

UNIVERSIDADE DE CAXIAS DO SUL
Centro de Ciências Exatas e Tecnologia
Curso de Bacharelado em Ciência da Computação

Rodrigo Pasini

**ESTUDO E PROJETO DE UMA ONTOLOGIA PARA A ÁREA DA
SAÚDE**

Caxias do Sul

2009

Rodrigo Pasini

**ESTUDO E PROJETO DE UMA ONTOLOGIA PARA A ÁREA DA
SAÚDE**

Trabalho de Conclusão de Curso
para obtenção do Grau de
Bacharel em Ciência da
Computação da Universidade de
Caxias do Sul.

**Carine Geltrudes Webber
Orientadora**

Caxias do Sul

2009

RESUMO

Este trabalho apresenta um estudo sobre a aplicação de ontologias como solução para organizar os diversos vocabulários existentes na área da Saúde. Essa proposta vem de encontro a proposta do projeto OntoData, o qual visa estudar ontologias na descoberta de conhecimento em Saúde. Tal estudo apresenta como fundamentação teórica uma breve descrição sobre as teorias existentes no estudo de ontologias, como conceitos, tipos de ontologias, critérios, projetos, recursos e metodologias utilizadas no desenvolvimento de ontologias. Como aplicação do conhecimento adquirido com base nas teorias apresentadas, este trabalho também apresenta o desenvolvimento de duas ontologias utilizando o *software Protégé*, a primeira ontologia é totalmente baseada na Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde (CID10), a segunda ontologia reaproveita a ontologia CID10 e apresenta os Indicadores de Dados Básicos para a Saúde no Brasil (IDB) como base para seu desenvolvimento. O processo do desenvolvimento de ambas as ontologias é detalhado passo a passo com base na metodologia escolhida para desenvolvimento das mesmas.

Palavras-chaves: ontologia, CID10, conhecimento em saúde, Protégé

ABSTRACT

This work presents a study on the application of ontologies as solution to organize the diverse existing vocabularies in the area of the Health. This proposal comes of meeting the proposal of the OntoData project, which aims at to study ontologies in the discovery of knowledge in Health. Such study one brief description presents as theoretical recital on the existing theories in the study of ontologies, as concepts, types of ontologies, criteria, projects, resources and methodologies used in the development of ontologies. As application of the acquired knowledge on the basis of the presented theories, this work also presents the development of two ontologies using Protégé software, the first ontology is total based on the *Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde (CID10)*, the second ontology reuses the ontology CID10 and presents the *Indicadores de Dados Básicos para a Saúde no Brasil (IDB)* as base for its development. The process of the development of both the ontologies is step by step detailed on the basis of the methodology chosen for development of the same ones.

Keywords: ontology, CID10, health knowledge, Protégé

LISTA DE FIGURAS

Figura 1 – Classificação das ontologias quanto a Generalidade (GUARINO, 1998).....	7
Figura 2 – 7 Passos no desenvolvimento de uma ontologia pelo guia Ontology Development 101.....	17
Figura 3 – Definição de taxa de mortalidade pelo DeCS.....	20
Figura 4 – Exemplo de Capítulos, Grupos, Categorias e Subcategorias da CID10.....	21
Figura 5 – Tela do Protégé.....	23
Figura 6– Plugin Instance Tree.....	25
Figura 7 – Definição das tabelas temporárias onde é importado os arquivos da CID10.....	32
Figura 8 – Importando os arquivos nas tabelas temporárias.....	33
Figura 9 – Criar o arquivo e cabeçalho da owl.....	33
Figura 10 – Inserir classes e propriedades de classes no arquivo.....	34
Figura 11 – Gerar cada instância de classe e valorar as referências das instâncias.....	34
Figura 12 – Exemplo de estrutura gerada no arquivo CID10.owl.....	35
Figura 13 – Hierarquia de classes da ontologia CID10 visualizada pelo plugin Jambalaia.....	35
Figura 14 – Exemplo da árvore de instâncias da ontologia CID10 visualizada pelo plugin Instance Tree.....	36
Figura 15 – CID10 importada na nova ontologia.....	37
Figura 16 – Matriz de indicadores – Mortalidade.....	38
Figura 17 – Modelagem das Classes referentes aos Indicadores no Protégé.....	40
Figura 18 – Propriedades modeladas no Protégé.....	41
Figura 19 – Restrição da propriedade causada_por.....	42
Figura 20 – Instâncias da classe Unidade_geográfica.....	45
Figura 21 – Propriedades a serem valoradas para cada indicador de mortalidade.....	45
Figura 22 – Valores atribuídos aos comentários do indicador de mortalidade Taxa de Mortalidade por Diabete Melito.....	46
Figura 23 – Valores atribuídos às propriedades do indicador de mortalidade Taxa de Mortalidade por Diabete Melito.....	46
Figura 24 – Visualização da ontologia de Domínio da Saúde através do plugin Jambalaia.....	47

LISTA DE TABELAS

Tabela 1 – Lista de termos encontrados na CID10.....	29
Tabela 2 – Classes da ontologia CID10.....	29
Tabela 3 – Propriedades da ontologia CID10.....	30
Tabela 4 – Restrições das propriedades de objeto.....	30
Tabela 5 – Lista de termos encontrados ao analisar a matriz de indicadores e a ficha de qualificação dos indicadores.....	38
Tabela 6 – Classes da ontologia Indicadores.....	40
Tabela 7 – Propriedades das classes definidas.....	41
Tabela 8 – Restrições das propriedades de objeto.....	42
Tabela 9 – Lista de classes e instâncias.....	43

LISTA DE ABREVIATURAS E SIGLAS

Sigla	Significado
ABL	<i>Advanced Business Language</i>
CID10	Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde
CRM	<i>Customer Relationship Management</i>
DAML	<i>DARPA Agent Markup Language</i>
DATASUS	Departamento de Informática do Sistema Único de Saúde (SUS)
DeCS	Descritores em Ciência da Saúde
ERP	<i>Enterprise Resource Planning</i>
HL7	<i>Health Level Seven</i>
HPKB	<i>High Performance Knowledge Bases</i>
IDB	Indicadores de Dados Básicos para a Saúde no Brasil
KB	<i>Knowledge base</i>
KIF	<i>Knowledge Interchange Format</i>
Mesh	<i>Medical Subject Headings</i>
OIL	<i>Ontology Interchange Language</i>
OTICSSS	Observatório das Tecnologias de Informação e Comunicação em Sistemas e Serviços de Saúde
OWL	<i>Web Ontology Language</i>
RDF	<i>Resource Description Framework</i>
RIPSA	Rede Interagencial de Informações para a Saúde
SNOMED	Nomenclatura Sistematizada de Termos Clínicos na Medicina
UMLS	<i>Unified Medical Language System</i>
W3C	<i>World Wide Web Consortium</i>
XML	<i>Extensible Markup Language</i>

SUMÁRIO

1	Introdução.....	1
1.1	Objetivos.....	2
1.2	Organização do Documento.....	2
2	Ontologias.....	4
2.1	Conceituação.....	4
2.2	Tipos de Ontologias.....	5
2.2.1	Quanto a Função.....	6
2.2.2	Quanto ao Grau de Formalismo.....	6
2.2.3	Quanto a Generalidade.....	7
2.2.4	Quanto a Aplicação.....	7
2.2.5	Quanto a Estrutura.....	8
2.2.6	Quanto ao Conteúdo.....	8
2.3	Critérios para Construção de Ontologias.....	9
2.4	Projetos de Desenvolvimento de Ontologias.....	9
2.5	Recursos para desenvolvimento de Ontologias.....	10
2.6	Metodologias.....	12
2.6.1	Methontology.....	13
2.6.2	Metodologia on-to-knowledge.....	14
2.6.3	Cyc.....	14
2.6.4	Metodologia de Uschold and King.....	15
2.6.5	Metodologia de Gruninger and Fox.....	16
2.6.6	Kactus.....	16
2.6.7	Sensus.....	16
2.6.8	Ontology Development 101.....	17
2.7	Vocabulários na área da Saúde.....	18
2.7.1	UMLs.....	18
2.7.2	DECS.....	19
2.7.3	SNOMED.....	20
2.7.4	HLL7.....	20
2.7.5	CID 10.....	21
2.8	Ferramentas para Desenvolvimento de Ontologias.....	22
2.8.1	Ferramenta Protégé.....	22
2.9	Considerações Finais.....	25
3	Desenvolvimento.....	26
3.1	Indicadores de Dados Básicos para a Saúde no Brasil (IDB).....	26
3.2	Metodologia Utilizada no Desenvolvimento.....	27
3.3	Ontologia CID10.....	28
3.4	Ontologia Domínio da Saúde.....	36
3.5	Considerações Finais.....	48
4	Conclusão.....	49

4.1 Síntese.....	49
4.2 Contribuições.....	50
4.3 Trabalhos Futuros.....	50
5 Referências.....	51

1 INTRODUÇÃO

Com a crescente utilização de sistemas informatizados na área da saúde há uma grande demanda pela integração dos dados oriundos de aplicações nessa área. Diferentes vocabulários são utilizados diariamente pelos profissionais e pelos sistemas, o que faz com que existam diversas terminologias em uso simultaneamente na área. Isso pode ser atestado pela grande variedade de dicionários de termos encontrados em livros e na internet (DeCS, Mesh, UMLS, Snomed, entre outros).

Por outro lado, o desenvolvimento do conhecimento na área da saúde demanda aos profissionais da área que cada vez mais mantenham seus conhecimentos atualizados. Porém com tanta informação disponível e com a infinidade de formas de como a mesma informação pode ser representada acaba ficando difícil o compartilhamento do conhecimento, visto que devido a evolução da informática na área da saúde as informações são disponibilizadas em bases heterogêneas e essas bases rodam em plataformas de *hardware* distintos.

Como pesquisar em tanta informação de forma que o resultado obtido como resposta seja suficiente e completo para auxiliar à uma tomada de decisão rápida e precisa, tornando o atendimento a pacientes mais rápido e eficaz?

Na computação existem ferramentas que permitem a representação do conhecimento fazendo uso de terminologias que se aplicam também a integração de sistemas. As ontologias constituem um caso particular destas ferramentas, sendo já utilizadas na área da saúde. Uma ontologia pode ser vista como uma estrutura de dados e suas regras, onde juntos representam um conjunto de conceitos dentro de um domínio e os relacionamentos entre estes. São também descritas por esquemas formados em sistemas que se relacionam entre si, onde estes fornecem uma descrição coerente dos dados compartilhados, permitindo programas que agem e compartilham informações (Gruber, 1993).

Algumas instituições propuseram padrões de documentação como uma iniciativa para tentar organizar o conhecimento na área da saúde, criando recursos como a UMLs (*Unified Medical Language System*), o DeCS (Descritores em Ciência da Saúde), o HL7 (*Health Level Seven*), entre outros. Algumas dessas instituições, porém, utilizam vocabulários proprietários, ou não codificam algumas informações, tornando o compartilhamento do conhecimento algo muito difícil de ser realizado.

1.1 OBJETIVOS

Este trabalho contextualiza-se no projeto OntoData (ontologias para Sistemas de Informação em Saúde). O projeto OntoData é um projeto em desenvolvimento pela Universidade de Caxias do Sul sob coordenação da professora Carine Geltrudes Webber. Ele tem por objetivo estudar ontologias e descoberta de conhecimento em Saúde a fim de investigar métodos de representação de conhecimento para apoio a processos de aprendizagem, análise e gestão de conhecimento. Sistemas da área da saúde são sistemas distintos dos sistemas tradicionais como ERP (*Enterprise Resource Planning*), CRM (*Customer Relationship Management*) e outros que apenas computam dados numéricos. Sistemas da área da saúde manipulam informações clínicas sobre pacientes, relações entre conceitos da área médica, raciocínio de um processo de tomada de decisão, etc.

Este trabalho de conclusão tem por objetivo pesquisar e desenvolver uma ontologia para a área da saúde, no domínio de mortalidade. A ontologia proposta serve como base para construção de futuras ontologias a serem aplicadas no projeto OntoData.

Ainda este trabalho tem por objetivo apresentar o *software Protégé*. Este *software* é um ambiente gráfico desenvolvido pelo Departamento de Informática Médica da Universidade de *Stanford*, livre e de código aberto utilizado na criação e edição de ontologias, bem como de bases de dados guiadas por uma ontologia. O *Protégé* possui um *plugin* para o UMLS (*Unified Medical Language System*), que juntamente com outros padrões como o DeCS (Descritores em Ciência da Saúde) e o CID10 (Décima Revisão da Classificação Internacional de Doenças e de Problemas Relacionados à Saúde) servem como fonte de pesquisa para o desenvolvimento da ontologia proposta.

1.2 ORGANIZAÇÃO DO DOCUMENTO

O Capítulo 2 apresenta os referenciais teóricos seguidos neste trabalho. Define-se nele o que é uma ontologia, quais suas principais metodologias e principais aplicações, e ainda relaciona-se alguns padrões de vocabulários da área da saúde. Neste capítulo também é apresentado o *software* de apoio a modelagem de ontologias *Protégé*, amplamente utilizado para construção de ontologias no mundo todo.

O Capítulo 3 define a metodologia aplicada no desenvolvimento da ontologia, explora o material utilizado como fonte de pesquisa do domínio da ontologia, demonstra o desenvolvimento da ontologia no *software Protégé*, e ainda apresenta a utilização de alguns *plugins* do *Protégé* na

visualização da ontologia desenvolvida.

Por fim, o Capítulo 4 apresenta uma síntese geral do trabalho, contribuições e trabalhos futuros que podem ser realizados com base nos dados deste trabalho.

2 ONTOLOGIAS

Esta revisão bibliográfica tem como objetivo apresentar definições e conceitos básicos sobre ontologias, os tipos, as aplicações em projetos desenvolvidos até o momento, as ontologias existentes, as metodologias e recursos utilizados no seu desenvolvimento, além dos principais padrões de vocabulários utilizados na construção desta na área médica; e a ferramenta de apoio à construção de ontologias, *Protégé*.

2.1 CONCEITUAÇÃO

As ontologias estão sendo utilizadas ao decorrer da história com a finalidade de descrever estruturas, conceitualizá-las e classificá-las seguindo padrões para informatizar e compartilhar os dados acerca da inteligência artificial. Vários são os conceitos utilizados pelos autores que cada vez mais se aprofundam no assunto, mas um deles ficou conhecido e é a base para tais definições.

Gruber (1993) conceitua uma ontologia como “especificação formal e explícita de uma conceituação, na qual definições associam nomes (no universo de discurso) com todos que descrevem seus significados e axiomas que restringem sua interpretação e uso”. Conceituação (Grubber, 1993) é definida como uma coleção de objetos e suas interações, explícito dentro desta definição assume uma posição de restrição ao uso de alguns conceitos e formais é inserido como a leitura feita pela máquina de maneira clara e transparente sobre a ontologia. Varias outras são as definições, mas todas caem na mesma proposta descrita por Gruber, alguns autores complementam o seu conceito, mas o conceito do autor descrito acima é o mais utilizado e de fácil entendimento.

Ainda segundo Gruber (1993), uma ontologia é formada pelos seguintes componentes:

- Conceitos: podem representar qualquer coisa em um domínio, como uma tarefa, uma função, uma estratégia, etc;
- Relações: representam um tipo de interação entre os conceitos no domínio, sendo a cardinalidade sempre n:n;
- Funções: são um caso especial de relações, sendo a cardinalidade n:1;

- Axiomas: são as sentenças que são sempre verdadeiras;
- Instâncias: são utilizadas para representar os elementos do domínio.

Para Salé e Cafe (2009), ontologias consistem em uma reunião de elementos desenvolvidos pela engenharia computacional, especificando formas que visam descrever estruturas conceituais de domínios exclusivos.

São também descritos por esquemas formados em sistemas que se relacionam entre si, onde estes fornecem uma descrição coerente de dados compartilhados, permitindo programas que agem e compartilham informações (Gruber, 1993).

Ontologias são descritas por Guarino e Giaretta (1995), como uma teoria lógica que fornece um relato explícito e parcial de uma conceitualização. Para estes autores a ontologia pode ser considerada sinônimo de conceitualização sendo uma estrutura que se refere ao estudo da palavra, ou seja, do seu significado, que codifica as regras subentendidas, em um padrão e princípios estabelecidos para uma estrutura ou parte da sua realidade.

