



UNIVERSIDADE DE ÉVORA

MESTRADO DE ENGENHARIA INFORMÁTICA

Criação de uma ontologia e respectiva povoação
a partir do processamento de relatórios médicos

Luís Carlos Moreira Borrego

orientador: *Prof. Doutor Paulo Quaresma*

Dezembro de 2010

Prefácio

Este documento contém uma dissertação intitulada "Criação de uma ontologia e respectiva povoação a partir do processamento de relatórios médicos", um trabalho do aluno Luís Carlos Moreira Borrego¹, estudante de Mestrado em Engenharia Informática na Universidade de Évora.

O orientador deste trabalho é o Professor Doutor Paulo Quaresma², do Departamento de Informática da Universidade de Évora.

O autor do trabalho é licenciado em Engenharia Informática, pela Universidade de Évora. A presente dissertação foi entregue em Outubro de 2010.



¹luis_borrego@hotmail.com

²pq@di.uevora.pt

Agradecimentos

Após concluir com toda a dedicação esta longa etapa, quero deixar umas palavras de apreço e gratidão a todas as pessoas que contribuíram para o sucesso deste trabalho.

Um reconhecimento especial ao Professor Doutor Paulo Quaresma, meu orientador, pela sua disponibilidade na discussão do trabalho, e por todo o apoio e incentivo prestado.

Agradeço ao Hospital do Espírito Santo de Évora, em particular ao Director Clínico Doutor Manuel Carvalho e à equipa do Laboratório de Ultrassons Cardio e Neurovascular, pelo fornecimento dos relatórios médicos bem como toda a disponibilidade que demonstraram no esclarecimento de dúvidas.

Agradeço também aos Professores do Departamento de Informática da Universidade de Évora, que directa ou indirectamente contribuíram no desenvolvimento deste trabalho.

Um especial agradecimento à minha família que sempre me apoiou e incentivou a seguir em frente. Obrigado Pai, Mãe, Lena, Avós e Virgínia!

À Joana Correia agradeço o apoio incondicional e compreensão que sempre demonstrou. Obrigado por tudo!

Não podia deixar de agradecer a todos os amigos e colegas pelo apoio prestado ao longo desta caminhada académica. Um especial obrigado aos amigos Gaspar Brogueira e Bruno Antunes!

Parte deste trabalho foi realizado no âmbito do projecto "*MEDON - Ontologias para a modelação de dados e procedimentos médicos PTDC/EIA/80722/2006*", tendo recebido uma bolsa de investigação co-financiada pela FCT/FEDER através do programa POCI 2010.

Resumo

A evolução tecnológica tem provocado uma evolução na medicina, através de sistemas computacionais voltados para o armazenamento, captura e disponibilização de informações médicas. Os relatórios médicos são, na maior parte das vezes, guardados num texto livre não estruturado e escritos com vocabulário proprietário, podendo ocasionar falhas de interpretação.

Através das linguagens da Web Semântica, é possível utilizar ontologias como modo de estruturar e padronizar a informação dos relatórios médicos, adicionando-lhe anotações semânticas. A informação contida nos relatórios pode desta forma ser publicada na *Web*, permitindo às máquinas o processamento automático da informação. No entanto, o processo de criação de ontologias é bastante complexo, pois existe o problema de criar uma ontologia que não cubra todo o domínio pretendido.

Este trabalho incide na criação de uma ontologia e respectiva povoação, através de técnicas de PLN e Aprendizagem Automática que permitem extrair a informação dos relatórios médicos. Foi desenvolvida uma aplicação, que permite ao utilizador converter relatórios do formato digital para o formato OWL.

Palavras chave: Ontologias na Saúde, Extração de Informação, Processamento de Linguagem Natural, Aprendizagem Automática, Web Semântica

Creation of an ontology and its population from the processing of medical reports

ABSTRACT

Technological evolution has caused an medicine evolution through computer systems which allow storage, gathering and availability of medical information. Medical reports are, most of the times, stored in a non structured free text and written in a personal way so that misunderstandings may occur.

Through Semantic Web languages, it's possible to use ontology as a way to structure and standardize medical reports information by adding semantic notes. The information in those reports can, by these means, be displayed on the web, allowing machines automatic information processing. However, the process of creating ontology is very complex, as there is a risk creating of an ontology that not covering the whole desired domain.

This work is about creation of an ontology and its population through NLP and Machine Learning techniques to extract information from medical reports. An application was developed which allows the user to convert reports from digital format to OWL format.

Keywords: Health Ontology, Information Extraction, Natural Language Processing, Machine Learning, Semantic Web

Conteúdo

Prefácio	i
Agradecimentos	ii
Resumo	iii
Abstract	iv
Lista de Figuras	vii
Lista de Tabelas	ix
Lista de Abreviaturas	x
1 Introdução	1
1.1 Motivação e Enquadramento	1
1.2 Objectivos	2
1.3 Contribuições desta Dissertação	3
1.4 Estrutura da Dissertação	3
2 Fundamentação Teórica	5
2.1 Web Semântica	5
2.1.1 Objectivo	6
2.1.2 Mudanças da Web Actual para a Web Semântica	7
2.1.3 Arquitectura da Web Semântica	8
2.2 Ontologia	10
2.2.1 Componentes de uma Ontologia	10
2.2.2 Tipos de Ontologia	11
2.2.3 Metodologias para Construção de Ontologias	11
2.2.4 Linguagens de Representação de Ontologias	13
2.2.5 Ferramentas	19
2.2.6 Jena	20
2.2.7 Ontologias Médicas	21
2.3 Mineração de Textos	22
2.3.1 Processamento de Linguagem Natural	23
2.3.2 Aprendizagem Automática	25
3 Trabalho Relacionado	27
3.1 Ontologias na Saúde	27
3.2 Criação de Ontologias em Relatórios Médicos	28

