as a regular observation, or a quality of the organism the heart is a part of, in the same fashion as body temperature. Considering the heart rate as a quality of the (human or animal) organism seems reasonable: at least in theory, it would be possible to calculate the heart rate based on complete knowledge of the chemical balance of the heart, the breathing cycle and some instantaneous measurement of heart contraction speed, to avoid the time-dependence. However, the situation becomes more complicated in cases in which we have to know the exact relation between some entity, like an event or a substance, and a resulting *change* in a heart rate. For example, a drug might cause a change in the duration of a heart cycle, directly affecting the heart rate, and there should be some way to capture the knowledge of the effect that such a drug can have or to relate such a change back to substances which can cause it. This requires explicit representation of process attributes. We can also find situations in which process attribute changes must be recorded in medical records. For instance, the concept of heart rate variability, recently implicated in worse outcomes of cardiovascular diseases, is measured by the change of the heart rate in resting position and after heart rate decreasing-increasing maneuvers (Valsalva) (TASK FORCE OF THE EUROPEAN SOCIETY OF CARDIOLOGY THE NORTH AMERICAN SOCIETY OF PACING ELECTROPHYSIOLOGY, 1996). We are, therefore, interested in describing the process of change of the heart behaviour, finally identifying whether the patient's heart is capable of varying its rhythm accordingly.

10.2.2.2 Evaluation criteria

In order to produce objective results, we will use pre-defined criteria to evaluate the quality of the resulting representation artifact. Since we are dealing with a foundational issue, it is hard to create clear quantitative metrics for evaluation. However, by using a simplified model it is possible to state the main advantages and problems, and allow for future comparison with other methods. The competency questions⁷¹ can be found at **Table XX**.

⁷¹ Some definitions of medical terms used to formulate competency questions:

<p>1. Given that we know the number and duration of n sequential heart cycles, can we categorize the heart beating process instance under the following classes:</p> <ol style="list-style-type: none"> Normal heart rate for a 30 year-old Fast heart rate for a 30 year-old
<p>2. If the beating process is properly classified:</p> <ol style="list-style-type: none"> Drug X is contra-indicated in bradycardia (external rule); patient Y has bradycardia (ontological representation) -> Drug X is contra-indicated for patient Y Knowing that the patient took drugs X, Y and Z, and knowing that the heart frequency of the patient has increased from base line, query which of the drugs could have caused such increase. Querying a database for diseases that co-occur with some arrhythmia.
<p>3. Can we represent the following medical statements (from real medical records):</p> <ol style="list-style-type: none"> "Paroxysmal atrial fibrillation (diagnosis)"; "Regular cardiac rhythm"; "Sudden onset of palpitation"; "History of supraventricular tachyarrhythmia"; "No atrium-ventricular or intra-ventricular conduction abnormalities (ECG finding)"; "Chronic atrial fibrillation (diagnosis)";

The answers will be qualitatively analyzed with the following generic questions in mind, based on (DRUMMOND *et al.*, 2006) :

- Is it useful?
- Does it produce the correct inferences?
- How expressive is it relative to the alternatives, in particular to regular expressions?
- Are there computationally more efficient solutions?
- Which pattern should be chosen for a particular application?

Paroxysmal atrial fibrillation: Atrial fibrillation that occur in episodes, separated by periods of normal heart beating

Palpitation: Sudden increase in heart rate

Tachyarrhythmia: Cardiac rhythm disorder in which the heart rate is abnormally high

10.2.2.3 Implementation

For the implementation of the use case, an OWL DL ontology was created using Protégé 4 OWL Editor <version 4.1.0, Build = 239>. The ontology extends OGMS version of 2011-09-20, which is linked to BFO version 1.1 and the OBO Relation Ontology. Time-related information artifact classes were extracted from OBI through the MIREOT methodology (COURTOT *et al.*, 2011).

The reasoners used were FaCT++ (version 1.5.3) and Hermit (version 1.3.5). They were run on an AMD A6-3410MX processor / 8 GB RAM computer, performing automatic classification of classes and their members, and probing for inconsistencies.