Diversos grupos utilizam abordagem própria para engenharia de ontologias, (Gomez-Perez, 1999), muitas delas desenvolvidas fora do contexto da web semântica.

A web semântica é entendida como a cooperação de trabalho entre computadores e seres humanos, ela é uma extensão da web onde interliga as palavras com o intuito de atribuir um significado de um modo que seja perceptível tanto para humano quanto para o computador.

Conforme define Jacob (2003), no ambiente da web semântica uma ontologia não é conceituada parcialmente por um domínio de conhecimento. Ela é uma proposta explícita de uma estrutura semântica compartilhada por uma congregação de usuários, por meio de sistemas de informações recuperadas e automatizadas.

Dessa forma a relação entre ontologia e Web semântica se torna direta, fornecendo meios para padronização de vocábulos, servindo como base para realização de sistemas completos autônomos.

Para Russel e Norvig (2004), a palavra ontologia significa uma teoria particular do ser ou da existência, também definida como uma lista informal de conceitos sobre um domínio.

2.2 TIPOS DE ONTOLOGIAS

Vários são os tipos de ontologias. Na maioria das vezes seguem a mesma linha e contem os

mesmos componentes mas, nem sempre configuram a mesma estrutura. Da mesma forma que possuem propriedades semelhantes, identificamos tipos bem variados, conforme a função, o seu grau de formalismo, a estrutura do seu conteúdo.

2.2.1 QUANTO A FUNÇÃO

Mizoguchi e colegas (Mizoguchi et al., 1995), classificaram as ontologias em três tipos: domínio, tarefa e geral:

A ontologia de domínio fornece um vocabulário conceitual incluindo relacionamentos, atividades e regras que as dominam.

A ontologia de tarefa é um sistema de vocabulário para descrever a resolução de problemas e a estrutura de todas as tarefas existentes no domínio. Ela inclui verbos que fornecem a solução destes problemas.

A ontologia geral avalia vocabulários relacionados a um ciclo de eventos, como por exemplo: tempo, espaço, comportamento, função entre outros.

2.2.2 QUANTO AO GRAU DE FORMALISMO

Uschold e Grüninger (1996) classificaram as ontologias quanto a forma da linguagem, utilizando quatro tipos: altamente informal, semi-informais, semi formais e rigorosamente formal.

A ontologia altamente informal expressa ontologias livremente em uma linguagem natural.

A ontologia semi-informal expressa um forma restrita e estruturada da linguagem natural, aumentando a sua clareza.

A ontologia semi formal utiliza uma linguagem artificial definida formalmente.

A ontologia rigorosamente formal exige termos meticulosamente definidos com a semântica formal, os teoremas e as provas de propriedades como a solidez e a perfeição. Por exemplo *TOVE*¹.

1 <http://www.eil.utoronto.ca/enterprise-modelling/tove/>

2.2.3 QUANTO A GENERALIDADE

Guarino (1998) classifica as ontologias em 4 níveis de generalidade conforme a figura 1:

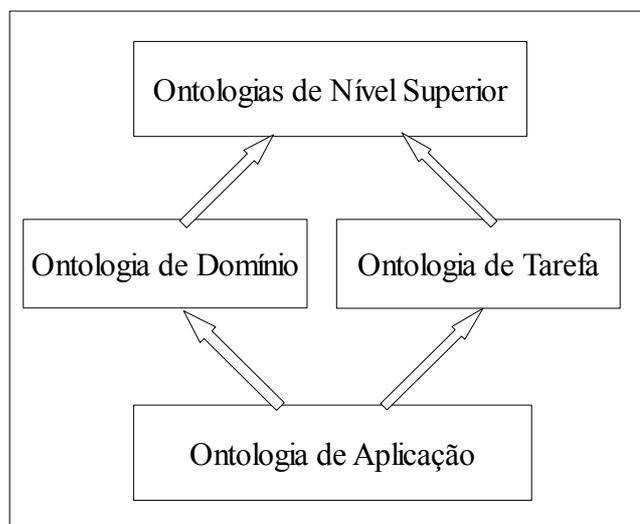


Figura 1 – Classificação das ontologias quanto a Generalidade (GUARINO, 1998)

Ontologias de Nível Superior: descrevem conceitos muito amplos como espaço, tempo, matéria, objeto, evento, ação, etc., os quais são independentes de um problema particular ou de domínio. Dessa forma é possível o compartilhamento de ontologias de nível superior por diversas comunidades de usuários.

As ontologias de Domínio: descrevem vocabulários relacionados com um domínio genérico (como medicina ou automóveis), especializando os termos introduzidos na ontologia de nível superior.

Ontologias de Tarefa: descrevem vocabulários relacionados com uma tarefa ou atividade genérica (como diagnosticar ou vender), especializando os termos introduzidos na ontologia de nível superior.

Ontologias de Aplicação: geralmente descrevem conceitos definidos em ontologias de domínio ou de tarefas, como maneira de especializar ambas as ontologias relacionadas.

2.2.4 QUANTO A APLICAÇÃO

No ano de 1999, Jasper e Uschold descreveram os tipos de ontologia quanto a autoria neutra, quanto a especificação e o acesso comum a informação.

Quanto a autoria neutra, o sistema é apresentado em uma única linguagem, e posteriormente

será convertido para uso em diversos sistemas, sem alterar a sua estrutura.

Quanto a especificação a ontologia, é utilizada para um domínio onde pode-se criar, manter e documentar o desenvolvimento dos programas.

Quanto ao acesso comum a informação, ela proporciona o conhecimento compartilhado das palavras inacessíveis ao vocabulário de forma inteligível.

2.2.5 QUANTO A ESTRUTURA

Haav e Lubi (2001), identificaram três tipos de ontologias, as de alto nível, as de domínio e as de tarefa.

A ontologia de alto nível descreve uma relação de todos os elementos da ontologia (espaço, tempo, etc.) independente do domínio.

A ontologia de domínio relaciona o vocabulário a um domínio específico (matemática, física, etc.).

A ontologia de tarefa descreve uma tarefa sobre termos especializados na ontologia.

2.2.6 QUANTO AO CONTEÚDO

Em 1997, Van Heijst, Schrelber e Wietinga determinaram sete tipos de ontologias, as terminológicas, as de informação, as de modelagem do conhecimento, as de aplicação, as de domínio, as genéricas e as de representação.

Ontologias terminológicas: especificam os termos que representam o conhecimento de um domínio.

Ontologias de informação: especificam a estrutura de registro das bases de dados. Exemplo banco de dados.

Ontologias de modelagem do conhecimento: é especificada pela conceitualização do conhecimento onde apresenta uma rica estrutura.

Ontologias de aplicação: contém definições que são necessárias para modelar o conhecimento através de uma aplicação específica. Estas ontologias são uma mistura de conceitos tomados a partir da ontologia de domínio e da ontologia genérica especificadas abaixo.

Ontologias de domínio: expressam conceitualizações que são específicas para determinados domínio. Ontologias genéricas: apresenta similaridade a ontologia de domínio, mas é definida por conceitos considerados genéricos e comuns em vários campos.

Ontologias de representação: explicam as conceituações que representam o conhecimento através do formalismo. Elas se destinam a ser neutras através das entidades do mundo, ou seja, oferecem a estrutura do trabalho sem a apresentação das críticas sobre o mundo.

2.3 CRITÉRIOS PARA CONSTRUÇÃO DE ONTOLOGIAS

Gruber (1993) aponta alguns critérios essenciais para a construção de ontologias, estes critérios quando utilizados, auxiliam para que o projeto atinja os resultados esperados:

- Clareza: uma ontologia deve comunicar eficazmente o significado pretendido de termos definidos. As definições devem ser objetivas, completas e documentadas em linguagem natural;
- Coerência: uma ontologia deve ser coerente, ou seja, as definições de axiomas devem ser logicamente consistentes e as inferências geradas não podem contradizer a documentação em linguagem natural;
- Extensibilidade: uma ontologia deve ser projetada prevendo o compartilhamento e atualização do vocabulário nela descrito, sem que haja a necessidade de revisar as definições já existentes;
- Limiar de Codificação Mínimo: a conceitualização deve ser especificada no nível de conhecimento, sem depender de uma linguagem especial;
- Compromisso Ontológico Mínimo: para que haja o compartilhamento de informações, uma ontologia deve definir apenas os termos extremamente necessários, independente do domínio que está sendo modelado. Isso torna a ontologia independente, permitindo sua especialização conforme a necessidade de outro utilizador.

2.4 PROJETOS DE DESENVOLVIMENTO DE ONTOLOGIAS

Diversos projetos relacionados ao uso de ontologias vêm sendo desenvolvidos em diferentes áreas de aplicação. A finalidade de estudar, pesquisar e desenvolver estas ontologias é o

compartilhamento da informação e a possibilidade de reuso do conhecimento sobre domínios específicos.

No Brasil há um forte movimento do meio acadêmico no uso de ontologias, sendo que, a partir do ano de 2008 iniciou-se um encontro anual dos pesquisadores de ontologias, denominado de Seminário de Pesquisa de Ontologias no Brasil, o qual teve a sua primeira edição em Setembro de 2008 e a segunda em Setembro de 2009. Como exemplo de trabalho apresentado neste seminário pode-se citar o artigo de Farias et al. (2006). Este artigo descreve o desenvolvimento de uma ontologia aplicada na área médica com ênfase na análise de dados referentes a doenças sexualmente transmissíveis.

Outro exemplo de proposta apresentada foi a dissertação de mestrado de Molossi, (2008), onde a autora apresenta uma dissertação sobre a construção de ontologias para auxiliar na recuperação de pesquisas utilizado pela Biblioteca de Teses e Dissertações do Instituto Brasileiro de Informação e Tecnologia.

Outros exemplos com a utilização de ontologias são facilmente encontrados. Estudá-los significa, reconhecer os trabalhos já existentes, e conhecer diferentes opções de processo de levantamento, modelagem e construção de ontologias.

2.5 RECURSOS PARA DESENVOLVIMENTO DE ONTOLOGIAS

Para desenvolver uma ontologia necessita-se de criatividade, integralidade, capacitação e exímio conhecimento. Apesar de ser um trabalho dispendioso, os benefícios podem ser bem significativos.

São várias as metodologias existentes para a construção de uma ontologia, mas também são desenvolvidos recursos que auxiliam no seu desenvolvimento. Um destes recursos é representado pela linguagem utilizada no desenvolvimento da ontologia.

Várias são as linguagens utilizadas no desenvolvimento de ontologia, algumas serão descritas a seguir:

Ontolingua (FARQUHAR et al., 1997), esta linguagem foi criada pela Universidade de Stanford em 1992, baseada em Knowledge Interchange Format (KIF). Combina paradigmas de frames e predicados de primeira ordem (Chaudhri et al., 1998 apud Almeida & Bax 2003). É uma linguagem de bastante expressão podendo representar conceitos, axiomas, taxonomias e procedimentos;

CycL (LENAT et al., 1990) denomina uma linguagem representada por meio de termos em forma de vocabulário, é utilizada para expressar o conhecimento do senso comum juntamente com o conhecimento armazenado nos banco de dados da *Cyc*. Foi criada pela companhia Cycorp e tem como vantagem a expressividade, a precisão e a representação de significado;

Loom (BRILL, 1993), abrange uma linguagem baseada em lógica descritiva e regras de produção, onde os conceitos, a taxonomia, as funções e axiomas são representados;

F-Logic (KIFER et al., 1995) foi desenvolvida na Universidade de *Karlsruhe* – Alemanha, em 1995, com o objetivo de representar o conhecimento através de um formalismo. Um de seus pontos fortes é a capacidade de representar diretamente um conceito. É uma linguagem clássica sustentada em cálculos de predicados;

RDF (Resource Description Framework) (BRICKLEY et al., 2004), é representada pela ontologia básica clássica e foi criada pela *W3 Consortium*, onde utiliza informações de fontes da internet, que descrevem os termos em expressões RDF escrito em XML e URIs para a identificação de fontes;

SHOE (HEFLIN & HENDLER, 2002), é utilizada para a construção de uma ontologia em forma de marcações da extensão HTML onde insere dados sobre outros dados (meta dados);

XOL (KARP et al., 2002) foi desenvolvida para projetar ontologias de domínio de biomedicina, onde especifica conceitos, taxonomias e relações binárias;

OIL (Ontology Interchange Language) (OIL, 2008) representa uma linguagem que fornece aprimoramento primitivo onde utiliza ontologias baseadas em *frames* e orientadas pela descrição lógica. Foi criada por investigadores europeus e utiliza como sintaxe RDF e XML;

DAML (DARPA Agent Markup Language) + OIL (CONNOLY et al., 2001) utiliza linguagem semelhante a OIL com aperfeiçoamento na inserção da sintaxe DAML;

OWL (Web Ontology Language) (OWL, 2004) foi criada em 10 de fevereiro de 2004 pelo W3C (*World Wide Web Consortium*). Seu projeto foi criado para o uso em aplicações que necessitam processar o conteúdo da informação, ao invés de apenas apresentar informações para os

seres humanos. Tem por característica facilitar uma maior possibilidade de interpretação por máquinas do conteúdo da Web do que XML, RDF e RSFS, por fornecer vocabulário adicional com uma semântica formal.

A OWL possui três sub linguagens projetadas para o uso por comunidades específicas de implementadores e usuários, sendo elas:

1. *OWL Lite*: designada a usuários que necessitam de uma classificação hierárquica e de restrições simples. Por exemplo, para restrições de cardinalidade só são suportadas restrições com valores 0 ou 1.
2. *OWL DL*: suporta os usuários que querem o máximo de expressividade, mantendo completude computacional (há garantias de que todas as conclusões são computáveis) e decidibilidade (todas as computações terminarão em tempo finito). OWL DL inclui todas as construções de linguagem OWL, mas eles só podem ser utilizados com certas restrições (por exemplo, enquanto uma classe pode ser uma subclasse de muitas classes, uma classe não pode ser uma instância de outra classe).
3. *OWL Full*: destinado a usuários que querem a máxima expressividade e liberdade sintática do RDF sem garantias computacionais. Por exemplo, em OWL Full uma classe pode ser tratada simultaneamente como uma coleção de indivíduos e como um indivíduo.

2.6 METODOLOGIAS

As metodologias existentes amplificam os estudos da ontologia por serem de vários tipos onde podem construir ontologias, construir também ontologias em grupos para seu aprendizado e integração. Os diversos tipos de metodologias aqui descritos não se relacionam entre si, para utilizá-los é necessário avaliar a ontologia e compará-la na aplicação de cada metodologia. Vários são os estudos que estabelecem metodologias para o desenvolvimento de ontologias, mas isso ainda está sendo construído, para que seja adotada uma metodologia existente que compreenda cada grupo no seu processo. Serão descritas a seguir algumas metodologias, métodos e processos para projeto e desenvolvimento de ontologias.

2.6.1 METHONTOLOGY

Segundo Fernandez-Lopez et al, (1997), esta metodologia foi desenvolvida no Laboratório de Inteligência Artificial da Universidade Politécnica de Madrid. *Methontology* permite a construção de uma ontologia no nível do conhecimento e inclui a identificação da ontologia no processo de desenvolvimento, no ciclo de vida com base na evolução de protótipos e técnicas específicas para realização de cada atividade. Ela utiliza como atividades principais a manutenção, especificação, conceitualização, formalização e implementação. Estas serão descritas a seguir, segundo Benjamins V., R, et al, (2005):

- A manutenção das atividades corrige a ontologia quando necessário.
- A especificação explica por que a ontologia está sendo construída, quais as suas finalidades e quem são seus usuários finais.
- A conceitualização organiza e converte a visão informal de um domínio em uma especificação semi formal através dos conjuntos de representações intermediárias. O resultado deste modelo define uma ontologia de modelo conceitual.
- A formalização transforma o modelo conceitual em um modelo formal ou semi computável.
- A implementação constrói modelos computáveis em uma linguagem ontológica, por exemplo (*Ontolingua*, RDF, OWL, etc). Automaticamente algumas ferramentas implementam modelos conceituais em idiomas variados, por exemplo (webODE, importações e exportações de ontologias e para os seguintes idiomas: XML, RDF, OIL, DAML+OIL, OWL, Carin, etc.).

METHONTOLOGY identifica também as atividades de gerenciamento (cronograma, controle e garantia de qualidade), e o apoio as atividades do conhecimento (aquisição, integração, avaliação, documentação e gerenciamento de configuração).