4	Proposta de Arquitectura	32
4.1	Apresentação	32
4.1.1	Objectivos	33
4.1.2	Metodologia	33
4.2	Arquitectura	34
4.2.1	Plataforma	35
5	Análise dos documentos	36
5.1	Documentos em análise	36
5.1.1	Descrição dos documentos	36
5.2	Enumeração dos termos	37
5.2.1	Termos específicos dos documentos	37
5.2.2	Termos médicos	39
6	Criação da Ontologia	42
6.1	Ontologia e Web Semântica	42
6.1.1	Escolha da Linguagem	42
6.2	Criação das Classes	43
6.2.1	Classes dos Termos Específicos	43
6.2.2	Classes dos Termos Médicos	49
6.3	Criação dos Relacionamentos e das Restrições entre Classes	56
7	Extracção da Informação e Povoamento da Ontologia	62
7.1	O Sistema de Extracção de Informação	62
7.2	Preenchimento dos Campos Vazios	62
7.3	Criação e Preenchimento da Base de Dados	64
7.4	Processamento da Informação e Criação das Instâncias	66
8	Utilização do Sistema	81
8.1	Interface	81
8.2	Testes	82
9	Conclusão	90
9.1	Trabalho Futuro	91
A	Código da Ontologia	92
B	Relatório no Formato Digital	117
C	Árvore de Decisão	118
	Anexo A - Relatório Neurovascular	119
	Bibliografia	120

Lista de Figuras

2.1	Estruturação dos recursos e links na Web actual e na Web Semântica [41]	7
2.2	Utilização de Agentes Computacionais na Web Semântica [2]	8
2.3	Arquitectura da Web Semântica [4]	9
2.4	Tipos de Ontologias	11
2.5	Passos da Metodologia para a Criação da Ontologia	12
2.6	Camadas das Linguagens de Ontologias	13
2.7	Interface da Ferramenta Protégé	20
2.8	Etapas da Mineração de Textos [53]	23
2.9	Metodologia de Identificação de Tokens [34]	25
2.10	Exemplo de Árvore de Decisão	26
3.1	Exemplo de um Relatório Radiológico [1]	29
3.2	Ontologia de Relatórios Radiológicos	30
4.1	Fases de Desenvolvimento do Trabalho	33
4.2	Arquitectura	34
5.1	Identificação do Paciente nos relatórios	37
6.1	Hierarquia das classes para os termos do cabeçalho dos relatórios . . .	44
6.2	Hierarquia das classes para os termos da identificação do paciente . .	45
6.3	Hierarquia das classes para os termos do cabeçalho dos exames	46
6.4	Hierarquia das classes para os termos do rodapé dos relatórios	48
6.5	Hierarquia das classes para os termos das Entidades Anatómicas . . .	49
6.6	Hierarquia das classes para os termos relacionados com as Entidades Anatómicas	51
6.7	Hierarquia das classes para os termos das Observações	53
6.8	Hierarquia das classes para os termos dos Diagnósticos	54
6.9	Relações com a classe que possui o objecto <i>Artéria</i>	56
6.10	Relações entre classes que representam o esquema global do domínio .	58
7.1	Esquema do Módulo do Sistema para o Preenchimento dos Campos Vazios	64
7.2	Esquema do Módulo do Sistema para o Preenchimento da Base de Dados	66
7.3	Estrutura de classes	70
7.4	Vector com as Instâncias do Parágrafo	71
7.5	Algoritmo que relaciona as instâncias Lado e Entidades Anatómicas .	72
7.6	Algoritmo que relaciona as instâncias Zona e Entidades Anatómicas .	73

7.7	Algoritmo que relaciona as instâncias Stent e Entidades Anatômicas .	74
7.8	Algoritmo que relaciona as instâncias Entidades Anatômicas e Diagnóstico	75
7.9	Algoritmo que relaciona as instâncias Entidades Anatômicas e Observações	75
7.10	Algoritmo que relaciona as instâncias Observação e Diagnóstico	78
7.11	Esquema do Módulo do Sistema para o Processamento da Informação e Criação de Instâncias	80
8.1	Exemplo da Utilização da Aplicação	81
8.2	Mensagens Exibidas pela Interface	82
8.3	Instância do Relatório	83
8.4	Instâncias do Paciente	83
8.5	Instância do Médico	84
8.6	Instâncias do Primeiro Exame	84
8.7	Instâncias do Primeiro Grupo do Conteúdo do Primeiro Exame	85
8.8	Instâncias do Segundo Grupo do Conteúdo do Primeiro Exame	86
8.9	Instância do Técnico	86
8.10	Instâncias do Segundo Exame	87
8.11	Instâncias do Conteúdo do Segundo Exame	88
8.12	Instância da Conclusão	89

Lista de Tabelas

5.1	Termos do cabeçalho dos relatórios	38
5.2	Termos da Identificação do Paciente	38
5.3	Termos do cabeçalho dos exames	38
5.4	Termos do rodapé dos relatórios	39
5.5	Termos das Entidades Anatômicas	39
5.6	Termos relacionados com as Entidades Anatômicas	40
5.7	Termos das Observações	40
5.8	Termos dos Diagnósticos	41
6.1	Descrição de alguma sintaxe do OWL	43
6.2	Especificação das Relações e Restrições da classe com o objecto <i>Artéria</i>	57
6.3	Especificação das Relações e Restrições do esquema global do domínio	59