10.2.2.4 Use case analysis

For the ontological analysis of the heart cycle, common medical expressions were collected during the execution of the Blood Project, a cooperation between the Ontology Research Group at the University at Buffalo and the Hemominas Foundation and the School of Information Science at the Federal University of Minas Gerais. Expressions were translated from Portuguese to English by a domain expert. Competence questions were developed loosely based on questions present in medical records (e.g. a diagnostic question concerning a drug as cause of bradycardia), and based on common functions present on Electronic Health Records (drug contra-indications). Use case requirements were broadly discussed between the authors, aiming to describe different aspects of medical reporting (patient state, disease progress and physician reasoning process).

10.2.3 Results

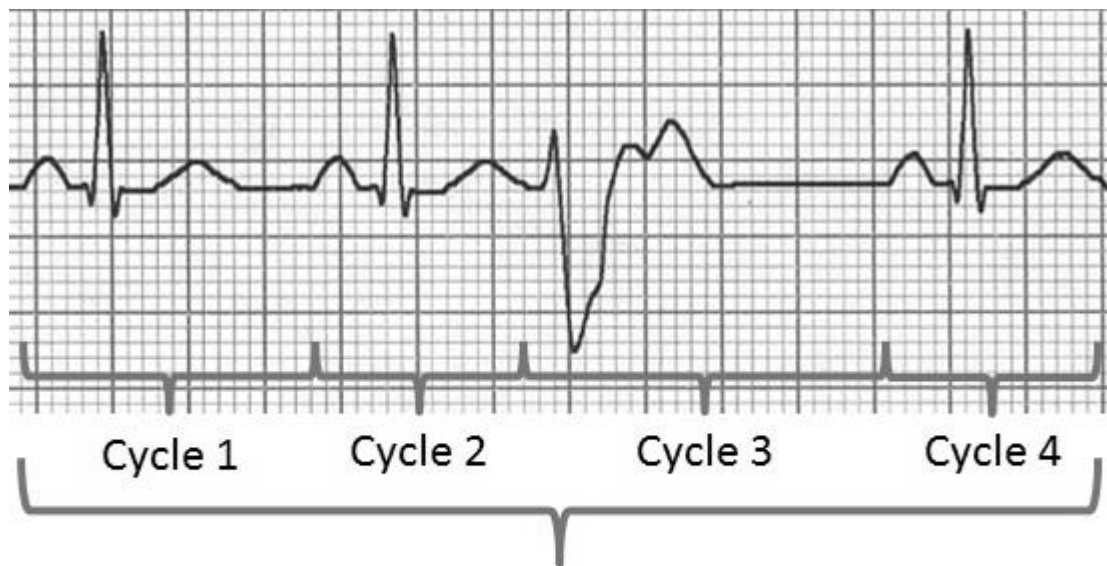
Based on the statements contained in medical records, medical literature analysis and ontological analysis of the process, several key terms were identified. They were represented as OWL classes, building upon the OGMS (SCHEUERMANN *et al.*, 2009).

10.2.3.1 Design pattern for representation of process attributes of cyclical processes and implementation in OWL

In the most general sense: our pattern for representation requires the following:

- That the overall process to which the attribute is being ascribed be composed of repeated sub-processes; and
- That the repeated sub-processes be enumerated and have a duration.