A metodologia *METHONTOLOGY* utiliza algumas informações que devem ser incluídas no seu desenvolvimento:

- O propósito da ontologia, incluindo o que se destina as suas utilizações, os cenários de

utilizações, e os usuários finais.

- O nível de formalidade da ontologia implementada.
- O Escopo que inclui o conjunto de codificações representadas, as suas características e granularidades.

2.6.2 METODOLOGIA ON-TO-KNOWLEDGE

A metodologia *On-To-Knowledge* foi desenvolvida e impulsionada pelas aplicações de desenvolvimento da ontologia na gestão de sistema de conhecimento (Fensel et al, 2000). Tem como objetivo a sustentação de um novo processo com base em questões humanas, em que os especialistas em contextos industriais são capazes de criar e gerir as suas próprias ontologias. O objetivo desta metodologia sustenta que todos os documentos eletrônicos na internet sejam transformados em termos e significados do seu conteúdo para os usuários da informação, *por padrões*, para que possam legitimamente falar de uma Web Semântica ou de um conhecimento.

A *On-To-Knowledge* permite filtrar e lidar com o significado em grandes repositórios de documentos, facilitando que o usuário tenha um acesso seletivo e significativo para os documentos eletrônicos.

2.6.3 CYC

A metodologia *Cyc* surgiu da experiência do desenvolvimento da base de conhecimento *Cyc* (KB), que contém uma enorme quantidade de conhecimento do senso comum (CYC, 2009). Três são as fases que devem ser seguidas para a construção da ontologia *Cyc* (Lenat, 1990), estas serão descritas a seguir:

A primeira fase propõe manualmente o código de conhecimento explícito e implícito que aparecem nas fontes de conhecimento em geral, sem a ajuda da linguagem natural e sistemas de aprendizagem. Esta fase é realizada manualmente, uma vez que o sistema de linguagem natural e as máquinas de aprendizagem não manipulam o suficiente o conhecimento do senso comum, mas, buscam o conhecimento do novo senso comum.

A segunda fase propõe codificação do conhecimento que é auxiliado pelas ferramentas que utilizam o conhecimento já armazenado na *Cyc KB*.

A terceira fase delega para as ferramentas a maioria do trabalho.

Duas tarefas são realizadas em cada fase do desenvolvimento da metodologia *Cyc*:

1. O desenvolvimento de uma representação do conhecimento e as ontologias de nível superior que contém a maioria dos conceitos resumidos. Termos como atributo ou valor do atributo são exemplos de termos de representação do conhecimento.
2. Representação do resto do conhecimento utilizando dados primitivos.

Cyc foi utilizado durante a experimentação no *High Performance Knowledge Bases* (HPKB, 2009), um programa de pesquisa para o avanço da tecnologia de computadores, que permite adquirir, representar e manipular o conhecimento.

Esta metodologia é usada apenas para construir a base de conhecimento *Cyc*; porém *Cyc* tem diferentes micro teorias que mostram o conhecimento dos domínios e de outros pontos de vista. Em várias áreas, estas micro teorias podem ser utilizadas, a partir de diferentes hipóteses e perspectivas.

No que diz respeito as aplicações em que são utilizadas as ontologias *Cyc*, existem vários módulos integrados no *Cyc KB*. Um desses módulos baseia-se no sistema de integração de bases de dados heterogêneas. Este módulo mapeia o vocabulário *Cyc* em esquemas de banco de dados, ou seja, os dados que são armazenados em banco de dados são interpretados de acordo com o vocabulário *Cyc*.

Finalizando, o módulo de recuperação de informações *web* utiliza a ferramenta de linguagem natural que disponibiliza acesso ao *Cyc KB*, permitindo estender o *Cyc KB*, utilizando informações disponíveis na *web*.

2.6.4 METODOLOGIA DE USCHOLD AND KING

A metodologia de Uschold e King (1995) propõe quatro atividades: identificar a finalidade da ontologia, construí-la, avaliá-la, e documentá-la.

Durante a atividade de construção os autores propuseram captar o conhecimento, a codificação e a integração de outras ontologias dentro da atual.

Eles também propuseram três estratégias para identificar os principais conceitos de uma ontologia.

A primeira abordagem denomina-se *top-down*, em que os conceitos mais abstratos são

identificados, e especializados em conceitos mais específicos; a segunda abordagem denomina-se *bottom-up*, no qual os conceitos mais específicos são identificados na primeira abordagem e em seguida generalizados em conceitos mais abstratos; a terceira e última abordagem denomina-se *middle-out*, no qual a maioria dos conceitos mais importantes são identificados e depois generalizados e especializados em outros conceitos.

2.6.5 METODOLOGIA DE GRUNINGER AND FOX

Grüninger e Fox (1995) propõem, uma metodologia inspirada no desenvolvimento de sistemas baseados no conhecimento, utilizando a lógica de primeira ordem. Os autores propõem identificar intuitivamente primeiro os principais cenários (possíveis aplicações em que a ontologia será utilizada). Em seguida, determinam um conjunto de questões de competência, usadas para identificar o escopo da ontologia. Estas questões de competência são utilizadas tanto para extrair os principais conceitos e suas propriedades, como as relações e os axiomas de uma ontologia. Estes componentes mencionados são formalmente expressos em lógica de primeira ordem. Portanto, este é um método muito formal, que tira proveito da lógica clássica robusta. Podendo ser usado como guia para transformar cenários informais em modelos computáveis.

2.6.6 KACTUS

No projeto *KACTUS*, segundo Bernaras (1996), o método proposto constitui uma ontologia construída através de uma base de conhecimento (*KB*), por meio de um processo de abstração (ou seja, seguindo uma estratégia *bottom-up*).

As demais aplicações são construídas, em geral, propondo iniciar uma construção de uma *KB* para uma aplicação específica, mais tarde quando uma nova base de conhecimento em um domínio semelhante é necessária, propõe-se a generalizar a primeira base de conhecimento desenvolvida e adaptá-la para ambas as aplicações.

2.6.7 SENSUS

Segundo Swartout (1997), *Sensus* é uma ontologia baseada em linguagem natural baseada em uma ontologia desenvolvida para fornecer uma ampla estrutura conceitual para trabalhar com

tradução automática. Ela inclui os termos de nível elevado (como “objeto inanimado”), bem como os termos específicos (como “submarino”).

Cada conceito da estrutura corresponde a um sentido da palavra. Assim, para a palavra “strut”, *SENSUS* contém três conceitos correspondentes aos diferentes significados da palavra “strut”. Atualmente, *SENSUS* contém mais de 50.000 conceitos.

Em contraste com outras ontologias de ampla cobertura, como a ontologia que faz parte do *CYC*, *SENSUS* foi desenvolvido pela extração e concentração de informações de recursos eletrônicos existentes, ao invés de construir do zero.

2.6.8 ONTOLOGY DEVELOPMENT 101

Para Noy e MacGuinness (2001) não há uma metodologia ideal de desenvolvimento de ontologias a ser seguida como modelo. Por isso, os autores sugerem um processo para auxiliar o desenvolvimento de ontologias. Esse processo foi denominado como *Ontology Development 101*. Ele consiste em um guia de 7 passos iterativos os quais podem ser executados livremente no desenvolvimento de ontologias conforme ilustra a Figura 2.

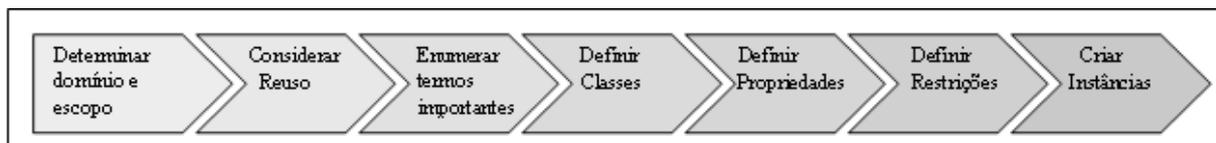


Figura 2 – 7 Passos no desenvolvimento de uma ontologia pelo guia *Ontology Development 101*

Cada passo tem objetivos específicos :

1. *Determinar o domínio e o escopo da ontologia*: neste passo deve-se determinar claramente o propósito e os cenários de utilização da ontologia a ser desenvolvida. Para isso, os autores identificaram algumas questões básicas que podem ser úteis a serem respondidas:

O que abrange o domínio da ontologia?

Para que se utilizará a ontologia?

Que questões a ontologia deveria responder?

Quem utilizará e manterá a ontologia?

2. *Considerar o reuso de ontologias existentes*: uma das propostas do desenvolvimento de ontologias é permitir o reuso das mesmas, então nesta fase procura-se encontrar ontologias relacionadas ao domínio e ao escopo da ontologia para uma possível reutilização;

3. *Enumerar termos importantes do domínio da ontologia*: é útil fazer uma lista de termos que são considerados importantes para o domínio sobre o qual será desenvolvida a ontologia, esses termos são importantes pois são utilizados nos próximos passos;
4. *Definir as classes do domínio e a hierarquia de classes*: a partir da lista de termos gerada no passo 3, extraem-se aqueles que descrevem objetos, os quais representam as classes. Após a definição das classes, deve-se organizá-las de forma hierárquica, essa organização pode ser feita através dos métodos *bottom-up*, *top-down*, ou uma junção dos mesmos, ficando ao critério do desenvolvedor escolher o melhor método para cada ocasião;
5. *Definir as propriedades das classes*: a partir das listas de termos remanescentes, define-se se esses termos correspondem a propriedades de dados ou de relações de classes;
6. *Definir as restrições das propriedades*: neste passo, deve-se definir quais restrições terão as propriedades;
7. *Criar as instâncias do domínio*: neste passo, criam-se as instâncias da ontologia, definindo valores para as propriedades de dados e relações;

2.7 VOCABULÁRIOS NA ÁREA DA SAÚDE

Atualmente existem diversos vocabulários disponíveis na área da saúde, tais vocabulários buscam padronizar e organizar periódicos científicos a fim de permitir a transferência de conhecimento, para isso, é necessário que o periódico esteja indexado pelo seu conteúdo em bases de dados referenciais e seguir as regras de normalização exigidas pelas mesmas. Alguns vocabulários são apresentados a seguir.

2.7.1 UMLS

A *Unified Medical Language System* (UMLS) foi projetada para facilitar o desenvolvimento de sistemas de computadores a compreender a linguagem biomédica e da área da saúde (UMLS, 2009).

A fonte UMLS utilizada pelos desenvolvedores tem o propósito de construir e melhorar os sistemas que criam, processam, recuperam e integram dados de saúde e informações biomédicas. Estas fontes de conhecimento são multifuncionais e são utilizadas em sistemas que executam diversas funções que envolvem tipos de informações, como prontuários, literatura científica, diretrizes e dados de saúde pública.

As ferramentas de *software* associadas (Rede semântica e *Metathesaurus*) ajudam os desenvolvedores a personalizar ou utilizar a base de conhecimento UMLS para fins particulares.

2.7.2 DECS

Os Descritores em Ciências da Saúde - DeCS constituem um vocabulário trilingue (Inglês, Português e Espanhol) e estruturado, desenvolvido pela BIREME (Centro Latino Americano e do Caribe de Informações em Ciências da Saúde) para indexar artigos de periódicos científicos, materiais técnicos com relatórios, anais de congressos e livros nas bases de dados *Lilacs* e *Medline* a fim de pesquisa e recuperação de assuntos da literatura científica.

Segundo a Biblioteca Virtual em Saúde, (BVS, 2009), o DeCS foi desenvolvido a partir do *Mesh – Medical Subject Headings* da *U.S. National Library of Medicine* (NLM) com o intuito de permitir em três línguas o uso de terminologias comuns para a pesquisa, proporcionando um meio consistente e único para a recuperação de informações independentemente do idioma.

Este vocabulário tem como principal finalidade unificar a linguagem para indexação e recuperação da informação entre os componentes da BIREME, permitindo uma pesquisa uniforme entre cerca de 750 instituições em 37 países na América Latina e no Caribe.

Estes descritores participam no projeto de desenvolvimento de terminologia única e rede semântica na saúde, UMLS com a responsabilidade de envio dos termos em português e espanhol.

Os conceitos que compõe o DeCS são organizados em um estrutura hierárquica onde permite a execução de termos mais amplos ou mais específicos ou todos os termos em uma pesquisa. A figura 3 – exibe a definição de taxa de mortalidade conforme o DeCS.

2 / 19 DeCS

Descritor Inglês: **Mortality Rate**

Descritor Espanhol: **Tasa de Mortalidad**

Descritor Português: **Coefficiente de Mortalidade**

Sinônimos Português: Razão de Mortalidade Proporcional
Taxa de Mortalidade

Categoria: [SP3.076.187.173.179](#)
[SP4.011.127.413.639.910](#)
[SP4.046.452.713.914](#)
[SP5.006.052.168.154.085](#)

Definição Português: É definido pelo número de óbitos em relação à [população](#), num período especificado.

Relacionados Português: [Mortalidade](#)

Qualificadores Permitidos Português: SN [estatística & dados numéricos](#) HI [história](#)
ST [normas](#)

Número do Registro: 16142

Ocorrência na BVS:

LILACS	1039
EQUIDAD	10
ADOLEC	170
BBO	2
BDENF	19
MedCarib	70
REPIDISCA	21
PAHO	264
DESASTRES	12

Similar: [DeCS](#) [CID-10](#) [SciELO](#) [LILACS](#) [LIS](#)

Figura 3 – Definição de taxa de mortalidade pelo DeCS

2.7.3 SNOMED

SNOMED (Nomenclatura Sistematizada de Termos Clínicos na Medicina) (SNOMED, 2009) é considerada a mais abrangente, terminologia multilíngue clínica de saúde do mundo. Ela visa melhorar o atendimento ao paciente, através do desenvolvimento de sistemas de precisão de cuidados de saúde.

Ela foi originalmente criada pelo *College of American Pathologists*, em setembro de 1993, onde inclui diagnósticos, sinais, sintomas e procedimentos; seu projeto permite a integração completa de todas as informações médicas, em um registro eletrônico dentro de uma estrutura única de dados.

2.7.4 HL7

HL7 – *Health Level Seven* (HL7, 2009), é uma organização sem fins lucrativos fundada em 1987, para desenvolver normas organizacionais dedicadas a fornecer um quadro global e normas relacionadas com o intercâmbio, integração, compartilhamento e recuperação de informações de

saúde eletrônicas que suportem a prática, a gestão, a execução e a avaliação dos serviços de saúde.

Ela define padrões de interoperabilidade que melhoram o atendimento, otimizando o fluxo de trabalho, reduzindo a ambiguidade e melhorando a transferência de conhecimentos entre todos os seus colaboradores como prestadores de cuidados de saúde, agências governamentais, a comunidade de fornecedores, e os pacientes.

2.7.5 CID 10

A CID10 – Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde (CID10, 2009), antigamente conhecida como 10º Revisão da Classificação Internacional de Doenças é um catálogo padronizado de doenças e problemas relacionados à saúde, tendo como referência a Nomenclatura Internacional de Doenças, estabelecida pela Organização Mundial de Saúde. Sua implementação para distribuição eletrônica foi efetuada pelo Departamento de Informática do SUS (Datusus, 2009), possibilitando assim, à todos os sistemas utilizados em território nacional que lidam com morbidade, seguirem um padrão de nomenclaturas internacional.

A CID10 possui 22 Capítulos, cada capítulo contém vários agrupamentos, sendo cada agrupamento um conjunto de categorias, e cada categoria subdividida em subcategorias, sendo essa organização exemplificada na Figura 4.

```
Capítulo I - Algumas doenças infecciosas e parasitárias
  A00-A09 - Doenças infecciosas intestinais
    A00 - Cólera
      A00.0 - Cólera devida a Vibrio cholerae 01, biótipo cholerae
      A01.1 - Cólera devida a Vibrio cholerae 01, biótipo El Tor
      A00.9 - Cólera não especificada
    A01 - Febres tifóide e paratifóide
      A01.0 - Febre tifóide
      A01.1 - Febre paratifóide A
      A01.2 - Febre paratifóide B
      A01.3 - Febre paratifóide C
    ...
  A15-A19 - Tuberculose
  ...
Capítulo IX - Doenças do aparelho circulatório
  I00-I02 - Febre reumática aguda
    I00 - Febre reumática sem menção de comprometimento do coração
      I00.0 - Febre reumática sem menção de comprometimento do coração
    I01 - Febre reumática com comprometimento do coração
      I01.0 - Pericardite reumática aguda
      I01.1 - Endocardite reumática aguda
      ...
      I01.9 - Doença cardíaca reumática aguda não especificada
    I02 - Coréia reumática
  ...
```

Figura 4 – Exemplo de Capítulos, Grupos, Categorias e Subcategorias da CID10.