Lista de Abreviaturas

CN	<i>Concept Name</i>
CUI	<i>Concept Unique Identifier</i>
DAML	<i>DARPA Agent Markup Language</i>
DAML+OIL	<i>DARPA Agent Markup Language+Ontology Inference Layer</i>
HTML	<i>HyperText Markup Language</i>
EI	Extracção de Informação
IA	Inteligência Artificial
MeSH	<i>Medical Subject Headings</i>
MT	Mineração de Textos
OCR	<i>Optical Character Recognition</i>
OIL	<i>Ontology Inference Layer</i>
OWL	<i>Ontology Web Language</i>
PLN	Processamento de Linguagem Natural
RDF	<i>Resource Description Framework</i>
RDFS	<i>Resource Description Framework Schema</i>
SGML	<i>Standard Generalized Markup Language</i>
SHOE	<i>Simple HTML Ontology Extensions</i>
SNOMED	<i>Systematized Nomenclature of Medicine</i>
SQL	<i>Structured Query Language</i>
UMLS	<i>Unified Medical Language System</i>
UMLSKS	<i>UMLS Knowledge Sources</i>
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
XML	<i>Extensible Markup Language</i>
W3C	<i>World Wide Web Consortium</i>

Capítulo 1

Introdução

Neste capítulo é feita a introdução sobre a área que abrange o trabalho exposto nesta dissertação. Na secção 1.1 é descrita a motivação e o enquadramento desta dissertação. Nas outras secções são descritos os objectivos (1.2), as principais contribuições (1.3) no desenvolvimento da dissertação, e a sua estrutura (1.4).

1.1 Motivação e Enquadramento

A maior parte das informações disponíveis na *Web* actual, estão representadas num formato textual e escritos em língua natural, destinados à análise, compreensão e interpretação do ser humano. Deste modo, o processamento automático da informação é um processo bastante difícil, pois tais documentos não possuem a informação de modo estruturado e com anotação semântica que possibilitem o processamento às máquinas. Este problema é bem visível quando, por exemplo, realizamos uma pesquisa num qualquer motor de pesquisa e o que ele nos devolve são dezenas ou centenas de documentos e páginas, muitas vezes sem possuírem a informação que desejamos obter.

Com o intuito de ultrapassar esta limitação, está a ser planeada uma nova *Web* que será uma evolução da *Web* actual, que se dá pelo nome de Web Semântica. Esta nova *Web*, aposta na estruturação dos recursos de informação acrescentando anotação semântica aos dados do seu conteúdo. Assim, todos os recursos da *Web* são passíveis de ser processados automaticamente por máquinas que passam a ter acesso à semântica dos dados.

Uma das formas de conseguir este objectivo, é através do uso de ontologias, que permitem estruturar e modelar conceitos e relações de um dado domínio, fornecendo uma base de conhecimento orientada para as máquinas e respectivo processamento automático. Para facilitar o preenchimento da ontologia a partir de um conjunto de textos de determinado domínio, podem ser empregues processos automáticos ou semi-automáticos. A área da Mineração de Textos pode ser aqui utilizada.

A Mineração de Textos veio ajudar a lidar com a sobrecarga de informação causada pelo crescente volume de informação na *Web*, pois era uma tarefa bastante complicada quando se pretendia analisar, localizar e aceder a essa informação de

modo manual.

A Mineração de Textos permite extrair informação de textos, através de técnicas, métodos e algoritmos de algumas áreas, tais como, processamento de linguagem natural e aprendizagem automática.

A crescente evolução que tem ocorrido na área da informática, provoca consequentemente uma evolução na área da medicina, nomeadamente através da criação de sistemas que permitem um melhor tratamento da informação médica. Por outro lado, cada vez mais nos deparamos com a descoberta de novas patologias, com o aparecimento de novos medicamentos, etc, que tem provocado uma grande expansão da informação da área médica.

A integração da informação da área médica é um dos grandes desafios da informática na saúde. Entre outros, os exames e consequentemente os relatórios médicos são muitas vezes efectuados em diversas instituições, que utilizam vocabulários proprietários, e se esta informação não for devidamente preparada e apresentada de maneira apropriada, pode ocasionar falhas de interpretação.

Para solucionar este problema pode-se recorrer ao uso de ontologias, que irão permitir estruturar e adicionar anotação semântica à informação dos relatórios médicos. Desta forma, a informação contida nos relatórios pode ser publicada nesta nova *Web*, permitindo que as diversas instituições utilizem sistemas que processem automaticamente essa informação, trazendo consequentemente uma melhor manutenção do cuidado de saúde do paciente. Essa informação pode ainda ser usada para fins de investigação ou para criar sistemas de auxílio aos médicos no apoio à decisão diagnóstica.

A criação de ontologias não é uma tarefa trivial. Quando se pretende construir uma nova ontologia surgem dificuldades pois não existe uma solução universal. Além disso, existe o problema de se criar uma ontologia que não consiga cobrir todo o domínio pretendido, tornando a sua utilização pouco viável.

1.2 Objectivos

O propósito geral desta dissertação é tratar a problemática da criação e preenchimento de ontologias, em particular, de relatórios médicos, utilizando técnicas de processamento de linguagem natural e de aprendizagem automática.

Foram analisados relatórios médicos neurovasculares, provenientes do laboratório LUSCAN do Hospital do Espírito Santo de Évora, com o propósito de extrair todos os termos que irão permitir formar o domínio da ontologia.