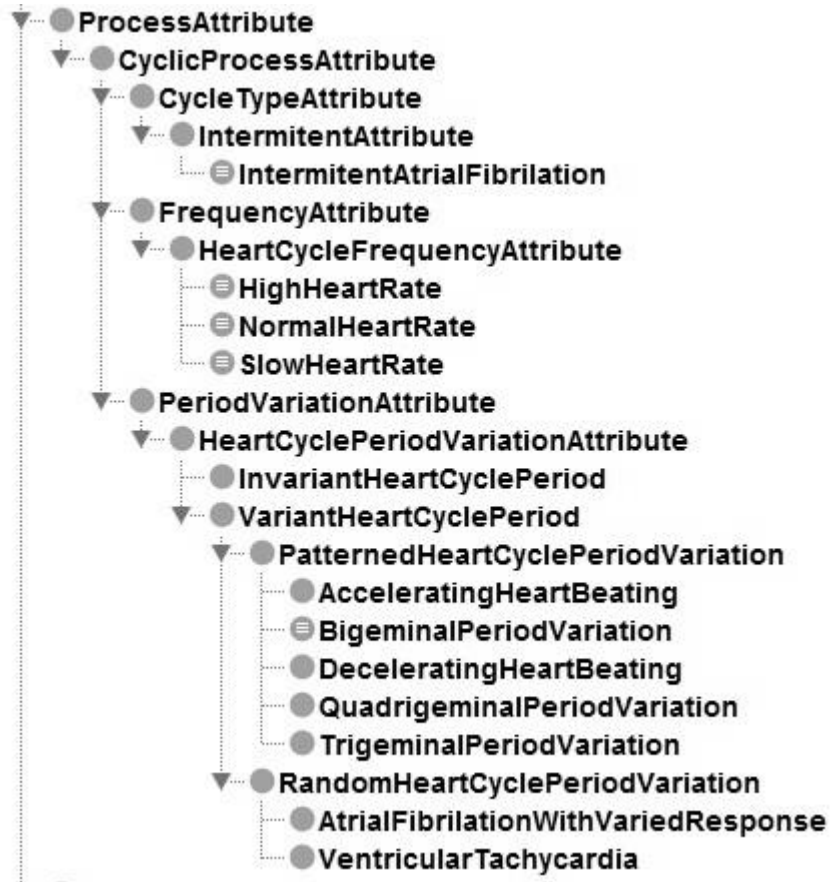
The OWL model is built around the two classes *heart beating process* and *heart cycle*. Whereas the former is a homomereous entity (there are parts of heart beating processes, which are, again heart beating processes) (GREWE, 2011), the latter refers to the events that occur from the beginning of one heartbeat to the beginning of the next. A sequence of members of the class *heart_cycle* constitutes all members of the class *heart_beating_process*. If we cut a temporal region, which spans over a minute of this beating process we can count the number of heart cycles that take place in that time period. The class *heart_cycle* has common properties to every cyclical process; and the heart beating process, which for any given period of time consists of a fiat collection of heart cycles. The generic classes were named “*single cycle*” and “*aggregate of cycles and their parts*”, respectively. This can be seen in **Figure XX and XX**. We relate the cycles and the aggregate using the BioTop(BEIBWANGER *et al.*, 2008) relation “hasGranularPart”, since our goal is to relate collectives (aggregate of cycles) and the grains (the cycles themselves) that compose it (JANSEN; SCHULZ, 2011a).



Heart Beating Process



The second pattern was a distinction between structural dimensions of aggregate of cyclic processes. Following a corollary of the physical properties of cycles, we created three classes, according to frequency of the cycle (number of cycles in a given time), the variation between their periods (being the lack of significant variation called “*regular cycle*”) and the types of cycles that compose the aggregate of cycles. This can be seen in **Figure XX**.



To test the validity of these distinctions, OWL axioms were created, which are described below.

Our first axiomatization attempt involved cardinality properties of the heart beating process. According to this view, every member of the class *heart_beating_process* is therefore the summation of individual cycles, members of the class *heart_cycle*.

HeartBeatingProcess **subClassOf** **hasGranularPart** min 2 *HeartCycle* (1)

HeartBeatingProcess **subClassOf** *process* (2)

HeartCycle **subClassOf** *process* (3)

process **subClassOf** *exactly 1* **hasDuration**

(4)

The calculation of a heart rate, though at first sight straightforward, has several possible measurement procedures, particularly for detecting heart rate variation (TASK FORCE OF THE EUROPEAN SOCIETY OF CARDIOLOGY THE NORTH AMERICAN SOCIETY OF PACING ELECTROPHYSIOLOGY, 1996). For the present purposes, we will consider that the measurement procedure does not alter the ontological description of a heart rate, namely, the description of how many individual heart cycles occurred in a process within a given time interval. Therefore, we define beating process according to their duration as measured in seconds, and we used the IAO to define the measurement as an information artifact.

SixtySecondHeartBeatingProcess **equivalentTo** *HeartBeatingProcess* and

hasDuration *some* *SixtySecondTemporalInterval*

(5)

OWL syntax allows us to express these counts by cardinality restrictions and pre-setting the duration of the analyzed process, in order to allow classification.