Ainda segundo a Datusus, a CID10 possui as seguintes funcionalidades:

- Disponibiliza download e consulta dos arquivos contendo os códigos e descrições utilizadas na CID10 através da internet;
- Classifica morfologicamente neoplasias, exhibe listas especiais de tabulação para mortalidade e para morbidade, além de fornecer as definições e os regulamentos da nomenclatura, através da Lista Tabular;
- Apresenta Manual de Instruções e Índice Alfabético, em que notas sobre certificação médica, classificações em geral, tabulações e planejamento, facilitam seu uso.

2.8 FERRAMENTAS PARA DESENVOLVIMENTO DE ONTOLOGIAS

Construir uma ontologia é uma tarefa complexa, para auxiliar no seu desenvolvimento, utiliza-se ferramentas que facilitam o seu desenvolvimento. Atualmente existem diversas ferramentas para auxiliar no desenvolvimento de ontologias, neste trabalho foi considerado o uso da ferramenta *Protégé* (PROTÉGÉ, 2009) versão 3.4.1.

2.8.1 FERRAMENTA *PROTÉGÉ*

A ferramenta *Protégé* (PROTÉGÉ, 2009) é uma plataforma escrita em linguagem de programação JAVA, de distribuição livre e de código aberto, desenvolvida na escola de medicina da universidade de Stanford pelo grupo de pesquisa Stanford Medical Informatics. Inicialmente o objetivo do *Protégé* era a aquisição de conhecimento limitado a um sistema especialista para oncologia. Foi atualizado gradativamente para acompanhar a evolução da tecnologia de Sistemas Baseados em Conhecimento.

Na Figura 5 é apresentada a interface principal do *Protégé*.

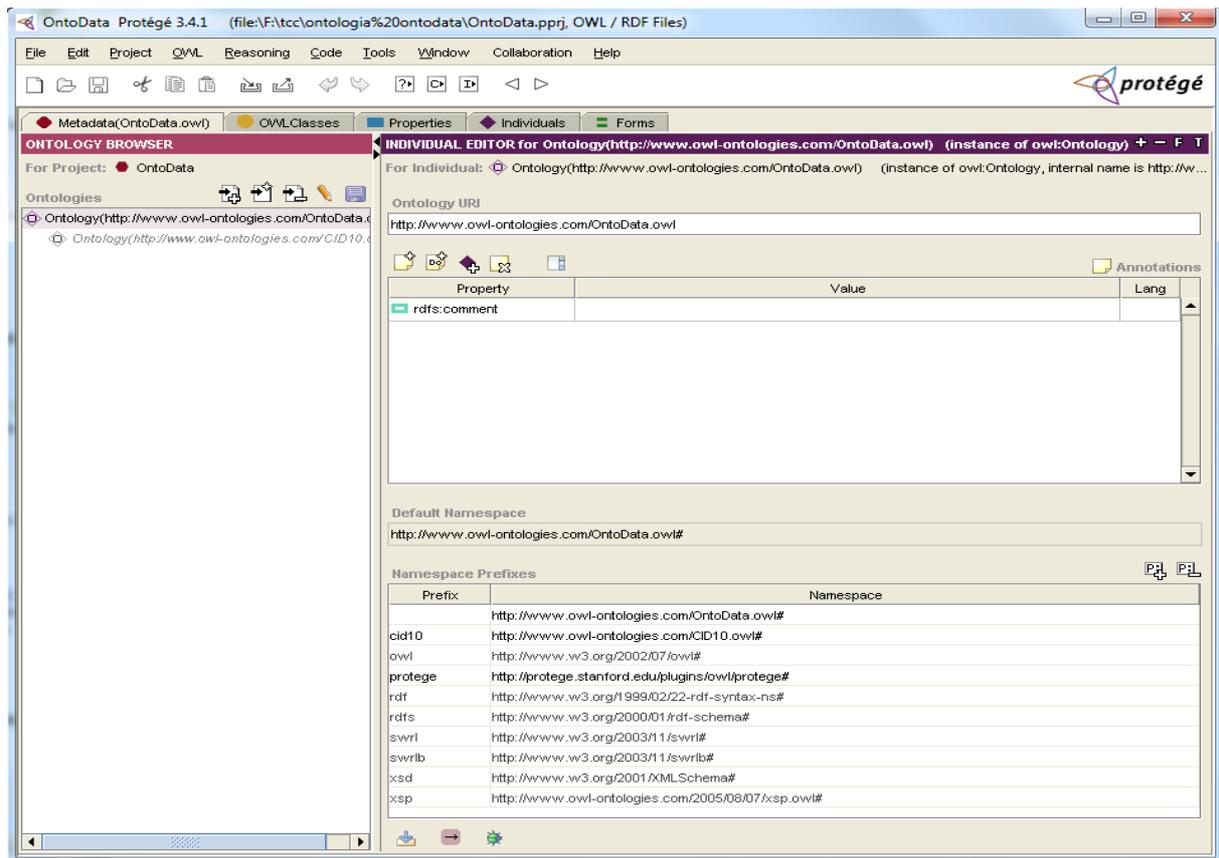


Figura 5 – Tela do Protégé

O conhecimento no *Protégé* é representado da seguinte forma:

1. *OWL/Classes*: é a definição de um conjunto de classes que representa um domínio de conceitos;
2. *Properties*: um conjunto de propriedades que associadas as classes formam as relações e funções de cada classe, uma propriedade pode ser de objeto (*object properties*), tipo de dados (*datatype properties*) ou de anotação (*annotation properties*). As propriedades de objeto também possuem propriedades, sendo elas:
 - Funcional: uma propriedade dita funcional pode apenas relacionar-se a um outro indivíduo a partir dela;
 - Funcional inversa: identifica a relação inversa de uma propriedade funcional;
 - Transitiva: identifica a relação entre indivíduos, sendo que se o indivíduo a se relaciona com o indivíduo b, o indivíduo b se relaciona com o indivíduo c, então o indivíduo a

também se relaciona com o indivíduo c;

- Simétrica: indica a simetria entre dois indivíduos.

3. *Individuals*: um conjunto de indivíduos, que formam as instâncias de cada classe, onde em cada indivíduo são informados os valores das propriedades.
4. *Forms*: esta aba permite indicar o tipo de dado a ser inserido em cada indivíduo, seja ele data, número ou *character*, ou definir se o indivíduo permite entrada de múltiplos valores (somente se a propriedade do indivíduo não for funcional) ou se permite que seja informado um único valor;

O *Protégé* ainda permite a inclusão de novas abas, essas abas são *plugins* desenvolvidos pelos mantenedores do *software* ou são *plugins* desenvolvidos por terceiros. Um exemplo é o *plugin* para a UMLS, já apresentada na seção 2.7.1 deste documento, ou do *plugin Instance Tree*, demonstrado na Figura 6. O *Instance Tree* exibe as classes e seus indivíduos em forma de árvore expansível, cada classe ou instância exibida pode ser expandida de forma a ser exibida todas as propriedades e instâncias ligadas a mesma, demonstrando todas as ligações vinculadas ao objeto que está sendo visualizado.

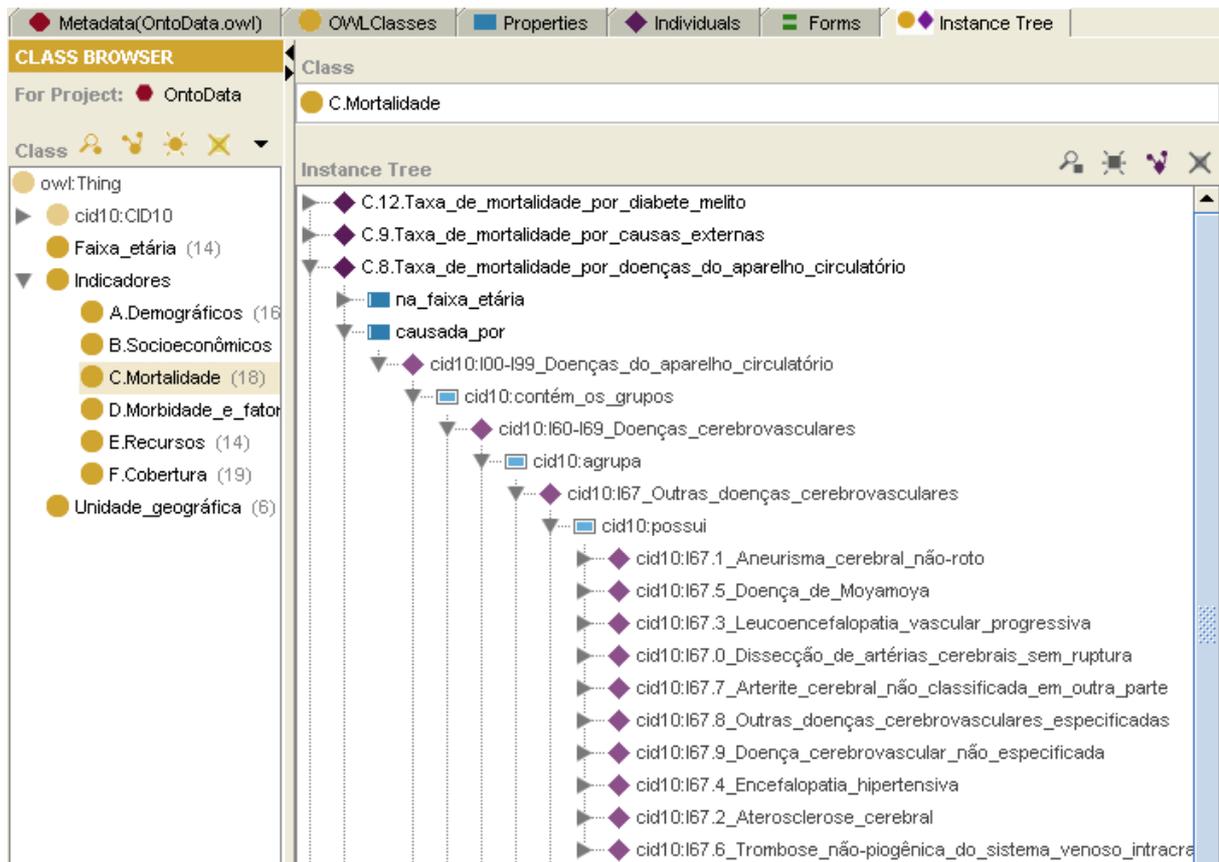


Figura 6– Plugin Instance Tree

2.9 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados os principais conceitos encontrados sobre ontologia, incluindo o conceito definido por Gruber, amplamente utilizado para definir ontologia por diversos autores. Foram detalhados os tipos de ontologia conforme a especificação de diversos autores, bem como foram elencados alguns critérios e metodologias para o desenvolvimento de ontologias.

Na área da Saúde foram apresentados alguns dos vocabulários disponíveis. Alguns destes vocabulários já podem ser encontrados no modelo de ontologia, outros ainda necessitam serem implementados.

Também foi apresentada uma versão do *software Protégé*, amplamente utilizada no desenvolvimento de ontologias.

3 DESENVOLVIMENTO

Neste capítulo é apresentado o desenvolvimento de uma ontologia para a área da Saúde utilizando o *software Protégé*. O domínio da ontologia desenvolvida neste capítulo é mortalidade, o qual é objeto de estudo nos Indicadores e Dados Básicos para Saúde no Brasil (IDB).

3.1 INDICADORES DE DADOS BÁSICOS PARA A SAÚDE NO BRASIL (IDB)

Os Indicadores de Dados Básicos para a Saúde no Brasil (IDB) compreendem um amplo conjunto de indicadores construídos a partir de bases de dados e pesquisas de âmbito nacional, cobrindo diversos aspectos da saúde no país, é um produto da Rede Interagencial de Informações para a Saúde (RIPSA, 2009).

A RIPSA disponibiliza o livro *Indicadores Básicos para a Saúde no Brasil: conceitos e aplicações – 2008 – 2ª Edição*. Este livro apresenta o conteúdo técnico completo das fichas de qualificação do IDB, trazendo conceitos utilizados, matriz de indicadores, fichas de qualificação e resumo das fontes de informação adotadas.

Para estudo do domínio mortalidade, foram utilizados o capítulo 2 – Matriz de indicadores e o capítulo 3 – Fichas de qualificação dos indicadores do livro. Ambos os capítulos tratam dos mesmos assuntos, sendo eles:

- *Indicadores Demográficos;*
- *Indicadores Socioeconômicos;*
- *Indicadores de Mortalidade;*
- *Indicadores de Morbidade e Fatores de Risco;*
- *Indicadores de Recursos;*
- *Indicadores de Cobertura;*

Apesar de ambos os capítulos tratarem dos mesmos indicadores, há diferenças entre o detalhamento das informações ali contidas, sendo assim, cada capítulo descreve os seguintes tópicos:

- *Capítulo 2 – Matriz de indicadores:*
 1. denominação;
 2. conceituação;
 3. método de cálculo;
 4. categorias de análise;
 5. fontes de dados.

- *Capítulo 3 – Ficha de qualificação dos indicadores:*
 1. conceituação;
 2. interpretação;
 3. usos;
 4. limitações;
 5. fontes;
 6. método de cálculo;
 7. categorias sugeridas para análise;
 8. dados estatísticos;
 9. comentários.

Ao analisar detalhadamente o material encontrado, percebe-se a utilização da CID10 na definição de conceitos relacionados a mortalidade. Visto que são muitas as referências a CID10, optou-se por também por implementar a CID10 na ontologia. A CID10 já foi apresentada na seção 2.7.5 deste documento. Para possibilitar uma utilização independente da CID10, optou-se por desenvolver duas ontologias, uma sobre a CID10 e outra sobre mortalidade.

3.2 METODOLOGIA UTILIZADA NO DESENVOLVIMENTO

Conforme um estudo nas metodologias apresentadas na seção 2.6 desse documento, foi selecionado o processo *Ontology Development 101* para desenvolver a ontologia, essa escolha foi

baseada no fato de o mesmo ser indicado na documentação do *site* do *Protégé*, e também por ser um processo simples e de fácil entendimento. Outra motivação veio de alguns projetos encontrados durante a pesquisa, como a ferramenta de apoio ao desenvolvimento de ontologias OntoKem². Esta ferramenta utiliza o processo *Ontology Development 101* como a principal metodologia de desenvolvimento analisada para a construção do *software*, sendo outras metodologias utilizadas para padronização dos documentos gerados através do mesmo.

3.3 ONTOLOGIA CID10

Segue os passos realizados na construção da ontologia, conforme determinado na seção 2.6 deste documento.

1. Determinando o Domínio e Escopo da Ontologia

O que abrange o domínio da ontologia?

O domínio da ontologia é a CID10.

Para que se utilizará a ontologia?

Inicialmente a ontologia CID10 será utilizada como classificador para os grupos de causa da ontologia Domínio da Saúde. Posteriormente, a ontologia poderá ser utilizada em demais ontologias ou para simples pesquisa sobre o assunto.

Que questões a ontologia deveria responder?

Neste momento, a ontologia classifica as doenças tal qual a CID10, posteriormente, complementando-a, será possível expandir a sua abrangência, respondendo assim a outras questões pertinentes a classificação de doenças.

Quem utilizará e manterá a ontologia?

A ontologia será utilizada por profissionais da área da Saúde, estudantes e pesquisadores com interesses relevantes ao assunto. Ela será mantida por pesquisadores do projeto OntoData e também por profissionais da área da Saúde.

2. Considerar o reuso das ontologias existentes

² <http://ontokem.egc.ufsc.br/>

Conforme o material pesquisado, não existem ontologias relacionadas ao domínio em questão, não possibilitando o aproveitamento de outras ontologias.

3. Enumerar termos importantes do domínio da ontologia

Analisando a CID10 foram encontrados os termos exibidos na Tabela 1.