O objectivo principal deste trabalho é criar uma ontologia que permita representar a informação expressa nos relatórios neurovasculares, e criar mecanismos automáticos que extraiam a informação e efectuem a referida "povoação" da ontologia. Deste modo, iremos obter uma base de conhecimento para este domínio, tornando possível uma nova utilização dos dados textuais através do processamento automático por máquinas.

1.3 Contribuições desta Dissertação

As principais contribuições apresentadas nesta dissertação são:

- Criação de uma ontologia para o processamento de relatórios médicos neurovasculares, e em particular:
 - Enumeração dos termos;
 - Definição de classes, propriedades e restrições sobre as classes.
- Aplicação das Árvores de Decisão para extracção de informação dos relatórios médicos;
- Desenvolvimento de um conjunto de algoritmos que permitem atribuir as relações entre os termos médicos presentes em cada frase dos relatórios (observações, diagnósticos, entidades anatómicas, etc);
- Desenvolvimento de uma aplicação, que permite aos utilizadores converter automaticamente um relatório médico neurovascular do formato digital para o formato OWL.

1.4 Estrutura da Dissertação

A dissertação está organizada do seguinte modo:

Capítulo 1

No presente capítulo é feita a motivação e o enquadramento do problema a solucionar, tais como os seus objectivos. É apresentada a estrutura da dissertação e as suas principais contribuições.

Capítulo 2

Neste capítulo é efectuada uma revisão da literatura das áreas referentes ao problema em análise. Em concreto é descrita a Web Semântica, as ontologias como forma de estruturar a informação e publicá-la na *Web*, e a área da Mineração de Textos como forma de extrair a informação.

Capítulo 3

O trabalho realizado no âmbito da área do problema é descrito neste capítulo, quer de um modo mais geral, quer de um modo mais próximo do problema em análise.

Capítulo 4

Neste capítulo é descrito o problema, e é apresentada uma proposta que visa

solucioná-lo. Esta proposta é aplicada nos capítulos seguintes.

Capítulo 5

São analisados os documentos utilizados, que correspondem a relatórios médicos. É feita uma descrição geral dos documentos, e são enumerados todos os termos necessários à criação da ontologia (capítulo 6).

Capítulo 6

A criação da ontologia que permite estruturar a informação proveniente dos relatórios é feita neste capítulo. A partir da enumeração dos termos, são criadas as classes, as suas propriedades e respectivos relacionamentos.

Capítulo 7

É criado um sistema de extracção de informação, que extrai a informação dos relatórios através do processamento de linguagem natural, de técnicas de aprendizagem automática e de alguns algoritmos desenvolvidos, que permitem povoar a ontologia.

Capítulo 8

Neste capítulo é apresentada uma aplicação, que permite uma interacção com o utilizador de forma simples e eficaz. São também realizados alguns testes, de modo a avaliar as instâncias criadas na ontologia após a conversão dos relatórios.

Capítulo 9

Por fim, são apresentadas as principais conclusões e descrevem-se os problemas em aberto para uma investigação futura.

Capítulo 2

Fundamentação Teórica

Este capítulo faz uma introdução aos conceitos teóricos e à área em que se enquadra o trabalho apresentado nesta dissertação.

A secção 2.1 descreve a Web Semântica, onde são descritos quais os seus objetivos (2.1.1), as suas mudanças em relação à *Web* actual (2.1.2) e a sua arquitectura (2.1.3).

A secção 2.2 corresponde ao conceito ontologia, onde são descritos os componentes de uma ontologia (2.2.1), os vários tipos (2.2.2), as metodologias para construção de ontologias (2.2.3), as linguagens ontológicas (2.2.4), algumas ferramentas usadas nesta área (2.2.5), um *framework* para manipulação de ontologias (2.2.6) e o uso de ontologias na área médica ou da saúde (2.2.7).

Na secção 2.3 é feita uma breve descrição à área da Mineração de Textos, e o seu uso na área da Inteligência Artificial através de algoritmos e técnicas das áreas do Processamento de Linguagem Natural (2.3.1) e da Aprendizagem Automática (2.3.2).

2.1 Web Semântica

Actualmente, a *Web* assume-se como o meio mais utilizado para comunicação, publicação e partilha de informação. As redes sociais, o comércio e os serviços têm provocado um crescimento exponencial da Internet, que conta cada vez com mais utilizadores.

A *Web* pode ser vista como uma rede de servidores que proporcionam o acesso a um conjunto de documentos ou páginas ligadas entre si. O constante aumento do número de servidores e de páginas *Web* faz com que a rede de informação assumam dimensões gigantescas, tornando a *Web* no maior repositório de informação da actualidade.

A informação contida na *Web*, tem sido pensada apenas para leitura por parte das pessoas, fazendo com que a *Web* seja voltada apenas para o consumo humano e não para os computadores [13]. Tais factores, aliados ao grande crescimento da *Web* tem provocado com que o ser humano não consiga aproveitar todas as suas potencialidades, perdendo-se facilmente entre documentos que possuem endereços URL (*Uniform Resource Locator*) e que referem muitas vezes outros documentos

com *links* para os respectivos endereços.

Mesmo com o auxílio dos motores de pesquisa, torna-se difícil ao ser humano encontrar aquilo que deseja ou necessita, pois o número de resultados devolvidos é exageradamente grande e impreciso. Tal acontece, pois a *Web* actual não fornece condições que possibilitem a estes motores de pesquisa distinguir entre os vários significados semânticos que um termo pode comportar.