NormalHeartRate **isProcessAttributeOf** *only* (*SixtySecondHeartBeatingProcess* and

(**hasPart** *min 60* *HeartCycle*) and (**hasPart** *max 100* *HeartCycle*)

(6)

The use of reasoners such as FaCT++ (version 1.5.3) and Hermit (version 1.3.5) showed that such axiomatization is practically infeasible, since the addition of this axiom increased classification time by more than 10 minutes. Both classifiers took very similar processing time in a quad-Core 8 GB RAM laptop computer. Even more importantly, the results of classification were never obtained – the classification process did not appear to terminate. Since the goal of this paper is not an evaluation of reasoners, this generic axiom was modified to use OWL data properties, maintaining the general modeling approach.

HeartCycleFrequencyAttribute and (**isProcessAttributeOf** some

(*SixtySecondHeartBeatingProcess* and (**hasNumberOfGrains** some int[≥ 60 , ≤ 100])))

(7)

Here, classification time was reduced; the time taken with the cardinality constraint was in the order of 1892256 ms (Fact++, but the results of both reasoners were comparable), but only 12967 ms for the data property processing. While this alternative representation of cardinality using data properties was successful in properly classifying heart rates, a different approach was required for representing acceleration and complex arrhythmic patterns. An acceleration pattern could be translated as “an attribute of a collection of sequential cycles, in which every cycle, except the first, has a shorter duration than the one that immediately precedes it” (more tolerant and realistic variations would allow a certain amount of exceptions). But already this most simple definition proved impossible to represent in OWL, since it requires keeping track of individuals across a variable number of cycles. Therefore, we represented rhythm patterns as primitive classes.

10.2.3.2 Evaluation

According to the pre-defined competency questions, we obtained the following results, addressing the above mentioned:

Problem	Representation	Comments
1.a. Normal heart rate for a 30 year-old	<i>HeartCycleFrequencyAttribute</i> and (isProcessAttributeOf only (<i>SixtySecondHeartBeatingProcess</i> and (hasPart min 60 <i>HeartCycle</i>) (hasPart max 100 <i>HeartCycle</i>)))	The model allowed proper representation and automatic classification of those classes. Such representation can be done through fully defined classes – therefore precluding the necessity of process attributes
1.b. Fast heart rate for a 30 year-old	<i>HeartCycleFrequencyAttribute</i> and (isProcessAttributeOf some (<i>SixtySecondHeartBeatingProcess</i> and (hasNumberOfGrains only int[> 100])))	
2.a. Drug X is contra-indicated in	<i>BradycardiacPatient</i> equivalentTo <i>Human</i> and (hasPart some (<i>Heart</i> and participatesIn some	

bradycardia; patient Y has bradycardia -> Drug X is contraindicated for patient Y	<i>(HeartBeatingProcess</i> and hasProcessAttribute some <i>SlowHeartRate</i>))	
2.b. Knowing that the patient took drugs X, Y and Z, and knowing that the heart frequency of the patient has increased from base line, query which of the drugs could have caused such increase.	<i>TachycardicEffect</i> equivalentTo <i>DrugFunction</i> and (inheresIn some (<i>Drug</i> and participatesIn some (<i>TherapeuticMedicationAdministration</i> and precedes some (<i>HeartBeatingProcess</i> and hasProcessAttribute some <i>AcceleratingHeartBeating</i>))))	
2.c. Querying a database for diseases that co-occur with some arrhythmia	SELECT ?Disease WHERE <i>?ParticularDisease</i> isInstanceOf ?Disease. <i>?ParticularDisease</i> inheresIn ?Human. <i>?Human</i> hasPart ?Heart. <i>?Heart</i> participatesIn ?HeartBeatingProcess. <i>?HeartBeatingProcess</i> hasProcessAttribute ?ProcessAttribute. <i>?ProcessAttribute</i> isInstanceOf <i>ArrhythmicHeartRate</i> .	
3.a. "Paroxysmal Atrial Fibrillation (diagnosis)"	<i>AtrialFibrillationHeartBeating</i> equivalentTo <i>HeartBeatingProcess</i> and (hasPart only <i>AtrialFibrillationCycle</i>) and hasNumberOfGrains some integer [>1] <i>SANodeHeartBeating</i> equivalentTo <i>HeartBeatingProcess</i> and (hasPart only <i>SANodeStartedHeartCycle</i>) and hasNumberOfGrains some integer [>1]	Atrial fibrillation can be modelled as a collection of a different type of cycle, where the atrial contraction is not regulated by the SA node. The paroxysmal can be modelled as a process attribute of processes that have as part more than 1 instance of the same type of process.