Tabela 1 – Lista de termos encontrados na CID10

Termo	Descrição
CID10	Representa o domínio da ontologia
Capítulos	Conjunto de vários agrupamentos
Grupos de categorias	Conjuntos de várias categorias
Categorias	Correspondem ao código com uma letra e dois dígitos e que são subdivididos em Subcategorias
Subcategorias	Código completo da doença
agrupa	Indica as categorias do grupo de categorias
contém os grupos	Indica os grupos de categorias do capítulo
possui	Indica as subcategorias da categoria

4. Definir as classes do domínio e a hierarquia de classes

Com base na lista de termos disponível, conforme estabelecido no 4º passo do processo de desenvolvimento da ontologia, deve-se separar os termos que descrevem objetos, esses termos serão as classes da ontologia. Analisando a Tabela de termos, foram selecionados os termos expostos na Tabela 2 para representar as classes da ontologia CID10.

Tabela 2 – Classes da ontologia CID10

Classe	Hierarquia
CID10	Superclasse da classe Capítulos
Capítulos	Superclasse da classe Grupos_de_categorias
Grupos_de_categorias	Superclasse da classe Categorias
Categorias	Superclasse da classe Subcategorias
Subcategorias	Subclasse da classe Categorias

5. Definir as propriedades das classes

Com a lista remanescente de termos, estabeleceu-se qual tipo de propriedade o termo representa. As propriedades das classes da ontologia CID10 definidas podem ser observadas na

Tabela 3.

Tabela 3 – Propriedades da ontologia CID10

Propriedade	Tipo da propriedade
agrupa	Propriedade de ligação entre objetos
contém os grupos	Propriedade de ligação entre objetos
possui	Propriedade de ligação entre objetos

6. Definir as restrições das propriedades

As propriedades definidas são propriedades de objeto, elas definem a relação entre as classes. A Tabela 4 demonstra as restrições de cada propriedade.

Tabela 4 – Restrições das propriedades de objeto

Propriedade	Domínio	Limites
contém os grupos	Capítulos	Grupos_de_categorias
agrupa	Grupos_de_categorias	Categorias
possui	Categorias	Subcategorias

7. Criar as instâncias do domínio

As instâncias das classes da CID10 obedecem a hierarquia da própria CID10, sendo:

1. *Capítulos*: estão organizados de maneira que possuem um código inicial e final, o qual indica quais grupos de categorias pertencem ao capítulo;
2. *Grupos de categorias*: estão organizados através de um código inicial e final, o qual representa o capítulo, como também representa quais categorias estão incluídas em cada Grupo de categoria;
3. *Categorias*: estão organizadas através de um código contendo dois dígitos;
4. *Subcategorias*: estão organizadas através de um código contendo dois dígitos, o qual representa em qual categoria a Subcategoria pertence, junto com um dígito no intervalo de 0 à 9, representando assim o código completo da doença.

A CID10 possui 22 capítulos, 275 grupos de categorias, 2045 categorias e 12451 subcategorias, tornando o trabalho de construção da ontologia no Protégé algo muito demorado e cansativo. Para auxiliar na construção da ontologia foi desenvolvido um programa na linguagem ABL (Advanced Business Language) da Progress (Progress, 2009). A escolha desta linguagem foi definida pela familiarização do autor deste documento com a mesma.

O programa desenvolvido, possui a propriedade de ler os arquivos da CID10 disponíveis no site da DATASUS, e após gerar um arquivo no formato *OWL*. Tais arquivos estão estruturados com seus conteúdos separados por ponto-vírgula, sendo eles:

- *CID-10-CAPITULOS.CSV*: contém os capítulos da CID10;
- *CID-10-GRUPOS.CSV*: contém os grupos de categorias da CID10;
- *CID-10-CATEGORIAS.CSV*: contém as categorias da CID10;
- *CID-10-SUBCATEGORIAS.CSV*: contém as subcategorias da CID10.

O programa possui a seguinte lógica:

1. ***Definição de tabelas temporárias onde serão carregados os arquivos:*** para cada arquivo é definido uma tabela temporária contendo uma estrutura de campos similar à estrutura do arquivo que nela é importado. A Figura 7 demonstra a estrutura de cada tabela temporária.

```

define temp-table capitulos
  field num-cap as char
  field cat-inic as char
  field cat-fim as char
  field descricao as char
  field desc-abrev as char
  field desc-final as char
  index ch-primar num-cap.

define temp-table grupos
  field cat-inic as char
  field cat-fim as char
  field descricao as char
  field desc-abrev as char
  field desc-final as char
  index ch-primar cat-inic cat-fim.

define temp-table categorias
  field cat as char
  field classif as char
  field descricao as char
  field desc-abrev as char
  field refer as char
  field excluidos as char
  field desc-final as char
  index ch-primar cat.

define temp-table sub-categorias
  field subcat as char
  field classifi as char
  field restrisexo as char
  field causaobito as char
  field descricao as char
  field descabrev as char
  field refer as char
  field excluidos as char
  field desc-final as char
  field cat as char
  index ch-primar cat.

```

Figura 7 – Definição das tabelas temporárias onde é importado os arquivos da CID10

2. ***Importar cada arquivo em sua respectiva tabela temporária:*** cada arquivo é lido e o seu conteúdo é salvo em uma tabela temporária correspondente. A Figura 8 demonstra o processo de importar os arquivos.

```

input from D:\Downloads\CID10CSV\cid-10-capitulos.csv.
repeat:
    create capitulos.
    import delimiter ";" capitulos.
end.
input close.
input from D:\Downloads\CID10CSV\cid-10-grupos.csv.
repeat:
    create grupos.
    import delimiter ";" grupos.
end.
input close.
input from D:\Downloads\CID10CSV\cid-10-categorias.csv.
repeat:
    create categorias.
    import delimiter ";" categorias.
end.
input close.
input from D:\Downloads\CID10CSV\cid-10-subcategorias.csv.
repeat:
    create sub-categorias.
    import delimiter ";" sub-categorias.
end.
input close.

```

Figura 8 – Importando os arquivos nas tabelas temporárias

3. *Abrir o arquivo CID10.owl e criar cabeçalho, classes e propriedades:* o arquivo CID10.owl é criado no caminho especificado. Seguindo o padrão owl é criado um cabeçalho que identifica a owl, após é criado as classes e as propriedades já apresentadas anteriormente no início da seção 3.3 deste documento. As Figuras 9 e 10 exibem o código correspondente a este processo.

```

/*cria o arquivo CID10.owl em c:\emp\*/
output to c:\temp\CID10.owl.

/*imprime no arquivo o cabeçalho padrão*/
put unformatted '<?xml version="1.0"?>
<rdf:RDF
    xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
    xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
    xmlns:swrl="http://www.w3.org/2003/11/swrl#"
    xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns="http://www.owl-ontologies.com/CID10.owl#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
xml:base="http://www.owl-ontologies.com/CID10.owl">
<owl:Ontology rdf:about=""/>' skip.

```

Figura 9 – Criar o arquivo e cabeçalho da owl

```

/*imprime as classes*/
put unformatted '<owl:Class rdf:ID="capitulos"/>' skip.
put unformatted '<owl:Class rdf:ID="grupos_de_categorias"/>' skip.
put unformatted '<owl:Class rdf:ID="categorias"/>' skip.
put unformatted '<owl:Class rdf:ID="sub-categorias"/>' skip.

/*imprime as propriedades, ja vinculando-as à cada classe*/
put unformatted '<owl:ObjectProperty rdf:ID="contém_os_grupos">' skip.
put unformatted '<rdfs:domain rdf:resource="#capitulos"/>' skip.
put unformatted '<rdfs:range rdf:resource="#grupos_de_categorias"/>' skip.
put unformatted '</owl:ObjectProperty>' skip.

put unformatted '<owl:ObjectProperty rdf:ID="agrupa">' skip.
put unformatted '<rdfs:domain rdf:resource="#grupos_de_categorias"/>' skip.
put unformatted '<rdfs:range rdf:resource="#categorias"/>' skip.
put unformatted '</owl:ObjectProperty>' skip.

put unformatted '<owl:ObjectProperty rdf:ID="possui">' skip.
put unformatted '<rdfs:domain rdf:resource="#categorias"/>' skip.
put unformatted '<rdfs:range rdf:resource="#sub-categorias"/>' skip.
put unformatted '</owl:ObjectProperty>' skip.

```

Figura 10 – Inserir classes e propriedades de classes no arquivo

4. *Criar as instâncias de classes, e atribuir os valores correspondentes à cada instância:* para cada classe são criadas as suas instâncias e para cada instância são atribuídos os seus valores. As Figuras 11 e 12 mostram o código correspondente a criação das instâncias de classes e a estrutura gerada no arquivo respectivamente.

```

/*Lê cada capítulo*/
for each capitulos no-lock:
  put unformatted '<capitulos rdf:ID="' capitulos.desc-final '">' skip.
  /*Lê os grupos de cada capítulo*/
  for each grupos no-lock
    where grupos.cat-ini >= capitulos.cat-ini
      and grupos.cat-fim <= capitulos.cat-fim:
        put unformatted '<contém_os_grupos>' skip.
        put unformatted '<grupos_de_categorias rdf:ID="' grupos.desc-final '">' skip.
        /*Lê as categorias dos grupos*/
        for each categorias no-lock
          where categorias.cat >= grupos.cat-ini
            and categorias.cat <= grupos.cat-fim:
              put unformatted '<agrupa>' skip.
              put unformatted '<categorias rdf:ID="' categorias.desc-final '">' skip.
              /*Lê as subcategorias das categorias*/
              for each sub-categorias no-lock
                where substring(sub-categorias.subcat, 1,3) = categorias.cat:
                  put unformatted '<possui>' skip.
                  put unformatted '<sub-categorias rdf:ID="' sub-categorias.desc-final '">' skip.
                  put unformatted '</sub-categorias>' skip.
                  put unformatted '</possui>' skip.
                end.
              put unformatted '</categorias>' skip.
              put unformatted '</agrupa>' skip.
            end.
          put unformatted '</grupos_de_categorias>' skip.
          put unformatted '</contém_os_grupos>' skip.
        end.
      put unformatted '</capitulos>' skip.
    end.
  end.
end.

```

Figura 11 – Gerar cada instância de classe e valorar as referências das instâncias

```

<capítulos rdf:ID=instância_do_capitulo>
  <contém_os_grupos>
    <grupos_de_categorias rdf:ID=instância_do_grupo_de_categorias>
      <agrupa>
        <categorias rdf:ID=instância_da_categoria>
          <possui>
            <sub-categorias rdf:ID=instância_da_sub_categoria>
          </sub-categorias>
        </possui>
      </categorias>
    </agrupa>
  </grupos_de_categorias>
</contém_os_grupos>
</capítulos>

```

Figura 12 – Exemplo de estrutura gerada no arquivo CID10.owl

5. **Fechar o arquivo:** por final, é fechada a owl e fechado o arquivo gerado. Neste momento a ontologia está apta a ser utilizada no *Protégé*.

Após gerado o arquivo CID10.owl, o mesmo é editado no *Protégé*, devido as tabelas importadas possuírem cabeçalhos os mesmos foram gerados na ontologia, sendo necessário que fossem retirados através do *Protégé*.

O processo de desenvolvimento da ontologia CID10 encontra-se finalizado. Na Figura 13 é exibida a hierarquia de classes através do *plugin Jambalaia*, e na Figura 14 é exibido um exemplo da árvore de instâncias através do *plugin Instance Tree*.

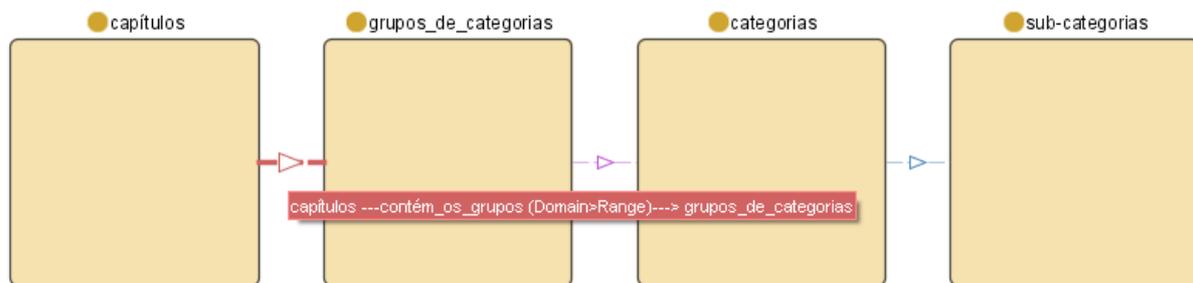


Figura 13 – Hierarquia de classes da ontologia CID10 visualizada pelo *plugin Jambalaia*

Nesta Figura, é exibida a ordem da hierarquia das classes através do relacionamento gerado pelas propriedades, sendo: **capítulos contém_os_grupos grupos_de_categorias agrupa categorias possui sub-categorias**.

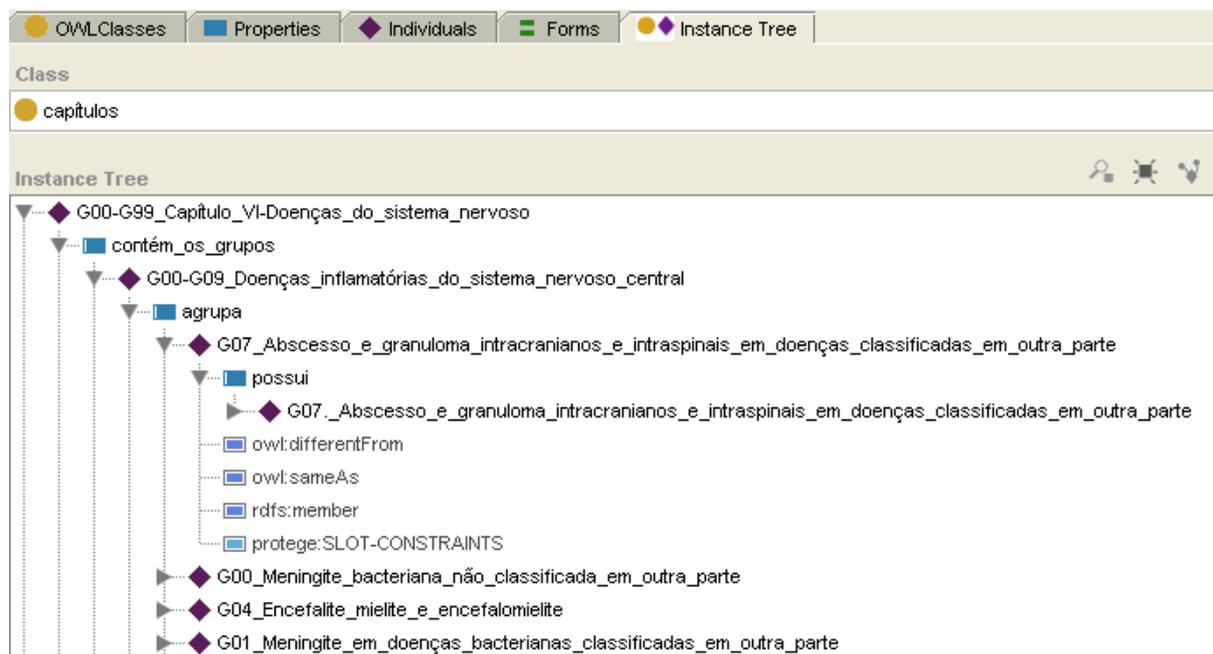


Figura 14 – Exemplo da árvore de instâncias da ontologia CID10 visualizada pelo plugin *Instance Tree*

3.4 ONTOLOGIA DOMÍNIO DA SAÚDE

Seguem os passos realizados na construção da ontologia de Domínio da Saúde, conforme determinado na seção 2.6 deste documento.

Determinando o Domínio e Escopo da Ontologia

Conforme explicado anteriormente, o domínio estabelecido para desenvolvimento neste trabalho é mortalidade, sendo este domínio referente ao IDB. Mesmo tendo esse domínio já definido, foram respondidas as questões definidas na seção 2.6

De acordo com o estabelecido para o primeiro passo, foram ser respondidas as seguintes perguntas:

O que abrange o domínio da ontologia?

O domínio da ontologia relaciona os conceitos referentes aos tipos de mortalidade relacionados no IDB.

Para que se utilizará a ontologia?