Com o objectivo de ultrapassar estas limitações está em curso uma nova geração da *Web*, a Web Semântica, que aposta na estruturação dos recursos de informação acrescentando informação complementar que represente a semântica do seu conteúdo, através de linguagens específicas que permitem o processamento por máquinas, facilitando assim a sua consulta através de aplicações desenvolvidas para o efeito.

Segundo Bernes-Lee [3], o primeiro passo para o desenvolvimento da Web Semântica é a inclusão dos dados num formato cujos sistemas computacionais possam compreender de forma directa ou indirecta.

Mais tarde surgiu uma nova definição, onde Berners-Lee et al. [2] afirmam que "a Web Semântica é uma extensão da *Web* actual, onde a informação possui um significado claro e definido, possibilitando uma melhor interacção entre computadores e pessoas". É assim evidente, que o grande objectivo da Web Semântica é permitir que as pessoas aproveitem ao máximo todo o potencial da *Web*, mas para isso torna-se necessário o uso de máquinas para o processamento automático da informação.

Para que a *Web* inclua a perspectiva das máquinas, há que encontrar soluções de estruturação, integração, troca e compreensão semântica da informação, tanto na óptica dos humanos como na óptica das máquinas, através dos seus agentes inteligentes [55].

2.1.1 Objectivo

A Web Semântica é uma iniciativa/projecto liderado pelo W3C (*World Wide Web Consortium*), que se tem dedicado ao desenvolvimento de tecnologias e *standards*, com o objectivo de criar um meio para troca de informação, atribuindo significado semântico ao conteúdo da *Web*, permitindo que esse significado seja compreendido não só por humanos como por máquinas.

Os objectivos principais deste projecto são [41]:

- Construir uma estrutura comum de representação dos dados de modo a facilitar a integração de diversas fontes de informação e a extrair um novo conhecimento;
- Aumentar a utilidade da informação conectando-a aos seus conceitos e ao seu contexto;
- Descoberta e análise da informação mais eficiente.

Deste modo, ao efectuar-se uma pesquisa os resultados serão mais precisos e de acordo com as pretensões e necessidades do utilizador.

Actualmente, já existem documentos na *Web* no formato da Web Semântica, mas é um quantidade insignificante para o tamanho que hoje em dia a *Web* possui. No entanto, grande parte dos utilizadores da *Web* já está consciente da importância que a Web Semântica pode ter no futuro, permitindo que a longo prazo a Web Semântica possa atingir as dimensões da *Web* actual.

2.1.2 Mudanças da Web Actual para a Web Semântica

Um dos princípios básicos que constituem a Web Semântica, é o princípio de que as pessoas, os animais, os lugares, os objectos e tudo o que existe no mundo pode ser identificado por um URI (*Uniform Resource Identifier*) [41]. Através deste identificador, é possível especificar um conjunto de características sobre uma entidade. Pode-se então identificar um lugar, como a Universidade de Évora, referindo-se ao URI da sua página *Web*, sendo também possível referir entidades de modo indirecto, como por exemplo, utilizando-se o URI da caixa de e-mail de uma pessoa.

A *Web* actual possui documentos pensados para a análise humana, sem meta-informação a explicar a sua relação com outros documentos. É relativamente fácil ao ser humano identificar um *link* num documento e perceber qual é a informação do seu conteúdo. Porém, tais informações não estão acessíveis para as máquinas, pois os *links* na *Web* actual não indicam quais os tipos de relações entre os recursos.

Na Web Semântica os *links* podem possuir diferentes tipos, possibilitando a definição de conceitos e permitindo que as máquinas façam o processamento dessa informação.

A figura 2.1, mostra como são referenciados os *links* na *Web* actual e alguns tipos de relacionamentos possíveis na Web Semântica.

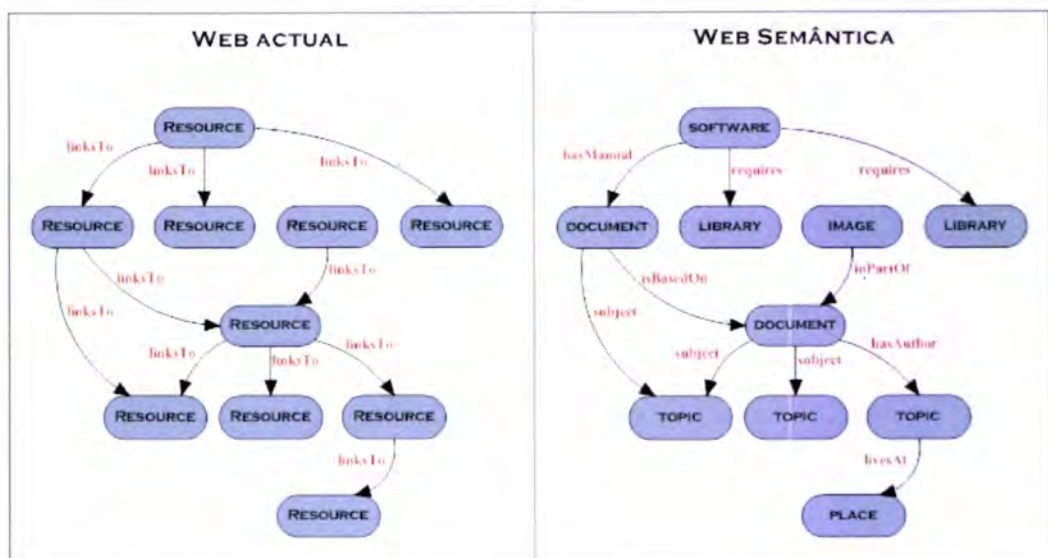


Figura 2.1: Estruturação dos recursos e links na Web actual e na Web Semântica [41]

A Web Semântica engloba as tecnologias da *Web* actual com a representação formal do conhecimento [24]. O conhecimento na Web Semântica deverá ser formado utilizando conceitos de meta-informação, agentes, ontologias, *Web Services*, linguagens e ainda outras tecnologias, possibilitando o processamento da informação [6].