	<p style="text-align: center;"><i>IntermittentAtrialFibrillationAttribute</i></p> <p>equivalentTo isProcessAttributeOf some <i>(HeartBeatingProcess</i> and (hasPart min 2 <i>AtrialFibrillationHeartBeating)</i> and (hasPart some <i>SANodeHeartBeating))</i></p>	
3.b. "Regular cardiac rhythm"	$(\forall x)(\forall y)(\forall z) (HeartBeating(x) \wedge HeartCycle(y) \wedge HeartCycle(z) \wedge has_attribute(x, Invariant_heart_cycle_period) \wedge (y \neq z) \wedge partOf(y,x) \wedge partOf(z,x) \rightarrow (durationOf(y) = durationOf(z)))$	Fully defined in FOL by stating that the duration of each cycle part of the heart beating has the same duration as any other cycle part of the heart beating.
3.c. "sudden onset of palpitation"	<p>hasProcessAttribute some <i>(NormalHeartRate</i> and precedes some <i>AcceleratingHeartBeating)</i></p>	Palpitation will require some primitive as acceleration has been shown as difficult to represent in OWL. However, the sudden can be expressed as a process attribute "normal heart rate" that precedes an acceleration attribute, and which the time instant of the end boundary of the former is the same as the time instant of the initial boundary of the latter.
3.d. "History of supraventricular tachyarrhythmia"	<p>isAbout some <i>HeartBeatingProcess</i> and (hasProcessAttribute some <i>HighHeartRate)</i> and (hasPart only <i>SANodeStartedHeartCycle)</i></p>	Would require a primitive, but the history can be expressed by IAO's solution of isAbout some supraventricular tachyarrhythmia.
3.e. "no atrium-ventricular or intra-ventricular conduction abnormalities (EKG finding)"	<p>isAbout some <i>HeartBeatingProcess</i> and (hasProcessAttribute only <i>NormalHeartRate)</i> and (hasProcessAttribute only <i>InvariantHeartCyclePeriod)</i> and (hasPart only <i>SANodeStartedHeartCycle)</i></p>	Already covered in b and d. However, it is important to note that in this modelling, every change in conduction on a single heart cycle requires a representation of a different type. Therefore, there is little or no information making a distinction between a SA started cycle with no AV block and one a heart of a Wolff-Parkinson-White syndrome.
3.f. "chronic	<p>isAbout some <i>ChronicAtrialFibrillationAttribute</i></p>	Similar to a, but in

atrial fibrillation (diagnosis)"	equivalentTo (<i>HeartBeatingProcess</i> and <i>AtrialFibrillationCycle</i>)	isProcessAttributeOf (hasPart only)	some only	this case the beating has only as part atrial fibrillation cycles.
----------------------------------	--	---	-----------	--

As can be seen in **Table XX**, most statements can be successfully expressed by describing the individual cycles and the number of cycles in a given time frame, making the language statements logically sound and internally coherent. However, some classes that refer to acceleration could only be formalized in first-order logic, which is further explored in the next section.

10.2.3.3 Formalization of process patterns:

We made an attempt to detail some points relating to a first-order logic (FOL) representation of the above process patterns, bearing in mind that most bio-ontologies are currently formalized in OWL or the OBO language and therefore would not be able to directly make use of such a formalization. Emerging tools for the integration of ontology modules specified in different logical frameworks could provide a pragmatic solution to this problem in the near future (KUTZ *et al.*, 2011).

A process q is constituted by a sequence of part processes $P = (p_1, p_2, \dots, p_n)$

All processes x have a duration $\text{dur}(x)$ which is denoted by a real number in some time scale

$$\bigwedge_{k=1}^n \text{hasPart}(q, p_k)$$

$$\bigwedge_{k=2}^n \text{follows}(p_k, p_{k-1})$$

(8)

Ideal accelerating process: $\text{follows}(p_k, p_{k-1}) \rightarrow \text{duration}(p_k) < \text{duration}(p_{k-1})$

Ideal decelerating process: $\text{follows}(p_k, p_{k-1}) \rightarrow \text{duration}(p_k) > \text{duration}(p_{k-1})$

Ideal even process: $\text{follows}(p_k, p_{k-1}) \rightarrow \text{duration}(p_k) = \text{duration}(p_{k-1})$

How much real rhythmic processes are to be classified requires a fiat division of the continuum between the ac- (de-)celerating and even processes. Considering the numerical representation of the durations of each individual cycle,

(5, 5, 4, 4, 5, 3, 3, 2, 2, 2, 3, 2, 1, 2, 1, 1) would be closer to an accelerating process than to an even process.