A utilização prevista da ontologia do Domínio da Saúde será:

- integração de informações no portal do projeto OTICSSS, em desenvolvimento por pesquisadores da UCS (OTICSS, 2009);
- para mineração em bases de dados na área da Saúde;
- estabelecimento de uma linguagem comum entre pesquisadores da área da Informática e da Saúde;
- compartilhamento de conhecimento entre profissionais da área saúde.

Que questões a ontologia deveria responder?

A ontologia deve responder questões tais como:

- a que doenças se refere uma classificação do cid10?
- Quais grupos de doenças estão contemplados em um determinado indicador de mortalidade da RIPSA?

Quem utilizará e manterá a ontologia?

A ontologia será utilizada por profissionais da área da Saúde, estudantes e pesquisadores com interesses relevantes ao assunto. Ela será mantida por pesquisadores do projeto OntoData e também por profissionais da área da Saúde.

Considerar o reuso das ontologias existentes

Conforme o material pesquisado, não existe ontologias relacionadas ao domínio em questão, porém, a ontologia CID10 também desenvolvida neste documento é aproveitada no desenvolvimento da ontologia de Domínio da Saúde. A Figura 15 demonstra a CID10 importada na nova ontologia.

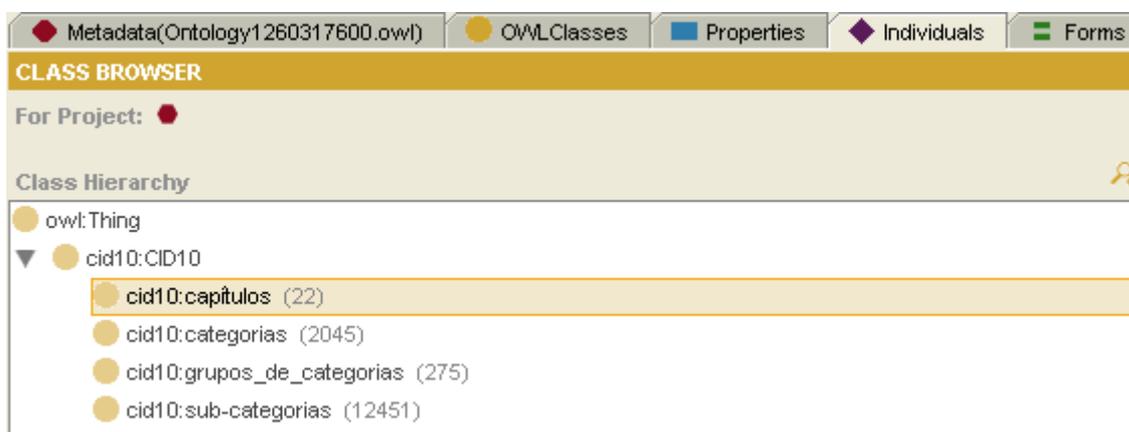


Figura 15 – CID10 importada na nova ontologia

Enumerar termos importantes do domínio da ontologia

Analisando a matriz de indicadores – mortalidade já apresentada na seção 3.1 deste documento, observou-se um padrão para determinar em qual indicador o óbito é classificado, sendo esse padrão composto pela causa do óbito, a idade e a região em que o óbito ocorreu. Isso pode ser visualizado na Figura 16.

DENOMINAÇÃO	CONCEITUAÇÃO	CATEGORIAS	FONTES
C.2. Taxa de mortalidade perinatal	Número de óbitos fetais (a partir de 22 semanas completas de gestação, ou 154 dias) acrescido dos óbitos neonatais precoces (0 a 6 dias) por mil nascimentos totais (óbitos fetais mais nascidos vivos), em determinado espaço geográfico, no ano considerado. Todos os valores referem-se à população residente.	<ul style="list-style-type: none"> Unidade geográfica: Brasil, grandes regiões, estados e Distrito Federal. 	<ul style="list-style-type: none"> MS/Cenepi: Sistema de Informações sobre Mortalidade (SIM) e Sistema de Informações sobre Nascidos Vivos (Sinasc).
C.3. Taxa de mortalidade materna	Número de óbitos femininos por causas maternas, por 100 mil nascidos vivos, na população residente em determinado espaço geográfico, no ano considerado.	<ul style="list-style-type: none"> Unidade geográfica: Brasil, grandes regiões, estados e Distrito Federal. 	<ul style="list-style-type: none"> MS/Cenepi: Sistema de Informações sobre Mortalidade (SIM) e Sistema de Informações sobre Nascidos Vivos (Sinasc).
C.4. Mortalidade proporcional por grupos de causas	Distribuição percentual de óbitos por grupos de causas definidas, na população residente em determinado espaço geográfico, no ano considerado. Grupos de causas (capítulos da CID): doenças infecciosas e parasitárias (códigos A00-B99); neoplasias (C00-D48); doenças do aparelho circulatório (I00-I99); doenças do aparelho respiratório (J00-J99); algumas afecções originadas no período perinatal (P00-P96); causas externas (V01-Y98); demais causas definidas (todos os demais capítulos, exceto o XVII e o XXI).	<ul style="list-style-type: none"> Unidade geográfica: Brasil, grandes regiões, estados, Distrito Federal, regiões metropolitanas e municípios das capitais. Sexo: masculino e feminino. Faixa etária: < 1 ano, 1-4, 5-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69 e 70 anos e mais. Para as afecções originadas no período perinatal, subdivisões da faixa etária de < 1 ano: 0-6, 7-27 e 28-364 dias de vida completos. 	<ul style="list-style-type: none"> MS/Cenepi: Sistema de Informações sobre Mortalidade (SIM).

Figura 16 – Matriz de indicadores – Mortalidade

Na Figura 15 são apresentadas as principais colunas avaliadas da Matriz de indicadores – mortalidade, nela pode ser observado, destacado em vermelho, o indicador de mortalidade, destacado em azul a faixa etária, destacado em amarelo a unidade geográfica e destacado em verde a causa do óbito. De forma a complementar a especificação da ontologia, também foram levantados termos a partir da Ficha de qualificação dos indicadores. A Tabela 5 mostra a lista de termos.

Tabela 5 – Lista de termos encontrados ao analisar a matriz de indicadores e a ficha de qualificação dos indicadores

Termo	Descrição
Indicadores	Representa a matriz de indicadores do IDB
Demográficos	Subconjunto temático dos indicadores
Socioeconômicos	Subconjunto temático dos indicadores
Mortalidade	Subconjunto temático dos indicadores

Termo	Descrição
Morbidade e fatores de risco	Subconjunto temático dos indicadores
Recursos	Subconjunto temático dos indicadores
Cobertura	Subconjunto temático dos indicadores
Unidade geográfica	Possíveis unidades geográficas onde ocorreu o óbito
CID10	Doenças da CID10
Faixa etária	Possíveis faixas etárias da mortalidade
causada por	Indica a causa da mortalidade
na faixa etária	Indica a faixa etária da mortalidade
na unidade geográfica	Indica a unidade geográfica em que a mortalidade ocorreu
Conceituação	Descrição do conceito do indicador
Interpretação	Descrição da interpretação do indicador
Limitações	Descrição das limitações impostas ao indicador
Usos	Descrição da finalidade de uso do indicador
Método de cálculo direto	Descrição do método de cálculo principal do indicador
Método de cálculo alternativo	Descrição do método alternativo de cálculo do indicador
Comentários	Comentários sobre o indicador

Apesar de o domínio da ontologia desenvolvida ser apenas mortalidade, todos os subconjuntos temáticos foram elencados, a representação dos mesmos enriquece a ontologia desenvolvida e amplia a possibilidade de projetos futuros.

Definir as classes do domínio e a hierarquia de classes

Com base na lista de termos disponível, conforme estabelecido no 4º passo do processo de desenvolvimento da ontologia, deve-se separar os termos que descrevem objetos, esses termos serão as classes da ontologia. Analisando a tabela de termos, foram selecionados os termos expostos na Tabela 6 para representar as classes.

Tabela 6 – Classes da ontologia Indicadores

Classe	Hierarquia
Indicadores	Superclasse das classes Demográficos, Socioeconômicos, Mortalidade, Morbidade e fatores de risco, Recursos e Cobertura
Demográficos	Subclasse da classe Indicadores
Socioeconômicos	Subclasse da classe Indicadores
Mortalidade	Subclasse da classe Indicadores
Morbidade e fatores de risco	Subclasse da classe Indicadores
Recursos	Subclasse da classe Indicadores
Cobertura	Subclasse da classe Indicadores
Unidade geográfica	Classe sem subclasses
Faixa etária	Classe sem subclasses
CID10	Doenças da CID10

De posse das classes definidas, iniciou-se o desenvolvimento da ontologia no *Protégé*, a Figura 17 demonstra o resultado da modelagem inicial.

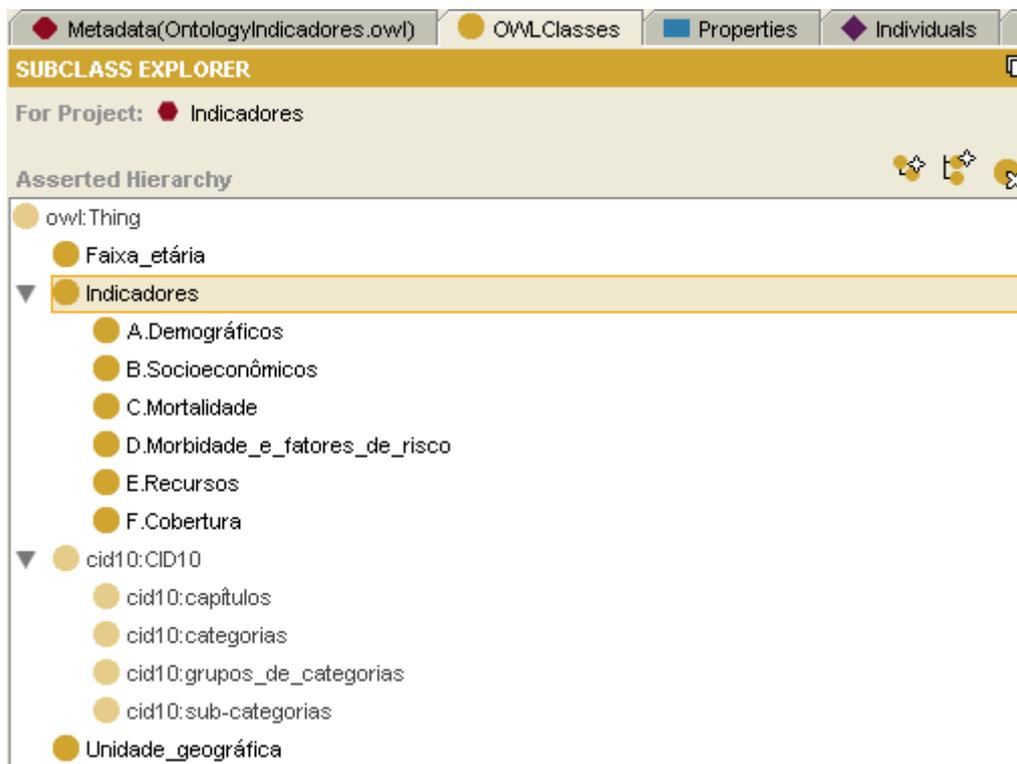


Figura 17 – Modelagem das Classes referentes aos Indicadores no *Protégé*

Definir as propriedades das classes

Com a lista remanescente de termos, estabeleceu-se qual tipo de propriedade o termo

representa. As propriedades definidas podem ser observadas na Tabela 7.

Tabela 7 – Propriedades das classes definidas

Propriedade	Tipo de propriedade
causada_por	Propriedade de ligação entre objetos
na_faixa_etária	Propriedade de ligação entre objetos
na_unidade_geográfica	Propriedade de ligação entre objetos
Conceituação	Propriedade de anotação
Interpretação	Propriedade de anotação
Limitações	Propriedade de anotação
Usos	Propriedade de anotação
Método de cálculo direto	Propriedade de anotação
Método de cálculo alternativo	Propriedade de anotação
Comentários	Propriedade de anotação

Após a definição das propriedades, as mesmas foram implementadas no *Protégé*, como é exibido na Figura 18. As propriedades de anotação estão destacadas em azul-escuro, as propriedades de objeto estão destacadas em azul claro. Junto também aparecem as propriedades herdadas da ontologia CID10.

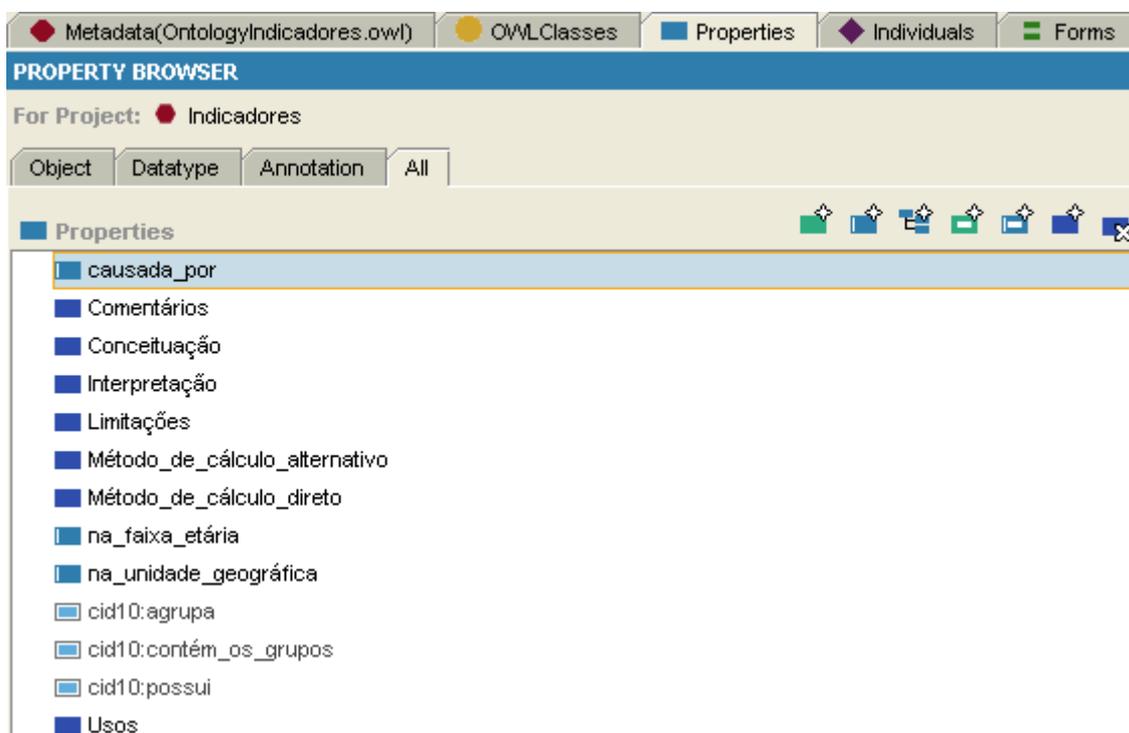


Figura 18 – Propriedades modeladas no *Protégé*

Definir as restrições das propriedades

As propriedades de anotação não possuem restrições, porém as propriedades de objeto necessitam de restrições, tais restrições podem definir o relacionamento entre classes, e ainda, podem definir limites de dados a serem informados nas propriedades. A Tabela 8 demonstra as propriedades com as restrições definidas.

Tabela 8 – Restrições das propriedades de objeto

Propriedade	Domínio	Limites
causada_por	Mortalidade	CID10
na_faixa_etária	Mortalidade	Faixa_etária
na_unidade_geográfica	Mortalidade	Unidade_geográfica

Observando a Tabela 8 é possível perceber a definição do limite da propriedade *causada_por* como sendo a CID10, isso ocorre devido a causa da mortalidade ser relacionada a itens presentes na CID10. Como a ontologia CID10 já foi importada para o projeto, as restrições relacionadas às instâncias da ontologia CID10 também podem ser estabelecidas. A Figura 19 exibe a restrição estabelecida com domínio o indicador Mortalidade e como limite a ontologia CID10.