Para Berner-Lee et al. [2], a Web Semântica poderá ser melhor explorada a partir do desenvolvimento de agentes computacionais que possibilitem armazenar as informações vindas de diversas fontes, relacioná-las automaticamente e retorná-las de maneira mais organizada aos utilizadores.

A figura 2.2 mostra como os agentes computacionais podem ser usados na recuperação de informação na Web Semântica.

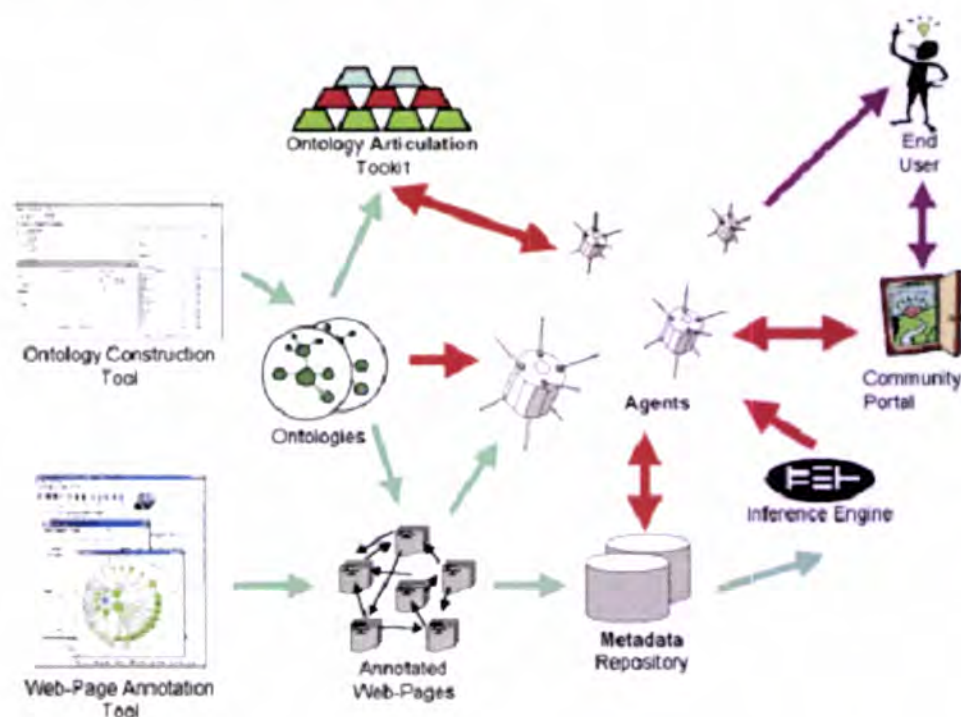


Figura 2.2: Utilização de Agentes Computacionais na Web Semântica [2]

2.1.3 Arquitectura da Web Semântica

Com o objectivo de facilitar o desenvolvimento e padronização das tecnologias relacionadas com a Web Semântica, o W3C divulgou uma arquitectura para a Web Semântica baseada em camadas sobrepostas, onde cada camada ou tecnologia deve obrigatoriamente ser complementar e compatível com as camadas inferiores, e não depender das camadas superiores.

A figura 2.3, representa as camadas lógicas da arquitectura da Web Semântica.

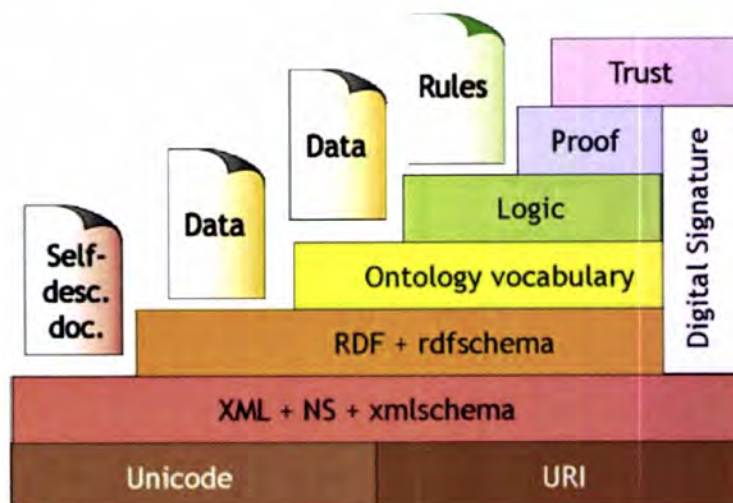


Figura 2.3: Arquitectura da Web Semântica [4]

As camadas inferiores desta arquitectura podem ser designadas de Camada Esquema (*Schema Layer*) que é responsável por estruturar os dados e definir o seu significado [7][9][35][40]. Esta camada genérica, engloba as camadas *Unicode*, *URI*, *XML+NS+xmlschema* e *RDF+rdfschema*.

A camada *URI* permite referir documentos, objectos e eventos através do respectivo endereço e de acordo com a norma *Unicode*.

A camada *XML* permite representar a estrutura dos dados baseada em *standards XML*, *namespaces* e *XML Schema* fornecendo um formato para intercâmbio de informação.