10.2.4 Discussion

While some authors argue that ontological realism should be relaxed in some aspects of ontology modeling (LORD; STEVENS, 2010) due to its apparent over-complexity, the creation of ad-hoc new primitive classes has unforeseeable downstream consequences. One of the main benefits of the realist approach is to allow modeling convergence despite domain-specific and application-specific perspective differences by using scientific results and the interdisciplinary bridging perspective of philosophical ontology as a methodology to arrive at more precise and unambiguous ontological structure as a substitute for unexamined natural language assertions such as form the strategy behind terminological resources such as MeSH. Representations of the world according to the consensual scientific discourse guarantee reliable and robust representation artifacts (JANSEN; SCHULZ, 2011b). Our description of process attributes maintains this principle, but at the same time provides sufficient expressivity to meet the domain requirements.

When analyzed more deeply, process attributes as used in domain terminologies reveal themselves to be a loosely related set of descriptions, reifications and analogies used to communicate some characteristics of events in natural language. The most basic characteristic of the heart cycle is the fact that it is a cycle and can be described according to its frequency and variation of sequential periods. A frequency of a cycle refers to its cardinality within some given time which can itself be fully expressed using primitive ontological constructs. Likewise, the duration of periods are real entities and require no special construct. Time intervals are occurrents, just as processes, and therefore not qualities. Durations of time intervals are numeric values.

Despite the underlying compatibility with primitive constructs, full definition of attributes in our use case was possible only in some cases due to limitations in the

underlying logical description language. The right degree of expressivity of a logical language has been subject to very long dispute. However, due to its balance between expressivity, decidability, and most importantly, due to its being an official W3C recommendation, the Web Ontology Language (OWL-DL) is the *de facto* standard for representing ontologies across many domains including that of biomedicine. OWL is based on description logics, which are a family of carefully selected decidable subsets of first order logic. Our main difficulty in defining some attributes (such as acceleration of heart rate) results from the lack of generalized support for arithmetical relations on individuals, such as *greater than* and *less than*, which are needed in our acceleration examples between instances of ordered cycles. Therefore, we could not assert that *cycle n* immediately precedes *cycle n+1* and has a greater duration than *cycle n+1*.

While not completely representable in OWL, we argue that the FOL definition we provided for some process attributes still contributes to restricting possible interpretations of the class. We can, for instance, distinguish between an strictly accelerating heart beating (every cycle is longer than the one it precedes), a constant accelerating cycle (every cycle is longer than the one it precedes, being the difference between any 2 sequential cycles duration always the same) and an accelerated heart beating, or palpitation (as defined in 3.c.i). The distinctions were shown to be quite adequate for representing the general meaning of common expressions and categories for cardiac arrhythmias, when tested against the criteria. It is important to emphasize, however, that natural language use of these words is much more relaxed and context-dependent, which may require local adaptations according to application needs. Also, many distinctions require the proper identification and classification of types of cycle, which is a rich subject area of its own and out of scope for the current work.

A special case is the description of rhythm patterns like the bigeminal⁷² rhythm. While at first classified under patterned period variation – since its rhythm is easily recognizable, with a short cycle followed by a long cycle followed by a short cycle, and so on – it can also be described according to the origin of the electrical impulse leading to heart contraction (supra-ventricular and ventricular).

⁷² Bigeminal rhythm: Heart beating characterized by a normal sinus beat succeeded by a premature beat – therefore, the beats occur in pairs, showing a particular rhythm.