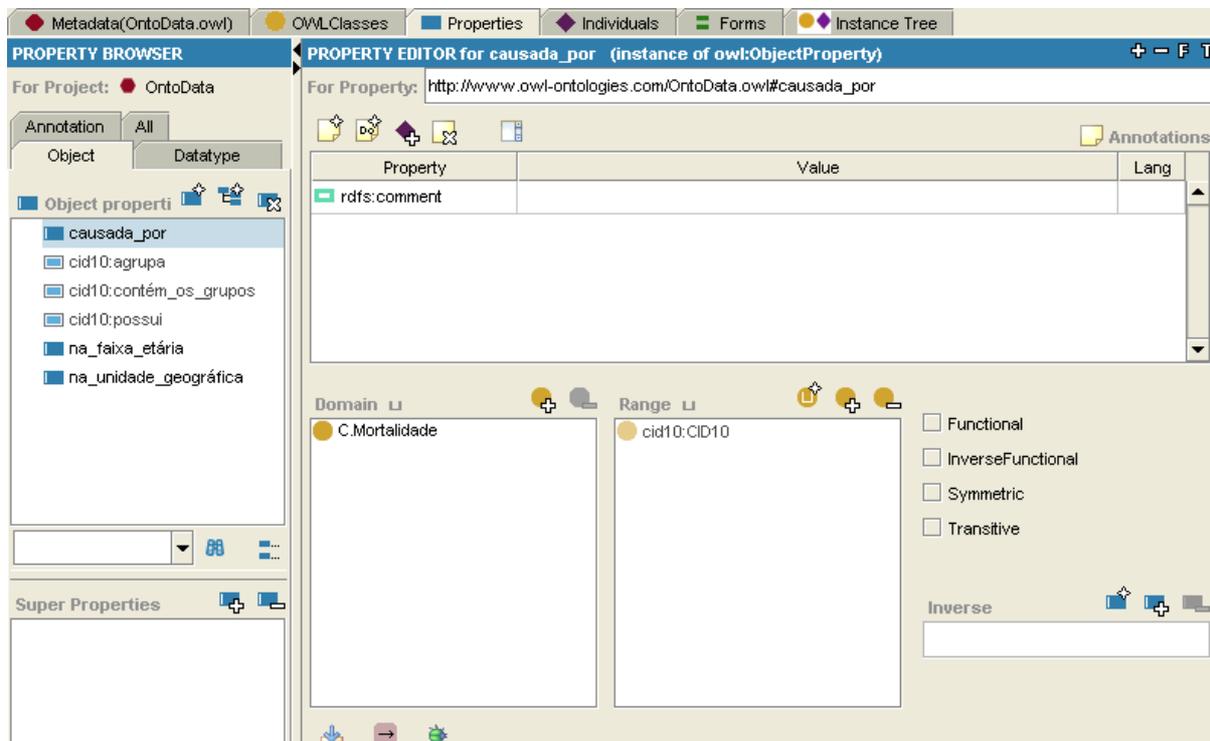


Figura 19 – Restrição da propriedade causada_por

Criar as instâncias do domínio

As instâncias de cada classe foram elencadas com base nos dados obtidos através da Matriz de indicadores apresentada na seção 3.1 deste documento. A coluna Denominação representa as instâncias das classes dos subconjuntos temáticos dos indicadores, as colunas Conceituação e Categorias servem para definir as instâncias das classes Faixa_etária e Unidade_geográfica. A Tabela 9 exhibe as principais classes e as suas instâncias.

Tabela 9 – Lista de classes e instâncias

Classe	Instância
Faixa_etária	
	a_partir_de_22_semanas_completas_de_gestação_ou_154_dias
	aos_28_e_mais_de_idade
	de_0-29_30-39_40-49_50-59_60-69_70-79_80_anos_e_mais_de_idade
	de_0-29_30-39_40-49_50-59_60-69_e_70_anos_e_mais_de_idade
	de_0-29_30-39_40-49_50-59_60_anos_e_mais_de_idade
	de_0-9_10-19-20-29_30-39_40-49_50-59_60_anos_e_mais_de_idade
	de_0_a_6_dias_de_vida
	de_10_a_49_anos
	de_28_a_364_dias_de_vida
	de_7_a_27_dias_de_vida
	menor_de_13_anos_13-14_15-19_20-29_30-39_40-49_50-59_60_anos_e_mais
	menor_de_15_anos_15-24_25-44_45-59_60_anos_e_mais_de_idade
	menor_de_um_ano_ou_de_1-4_5-9_10-19_20-29_30-39_40-49_50-59_60-69_e_70_anos_e_mais
	menores_de_5_anos_de_idade
C.Mortalidade	
	C.1.Taxa_de_mortalidade_infantil
	C.1.1.Taxa_de_mortalidade_neonatal_precoce
	C.1.2.Taxa_de_mortalidade_neonatal_tardia
	C.1.3.Taxa_de_mortalidade_pós_neonatal
	C.2.Taxa_de_mortalidade_perinatal
	C.3.Taxa_de_mortalidade_materna
	C.4.Mortalidade_proporcional_por_grupos_de_causas
	C.5.Mortalidade_proporcional_por_causas_mal_definidas
	C.6.Mortalidade_proporcional_por_doença_diarréica_aguda_em_menore

Classe	Instância
	s_de_cinco_anos_de_idade
	C.7.Mortalidade_proporcional_por_infecção_respiratória_aguda_em_menores_de_cinco_anos_de_idade
	C.8.Taxa_de_mortalidade_por_doenças_do_aparelho_circulatório
	C.9.Taxa_de_mortalidade_por_causas_externas
	C.10.Taxa_de_mortalidade_por_neoplasias_malignas
	C.11.Taxa_de_mortalidade_por_acidente_de_trabalho
	C.12.Taxa_de_mortalidade_por_diabete_melito
	C.13.Taxa_de_mortalidade_por_cirrose_hepática
	C.14.Taxa_de_mortalidade_por_aids
	C.15.Taxa_de_mortalidade_por_afecções_originadas_no_período_perinatal
Unidade_geográfica	
	Brasil
	Distrito_Federal
	estados
	grandes_regiões
	municípios_das_capitais
	regiões_metropolitanas

Com as instâncias de cada classe definidas, as mesmas foram modeladas no *Protégé*, neste momento, restou apenas a valoração das restrições definidas para a subclasse mortalidade. A Figura 20 exibe as instâncias da classe *Unidade_geográfica*, posteriormente, essas instâncias poderão ser inseridas como valores da propriedade *na_unidade_geográfica* de cada instância da subclasse *Mortalidade*.

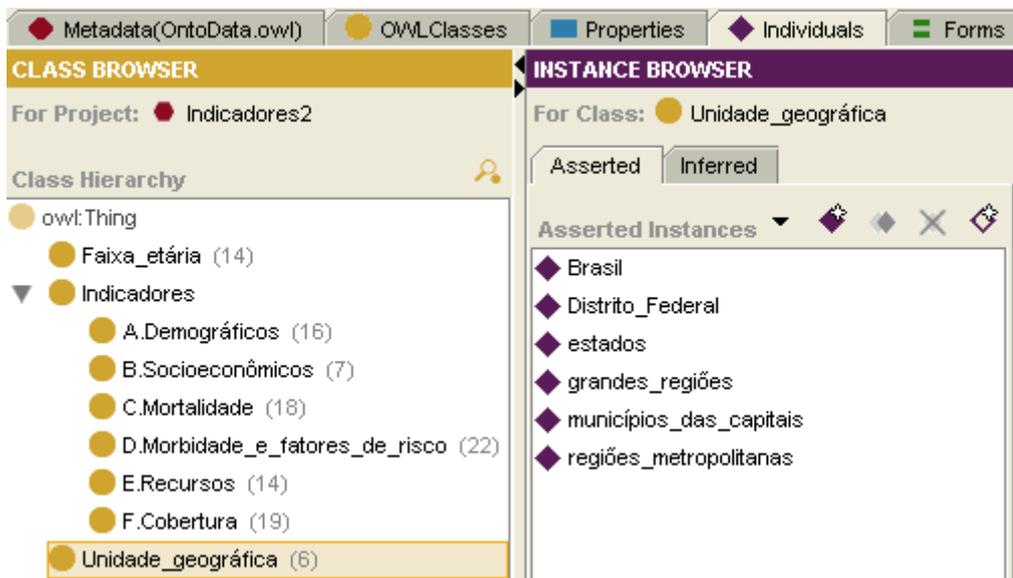


Figura 20 – Instâncias da classe Unidade_geográfica

As instâncias da subclasse mortalidade são consideradas indicadores de mortalidade. Conforme as propriedades definidas para a classe mortalidade, cada indicador pode ter uma ou mais propriedades de comentário ou de objeto valorada. A definição dos valores a serem atribuídos a estas propriedades será baseada no conteúdo da matriz de indicadores. A Figura 21 demonstra como exemplo as propriedades a serem valoradas no Indicador de Mortalidade – *C.12. Taxa_de_mortalidade_por_diabete_melito*.

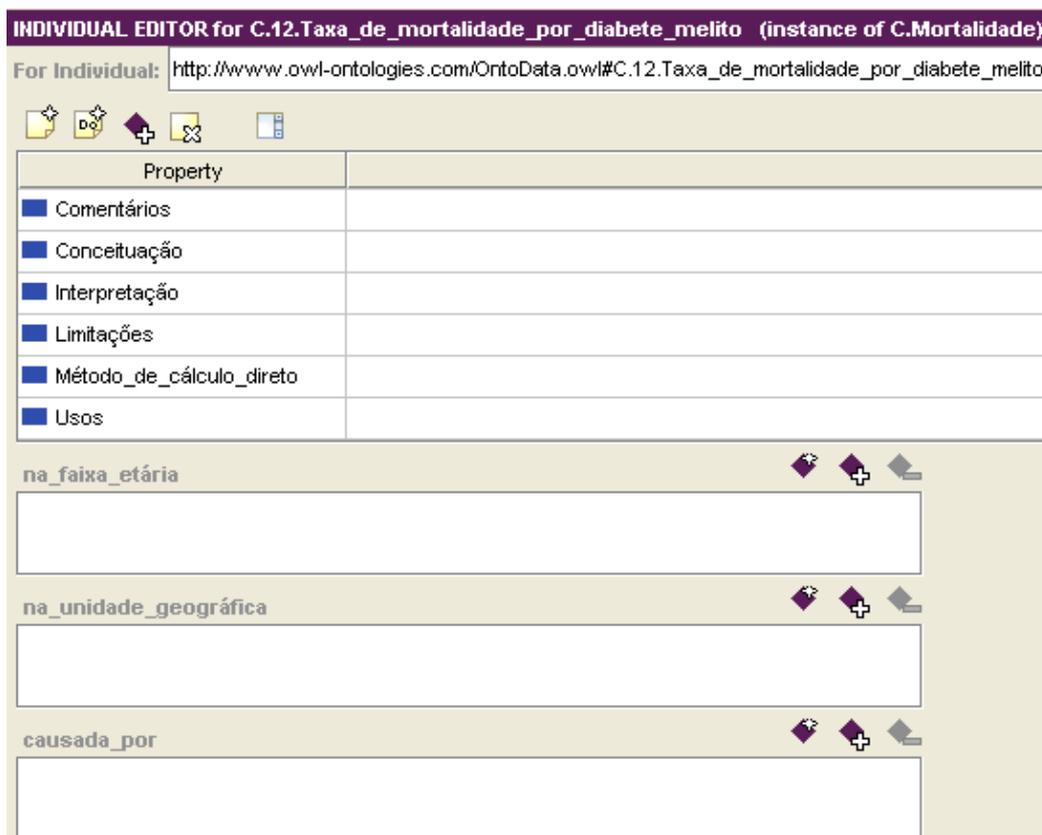


Figura 21 – Propriedades a serem valoradas para cada indicador de mortalidade

A Figura 22 apresenta alguns dos valores atribuídos aos comentários do Indicador de Mortalidade – *C.12. Taxa_de_mortalidade_por_diabete_melito* e a Figura 23 mostra os valores atribuídos as propriedades relacionadas ao mesmo indicador de mortalidade.

■ Conceituação	Número de óbitos por diabete melito (códigos E10 a E14 da CID-10), por 100 mil habitantes, na população residente em determinado espaço geográfico, no ano considerado.
■ Interpretação	Estima o risco de morte por diabete melito, em qualquer de suas formas clínicas. A elevação da taxa de mortalidade por diabetes acompanha o envelhecimento da população. No Brasil, mais de 85% dos óbitos por diabetes ocorrem a partir dos 40 anos de idade, em ambos os sexos. Variações da taxa de mortalidade específica podem também estar associadas à qualidade da assistência médica disponível, pois as complicações agudas do diabetes (códigos E-10.0 e E-10.1) são causas evitáveis de óbito. Em geral, as mortes por diabetes abaixo dos 40 anos de idade são consideradas evitáveis. A diabetes está associada à mortalidade por doenças do aparelho circulatório, em especial o acidente vascular cerebral, a doença hipertensiva e a doença isquêmica do coração.
■ Limitações	As estatísticas nacionais de mortalidade consideram apenas a causa básica do óbito, deixando de dimensionar a diabetes como importante causa associada. As bases de dados nacionais sobre mortalidade apresentam cobertura insatisfatória em muitos municípios do País, havendo expressiva subenumeração de óbitos nas regiões Norte e Nordeste. Imprecisões na declaração da "causa da morte" condicionam o aumento da proporção de causas mal definidas, comprometendo a qualidade do indicador.

Figura 22 – Valores atribuídos aos comentários do indicador de mortalidade Taxa de Mortalidade por Diabete Melito



Figura 23 – Valores atribuídos às propriedades do indicador de mortalidade Taxa de Mortalidade por Diabete Melito

Todas as classes, propriedades e instâncias definidas neste capítulo estão aplicadas às ontologias CID10 e Domínio da Saúde. Alguns indicadores não possuem definidos no material de consulta alguma das propriedades, como é o caso dos indicadores relacionados a mortalidade infantil, estes indicadores não possuem valores definidos para a propriedade *causada_por*.

A Figura 24 demonstra as ontologias CID10 e Domínio da Saúde visualizadas através do *plugin Jambalaia* do *Protégé*.

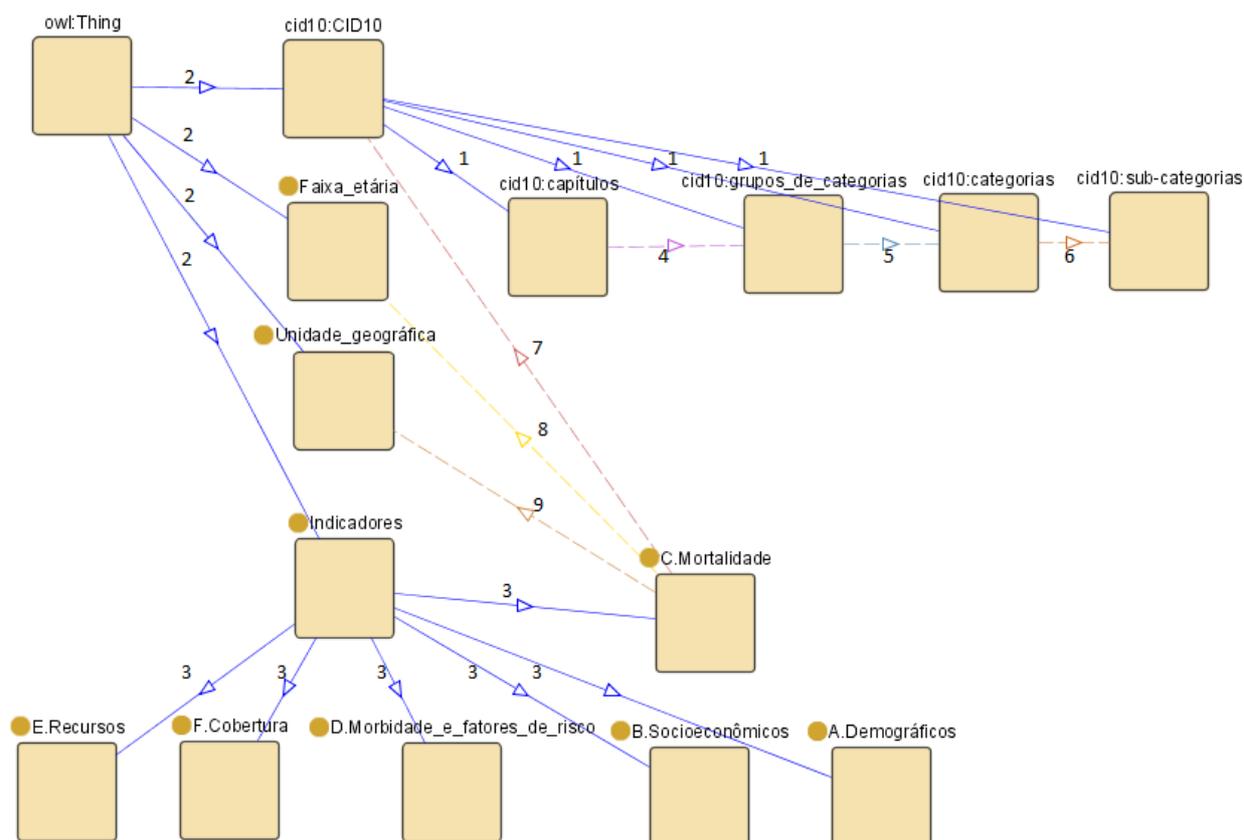


Figura 24 – Visualização da ontologia de Domínio da Saúde através do *plugin Jambalaia*