Já a camada *RDF* e *RDF Schema* engloba as linguagens da Web Semântica, fornecendo um conjunto de primitivas que permitem criar vocabulário ontológico (*Ontology vocabulary*) e representar o conhecimento de um determinado domínio.

A camada ontologia (*Ontology Layer*) é portanto constituída pelo vocabulário ontológico que fornece o significado dos termos de determinada área do conhecimento.

As camadas superiores *Logic*, *Proof* e *Trust* permitem definir um conjunto de regras lógicas que fornecem a troca de provas. A camada *Trust* possui mecanismos que permitem confiar ou não em determinada prova, tais como, o uso de assinaturas digitais como garantia da idoneidade da informação [21].

Nos últimos anos esta arquitectura tem sofrido pequenas alterações. Foi acrescentado a criptografia para garantir, juntamente com a assinatura digital, a confidencialidade das informações na Web Semântica. Uma das últimas alterações diz respeito à introdução da linguagem *OWL* que figura actualmente como recomendação para o desenvolvimento de ontologias, e foram também incorporadas tecnologias como *SPARQL (RDF Query Language and Protocol)* e *DPL (Description Logic Programs)* que proporcionam uma estrutura de conhecimento mais flexível e que facilitam a realização de consultas semânticas [10].

Numa *Web* futura, muitas camadas e muitas tecnologias serão acrescentadas a esta arquitectura, de modo a ser possível criar documentos com informação estruturada e com conteúdo semântico, possibilitando o processamento da informação por máquinas sem ser necessário a intervenção humana.

A secção seguinte assenta na camada ontologia.

2.2 Ontologia

Ontologia é uma palavra que tem origem no campo da filosofia e pode ser interpretado como o estudo (*-logia*) da existência ou do ser (*onto-*). Segundo Gruber, este termo significa "descrição explícita da existência" [25].

O termo ontologia foi posteriormente adoptado na área da Inteligência Artificial (IA) para designar artefactos usados como modelos formais de domínio [26].

As ontologias vêm sendo muito utilizadas na integração de informação, recuperação de informação e Web Semântica, sobre um determinado domínio ou nas diversas áreas do conhecimento [37], tais como a medicina.

Existem diversas definições para ontologia, porém há um consenso, as ontologias permitem estruturar e modelar conceitos e relações de forma correcta, evitando ambiguidades e estruturando o conhecimento existente de um dado domínio [51].

2.2.1 Componentes de uma Ontologia

Na Web Semântica, as ontologias são de grande utilidade, utilizando os seguintes componentes:

- classes para representar conceitos de um domínio que são geralmente organizadas em taxonomias¹, podendo haver herança de classes;
- propriedades de cada conceito (atributos que representam características das classes e relações que correspondem aos tipos de associações entre classes do domínio);
- instâncias das classes.

As ontologias são de facto um modo de representar o conhecimento de um dado domínio, orientadas para as máquinas e para o respectivo processamento automático da informação, principalmente a partir da semântica expressa em documentos *Web*.

¹Taxonomia: refere-se à classificação das coisas ou aos princípios subjacentes da classificação

2.2.2 Tipos de Ontologia

Segundo Guarino [26], as ontologias podem ter diferentes níveis, de acordo com os conceitos representados.

As chamadas ontologias de alto nível (*top-level ontologies*) ou ontologias superiores (*upper ontologies*), representam conceitos gerais, independentes do domínio.

As ontologias de domínio representam o conhecimento de um domínio específico.

As ontologias de tarefa descrevem como as tarefas ou actividades são conduzidas (independentemente do domínio).

Por fim, as ontologias de aplicação descrevem conceitos que dependem de igual modo de um domínio específico ou do tipo de tarefa.

A figura 2.4 representa a relação entre os quatro tipos de ontologias, do topo (mais geral) para baixo (mais específico).



Figura 2.4: Tipos de Ontologias

No trabalho apresentado nesta dissertação são utilizadas ontologias de domínio, pois o objectivo é compartilhar o conhecimento sobre um domínio delimitado.

2.2.3 Metodologias para Construção de Ontologias

Quando se pretende construir o esquema de uma ontologia, deve-se ter sempre em conta a utilização de uma ontologia pré-existente ao invés de construir uma ontologia nova. No entanto, e para ambos os casos, deve-se sempre aplicar uma metodologia durante o processo de desenvolvimento. Em seguida, são descritas duas das principais metodologias para a construção de ontologias.

As metodologias de desenvolvimento de ontologias podem ser classificadas de acordo com a abordagem de desenho e a partir dos conceitos do domínio [17], do seguinte modo:

- *Bottom-up* (de baixo para cima)
Começa com a definição das classes mais específicas, com subsequente agrupamento destas classes em conceitos mais gerais;
- *Top-Down* (de cima para baixo)
Começa com a definição dos conceitos mais gerais no domínio e posteriormente, é feita a especialização dos conceitos;
- *Middle-Out* (do centro para fora)
Começa com a definição dos conceitos mais relevantes para o domínio e segue tanto no sentido da generalização quanto da especialização.

Outra metodologia que é bastante referenciada na literatura é a proposta pelo grupo da Universidade de Stanford, que refere que a modelagem de uma ontologia pode ser realizada seguindo alguns passos que representam a sequência de criação e modelagem da ontologia [45]:

1. Determinar o domínio e escopo da ontologia, que pode ser realizado através das seguintes questões:
 - Qual o domínio e uso da ontologia?
 - Quais os tipos de questões que a ontologia deveria responder?
 - Quem usará e fará a manutenção desta ontologia?
2. Considerar a reutilização de ontologias existentes;
3. Enumerar termos importantes na ontologia;
4. Definir as classes e sua hierarquia;
5. Definir as propriedades das classes;
6. Definir os tipos de valores e cardinalidade das propriedades;
7. Criar instâncias das classes.