10.2.4.1 Application to other use cases

It is important to highlight that the mentioned design patterns apply exclusively to cyclic processes. Our approach can also be applied to attributes of biomedically relevant processes aside from heart cycles. For example, a pain process can be understood as the summation of nociceptor stimuli. Here, however, not the duration of the action potential matters but the frequency with which action potentials are produced by a group of nociceptors. Allowing the exact description of the process does not mean that such an exact measurement is possible in clinical practice. As discussed in (SCHULZ; KARLSSON, 2011), separation between the fact and information about the fact can be used to properly describe this situation (using the OWL 'only' operator). Also, many pain-related entities common to clinical practice are epistemological entities, which must be carefully evaluated for suitability in realist ontologies (ANDRADE; ALMEIDA, 2011b).

Pregnancy is another highly complex process, due to its mutually coordinated structural and functional changes in (at least) two organisms. The pregnancy process, focusing on the mother's organisms is commonly dissected by fiat into three trimesters, whereas the development of the offspring is split into embryogenesis and fetal development. The sub-process that terminates the pregnancy is the delivery, which again, can be split into a series of processes, such as the sequence of configuration of the baby's head and body within the birth channel, and the progress of the mother's labor. The variants of the pregnancy process are manifold in terms of

- duration
 - of the whole process, or process parts, such as labor
 - of repetitive phases such as uterine contractions in relation to the intermittent latency phases
- count
 - number of contractions
- intensity

- pain, contractions
- extra process parts
 - surgical interventions such as episiotomy or caesarean section
 - complications of pregnancy such as eclampsia or diabetes
- missing process parts
 - failure of descent of the fetal head
 - embryonic defects

The pregnancy is also characterized by its participants (mother, offspring), their related body parts and qualities, such as number of offspring, their size, missing or supernumerary parts etc.

Due to the myriad of determinants of a pregnancy process, a classification into “normal” and “abnormal” cannot be reduced to hard criteria. Apart from some extreme situations (e.g. fetal death, miscarriage), the boundary between the normal and the abnormal is fuzzy, as is common in medicine. We argue that the correct description of participants and sub-processes allow proper comparison of different abnormalities, without the arbitrary creation of different terms. Our approach promotes the precise description of each occurrent and participant of the pregnancy process, in order to maintain modeling coherence and accurate representation. For instance, premature labor could be defined according to the time span between each contraction cycle or the cardinality of contraction within a given time span, and the occurrence of these contractions within the time interval spanning from conception to the 37th week after conception. Therefore, it is clear what makes normal labor and premature labor pregnancies similar and what is the distinction between the normal and pathological parts.

As pointed out by the heart cycle example, the logical language (in this case OWL-DL) imposes limits on what can be adequately represented therein. The proper evaluation of this limitation in ontological representation and reasoning remains to be evaluated. However, it is

important to point out that ontological analysis here proposed is independent of particular representations, and is coherent with the philosophical view put forward in BFO foundational papers. Particularly, this approach is coherent with the view that “processes do not change, because processes are changes” put forward by Smith (SMITH). It is also compatible with BFO 2.0, which introduces *Process profile* as a special sort of processual parts (SMITH *et al.*, 2012). In this paper, we do not propose a different interpretation, but have rather outlined a complementary approach than a proper ontological definition of complex processes. Therefore, instead of determining profiles according to an ad hoc structural dimension of a process, process attributes require a precise definition in terms of the kinds of participants, participant qualities and sub-processes that characterize the (attributed) process.

10.2.5 Final remarks

Attributes of process – or process qualities – are common descriptions within most communities. While BFO 2.0 is now introducing a new category to fill this gap, modeling restrictions are still required to promote interoperability. Strict adherence to engineering guidelines and best practices of logical representation ensures that the resulting ontology will be adequate to the domain and useful for specific applications. We have shown that processes qualities can be successfully represented by the duration of the whole process, its parts (sub-processes), their participants, and the qualities of the participants. One could therefore argue that process qualities are not justified as first-class citizens in biomedical ontologies. They should rather be included for convenience as fully defined classes. However, as their full definition often requires logical machinery that exceeds the capabilities of current reasoning devices, and may adversely impact reasoning performance even in cases where the expressivity is supported, there may be a pragmatic need to accommodate them as primitives.

By describing in detail the application of this pattern to heart rate modeling, we have shown the suitability of cardinality, distinction of parts of process and cycle properties to define process attributes. Further work is required to refine these patterns of representation and increase coverage of the approach, while maintaining logical and philosophical consistency.