Nesta visualização é exibida a hierarquia de classes partindo da subclasse Mortalidade e a hierarquia de classes da ontologia CID10, sendo:

1. Relação entre a classe CID10 e as classes capítulos, grupos_de_categorias, categorias e sub-categorias;
2. Relação entre a classe *Thing* e as classes CID10, Faixa_etária, Unidade_geográfica e Indicadores;
3. Relação entre a classe Indicadores e as classes Demográficos, Socioeconômicos, Mortalidade, Morbidade_e_fatores_de_risco, Recursos e Cobertura.
4. Relação entre as classes capítulos e grupos_de_categorias definida pela propriedade contém_os_grupos;
5. Relação entre as classes grupos_de_categorias e categorias definida pela propriedade agrupa;
6. Relação entre as classes categorias e sub-categorias definida pela propriedade possui;
7. Relação entre as classes Mortalidade e CID10 definida pela propriedade causada_por;
8. Relação entre as classes Mortalidade e Faixa_etária definida pela propriedade na_faixa_etária;

9. Relação entre as classes Mortalidade e Unidade_geográfica definida pela propriedade na_unidade_geográfica.

3.5 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados os conceitos relacionados aos Indicadores Básicos de Saúde no Brasil (IDB), o processo selecionado como metodologia de desenvolvimento das ontologias, bem como foi apresentado detalhadamente o desenvolvimento da ontologia CID10, junto com o programa desenvolvido para auxiliar o desenvolvimento da mesma. Também foi apresentado o desenvolvimento da ontologia de Domínio da Saúde, o qual era a proposta deste trabalho. Ambas as ontologias foram demonstradas implementadas no *software Protégé*, inclusive com a utilização de alguns *plugins* disponíveis no mesmo.

É importante ressaltar que existem algumas vantagens detectadas pelo modo que foi desenvolvida a ontologia CID10 e posteriormente importada na ontologia de Domínio da Saúde, sendo elas:

- a ontologia CID10 poderá ser utilizada em projetos futuros sem a necessidade de também utilizar a ontologia de Domínio da Saúde;
- ter as ontologias separadas, possibilita trabalhar com arquivos menores, ao invés de trabalhar com um único arquivo contendo todas as informações;
- caso uma das ontologias não seja de domínio público, em caso de necessidade de compartilhamento de uma das ontologias, facilita preservar o sigilo dos dados contidos na outra ontologia.

4 CONCLUSÃO

O objetivo principal deste trabalho foi apresentar o *software Protégé*, alguns conceitos sobre ontologia e desenvolver uma ontologia para o Domínio da Saúde a ser utilizada no projeto OntoData.

4.1 SÍNTESE

Neste trabalho, inicialmente foram apresentados alguns dos problemas relacionados ao excesso de informação na área da Saúde. Também foi apresentado o objetivo do trabalho, que é pesquisar e desenvolver uma ontologia para a área da Saúde, bem como foi contextualizado o trabalho no projeto OntoData, fazendo a apresentação do mesmo.

Prosseguindo, no Capítulo 2 foi apresentado os conceitos relativos a ontologias, dentre eles um conceito geral definido por Gruber, que, apesar de existirem diversas definições sobre ontologias, dificilmente alguma a contradiz. Após foram detalhadas algumas definições de tipos de ontologias e seus autores. Para complementar o estudo conceitual sobre ontologias, também foram demonstrados alguns dos critérios, projetos, recursos e metodologias para desenvolvimento de ontologias. Como pesquisa relacionada à área da Saúde foram apresentados alguns dos vocabulários disponíveis. Por fim, neste capítulo, também foi apresentando o *software Protégé*, utilizado no desenvolvimento de ontologias.

No Capítulo 3, foi apresentado o processo *Ontology Development 101* como sendo a metodologia utilizada no desenvolvimento da ontologia proposta neste trabalho. Neste capítulo também foi apresentado o livro *Indicadores Básicos para a Saúde no Brasil: conceitos e aplicações – 2008 – 2º Edição*, o qual serviu como base para estudos sobre o domínio mortalidade, domínio este que faz parte da ontologia desenvolvida. Todo o trabalho despendido no desenvolvimento da ontologia de Domínio da Saúde foi documentado detalhadamente neste capítulo, tendo inclusive explicações sobre o motivo que levou o autor deste documento a desenvolver uma ontologia baseada na CID10, ontologia esta que posteriormente foi importada e reutilizada na ontologia de Domínio da Saúde. Para ilustrar todo o desenvolvimento, foi preciso demonstrar as principais telas de cada passo do desenvolvimento das ontologias, bem como foi preciso demonstrar alguns dos *plugins* disponíveis no *Protégé*.

4.2 CONTRIBUIÇÕES

Como contribuição, este trabalho fornece todo o material teórico sobre ontologias nele contido, e as duas ontologias nele desenvolvidas, a ontologia CID10 e a ontologia de Domínio da Saúde. Essas duas ontologias serão utilizadas no projeto OntoData, o qual já faz parte do objetivo principal deste trabalho. Além de servirem de contribuição no projeto OntoData, essas duas ontologias contribuem para a área de desenvolvimento de ontologias, pois o conhecimento nelas aplicado poderá ser reutilizado em outras ontologias relacionadas ao CID10 ou aos Indicadores de Dados Básicos de Saúde, bem como poderá servir como exemplo para o desenvolvimento de ontologias relacionadas a outras áreas.

4.3 TRABALHOS FUTUROS

Para possíveis trabalhos futuros ficam as seguintes propostas:

1. Ampliar a ontologia CID10, relacionando a causa de cada doença e possíveis tratamentos utilizados na tentativa de cura das mesmas;
2. Determinar quais localidades compõem cada Unidade Geográfica instanciada na ontologia de Domínio da Saúde;
3. Estabelecer os intervalos referentes a cada faixa de idade relacionada na ontologia de Domínio da Saúde, visto que da forma como foram modeladas, não está especificado o intervalo inicial e final de cada uma das faixas, como também existe faixas que sobrescrevem umas as outras.

5 REFERÊNCIAS

ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, Brasília, DF, v. 32, n. 3, p. 7-20, 2003.

BENJAMINS, V., R., CASANOVAS, P., BREUKER, J., GANGEMI, A. *Law and the Semantic Web*. Springer, 2005.

BERNARAS A., LARESGOITI I., CORERA J., Building and reusing ontologies for electrical network applications, in: Proc. European Conference on Artificial Intelligence (ECAI96), Budapest, Hungary, 1996, pp. 298–302.

BVS. Biblioteca Virtual em Saúde. Disponível em <<http://regional.bvsalud.org/php/index.php>>. Acesso em 05 de out. 2009.

BRILL, D. Loom reference manual. In: BACON, J. HAYTON; R. MOODY, K. *Middleware for digital libraries*. [S. l. : s. n.], 1998. Disponível em: <<http://www.dlib.org/dlib/october98/bacon/10bacon.html>>. Acesso em: 15 de set. 2009.

CID10 – Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde. Disponível em <<http://www.datasus.gov.br/cid10/v2008/cid10.htm>> . Acesso em 10 de ago. 2009.

CONNOLLY, D.; VAN HARMELEN, F.; HORROCKS, I. et al. DAM+OIL. Reference Description. W3C Note. Disponível em <<http://www.w3.org/TR/daml+oil-reference>>. Acesso em 15 de set. 2009.

CYC Cycorp Inc. Disponível em <<http://www.cyc.com>>. Acesso em 18 de out. 2009.

DATASUS Banco de Dados do Sistema Único de Saúde. Disponível em <<http://www.datasus.gov.br>>. Acessado em 20 de out. De 2009.

DECS Descritores em Ciências da Saúde. Atualizado em Abr. 2009. Disponível em <<http://decs.bvs.br/P/decswebp2009.htm>>. Acesso em 23 de nov. 2009.

FARIAS, R. F. et al. Ontologia para a gestão do conhecimento em saúde por meio da metodologia methontology. 2006. Disponível em: <<http://www.sbis.org.br/cbis/arquivos/857.pdf>>. Acesso em: 7 de ago. 2009.

FARQUHAR, A.; FIKES, R.; RICE, J. *The ontolingua server: USA: a tool for collaborative ontology construction*. Duluth : Academic Press, 1997. p. 707-727.

FENSEL, D., HARMELEN, F., KLEIN, M., AKKERMANS, H., *On-To-Knowledge: Ontology-based Tools for Knowledge Management*. Amsterdam, 2000.

FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A. Overview and analysis of methodologies for building ontologies. *The Knowledge Engineering Review*, Cambridge, v. 17, n. 2, p. 129-156, 2002.

FERNÁNDEZ, M., GÓMEZ-PÉREZ, A., JURISTINO, N. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. Symposium on Ontological Engineering of AAAI. Stanford (California). March 1997.

GÓMEZ-PÉREZ, A.; FERNÁNDEZ, M.; VICENTE, A. J. Towards a method to conceptualize domain ontologies. In: ECAI WORKSHOP ON ONTOLOGICAL ENGINEERING, 1996, Budapest. Proceedings... Disponível em: <<http://citeseer.ist.psu.edu/483876.html>>. Acesso em: 12 set. 2009

GRUBER, T. A translation approach to portable ontology specification. Knowledge Acquisition, Amsterdam, v.5, n. 2, p. 199-220, 1993.

GRÜNINGER M., FOX M.S., Methodology for the design and evaluation of ontologies, in: Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, 1995.

GUARINO, N. Formal ontology in information systems. In: First International Conference, 1., 1998, Trento, Itália. Anais... Trento IOS Press, 1998.

GUARINO, N.; GIARETTA, P. Ontologies and knowledge bases: towards a terminological clarification. In: Towards Very Large Knowledge Bases: knowledge building and knowledge sharing. Amsteram: IOS Press, 1996. p. 25-32.

HAAV, H. M.; LUBI, T. L. A survey of concept-based information retrieval tools on the web. In: PROCEEDINGS OF EAST-EUROPEAN CONFERENCE ADBIS. 5. 2001.

HEFLIN, J.; HENDLER, J. Searching the web with SHOE. In: ARTIFICIAL INTELLIGENCE FOR WEB SEARCH. 2000. Menlo Park : AAAI Press, CA, p. 35-40, 2000.

HPKB High Permormance Knowledge Bases. Disponível em <<http://reliant.teknowledge.com/HPKB/about/about.html>>. Acessado em 20 de out. de 2009.

HL7 Health Level Seven. Atualizado em 2009. Disponível em <<http://www.hl7.org/about/index.cfm>>. Acesso em 23 de nov. 2009.

JACOB, E. K. Ontologies and the Semantic Web. *Bulletin of the American Society for Information Science and Technology*, v. 29, n. 4, p. 19-22, April/May, 2003.

JASPER, R.; USCHOLD, M. A framework for understanding and classifying ontology applications. IJCAI-99, ONTOLOGY WORKSHOP, 1999, Stockholm, [S.l. : s. n.], 1999.

KARP, R.; CHAUDHRI, V.; THOMERE, J. XOL: an XML-based ontology exchange language. [S. l. : s. n.]. 1999. Disponível em: <<http://www.ai.sri.com/~pkarp/xol/xol.html>>. Acesso em: 20 de set. 2009.

KIFER, M.; LAUSEN, G.; WU, J. Logical foundations of object-oriented and frame-based languages. Journal of ACM, v. 42, p. 741-843, 1995.

LENAT, R.V. Guha, Building large knowledge-based systems. (Addison-Wesley Publishing Company, Inc. 1990).

MANICA, H.; DANTAS, M.A.R.; TEDESCO, J.L. Ontologia para compartilhamento e representação de conhecimento em saúde. *Diálogos & Saberes, Mandaguari*, v. 4, n. 1, p. 151-161, 2008.

MIZOGUCHI, R.; VANWELKENHUYSEN, J.; IKEDA, M. Task ontology for reuse of problem solving knowledge. In: *TOWARDS VERY LARGE KNOWLEDGE BASES*, 1994, Amsterdam. [S.I.] : IOS Press, 1995, p.46-59.

MOLOSSI, S. Inserção da biblioteca digital de teses e dissertações no contexto da web semântica: construção e uso de ontologias. 2008. 214 f. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal de Santa Catarina, Florianópolis, 2008.

NARDON, F., MOURA JR. L. Compartilhamento de conhecimento em Saúde utilizando ontologias e bancos de dados dedutivos. In: *International Information Technology Symposium, CBComp Best Papers Workshop*. Florianópolis, 2003.

NOY, N. F.; MCGUINNESS, D. L. Ontology development 101: a guide to creating your first ontology. Disponível em: <http://protege.stanford.edu/publications/ontology_development/ontology101.pdf>. Acesso em 23 de nov. 2009.

OTICSSS. Observatório das Tecnologias de Informação e Comunicação em Sistemas e Serviços de Saúde Disponível em: <<https://www.ucs.br/projetos/oticsss>>. Acesso em: 23 nov. 2009.

OWL. Ontology Web Language. Disponível em: <<http://www.w3.org/TR/owl-features/>> Acesso em: 20 de set. 2009.

PROGRESS 2009. Disponível em: <<http://www.progress.com/br/index.ssp>> Acesso em: 25 de jul. 2009.

PROTÉGÉ 2009. Disponível em: <<http://protege.stanford.edu>> Acesso em: 18 de jul. 2009.

RAUTENBERG, S., TODESCO, J.L., GAUTHIER, A.O; Processo de desenvolvimento de ontologias: uma proposta e uma ferramenta. *Revista. Tecnologia.*, Fortaleza, v.30, n.1, p.133-144, jun. 2009.

BRICKLEY, D.; MCBRIDE, B. RDF. Resource Description Framework. Disponível em: <www.w3c.org/rdf> Acesso em: 20 de set. 2009.

RIPSA. Observatório das Tecnologias de Informação e Comunicação em Sistemas e Serviços de Saúde. Disponível em: <<https://www.ripsa.org.br>>. Acesso em: 2 de out. 2009

RUSSEL, S. J.; NORVIG, P.: *Artificial Intelligence: A Modern Approach*, , Prentice Hall Series in Artificial Intelligence. Englewood Cliffs, New Jersey, 1995.

SALES, Rodrigo de; CAFE, Lígia. Diferenças entre tesauros e ontologias. *Perspect. ciênc. inf.*, Belo Horizonte, v. 14, n. 1, Apr. 2009 . Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362009000100008&lng=en&nrm=iso>. access on 20 Dec. 2009.

SNOMED. November 2009. Disponível em: <<http://www.ihtsdo.org/snomed-ct/>>. Acesso em : 23 de nov. 2009.

SWARTOUT B., RAMESH P., KNIGHT K., RUSS T., Toward Distributed Use of Large-Scale Ontologies, AAAI Symposium on Ontological Engineering, Stanford, 1997.

UMLS Documentation Knowledge Sources. September Release 2009AC. Disponível em: <http://www.nlm.nih.gov/research/umls/umlsdoc_intro.html#s1_5>. Acesso em: 23 nov. 2009.

USCHOLD, M.; GRUNINGER, M. Ontologies: principles, methods an applications. Knowledge Enginnering Review, v. 11, n. 2, 1996.

USCHOLD M., KING M., Towards a Methodology for Building Ontologies, in: IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, 1995.

VAN HEIJST, G.; SCHREIBER, A. T.; WIELINGA, B. J. Using explicit ontologies in KBS development. International Journal of Human-Computer Studies,, v. 46, n. 2-3, p. 183-192, fev./mar 1997.