A figura 2.5 corresponde aos passos da criação da ontologia deste tipo de metodologia, mas de modo simplificado.



Figura 2.5: Passos da Metodologia para a Criação da Ontologia

2.2.4 Linguagens de Representação de Ontologias

Com a constante evolução da Web Semântica, têm surgindo várias linguagens capazes de representar ontologias.

As principais linguagens ontológicas são: RDF, RDFS, SHOE, OIL, DAML, DAML+OIL e OWL. A figura 2.6 representa as camadas das linguagens de ontologia para a Web Semântica.

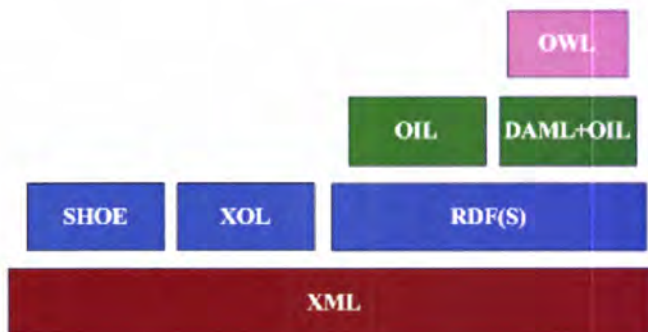


Figura 2.6: Camadas das Linguagens de Ontologias

A partir da análise à figura, percebe-se que o XML é base do desenvolvimento das linguagens SHOE, XOL e RDF, enquanto que para o OIL e DAML+OIL, a linguagem RDF é a base. Por fim, a linguagem OWL é uma evolução de DAML+OIL e por consequência da linguagem RDF.

Em seguida são descritas as principais linguagens ontológicas, embora a linguagem OWL seja descrita pormenorizadamente porque é hoje em dia a linguagem que encontra maior aceitação por parte dos utilizadores da Web Semântica, e porque foi a linguagem utilizada no trabalho descrito nesta dissertação.

RDF

A linguagem RDF² (*Resource Description Framework*) é utilizada para representar informações sobre os recursos da World Wide Web [38]. Esta linguagem foi desenvolvida para descrever meta-informação sobre recursos *Web*, tais como o título, autor e data de criação de uma página *Web*. O uso do RDF tem mais aplicabilidade através da generalização do conceito de um recurso *Web*, que possibilita o uso desta linguagem para expressar informações sobre as coisas que possam ser identificadas na *Web*.

A linguagem RDF permite a troca de informação entre aplicações, sem que ocorra perda de significado.

²<http://www.w3.org/TR/rdf-nt/>

A estrutura RDF é baseada na identificação de recursos na *Web*, e pode ser vista como uma tripla constituída por um sujeito, um predicado que descreve um conjunto de propriedades ao qual se atribui um valor e que expressa uma relação, e um objecto. Tanto o sujeito como o objecto, são identificados através de URIs, embora o objecto possa também ser identificado por um literal.

A sintaxe RDF é baseada em XML (*Extensible Markup Language*), chamada RDF/XML, que permite o processamento por máquinas em ambiente *Web*. No entanto, as URIs em RDF podem referir-se a qualquer coisa identificável, sem que tenha de ser um recurso *Web*. Assim, além de descrever páginas *Web*, com RDF é possível descrever pessoas, carros, motos, etc.

RDFS

A linguagem RDF permite expressar sentenças sobre recursos, no entanto, começou a existir a necessidade de definir vocabulários (termos) a serem utilizados nessas sentenças. Surgiu assim o RDFS³ (*RDF Schema*) que é uma extensão da linguagem RDF mas com vocabulário que permite descrever classes de objectos, as suas propriedades e o tipo de dados para o valor dessas propriedades [8].

Em RDFS, a definição de classes ocorre através do vocabulário da linguagem: **rdfs:Resource**, **rdf:Property** e **rdfs:Class**. Tudo o que é descrito em RDF são recursos e são instâncias da classe **rdfs:Resource**. A classe **rdf:Property** é utilizada para caracterizar as instâncias de **rdfs:Resource**. Por fim o **rdfs:Class** é utilizada para definir conceitos em RDFS.

Para as propriedades dos objectos, a linguagem RDFS possui também vocabulário específico: **rdfs:domain** e **rdfs:range** representam respectivamente a classe que possui a propriedade e a classe que está relacionada com a classe de domínio ou o tipo de valor dessa propriedade.

RDFS permite descrever conhecimento simples, não sendo uma linguagem suficientemente "poderosa" para representar conhecimento mais complexo, como por exemplo: afirmar que um relatório médico tem um único paciente, tem de ser realizado por um médico, ser executado por um ou vários técnicos e apresentar uma conclusão.

SHOE

SHOE⁴ (*Simple HTML Ontology Extensions*) é uma linguagem para representação de conhecimento baseada em ontologias para a *Web*, que foi desenvolvida por estudantes e investigadores da Universidade de Maryland [27].

³<http://www.w3.org/TR/rdf-schema/>

⁴<http://www.cs.umd.edu/projects/plus/SHOE/>

A linguagem SHOE é uma extensão de HTML (*HyperText Markup Language*) com tags que fornecem uma estrutura capaz de representar conhecimento, e como tal, necessárias para introduzir indicações semânticas.

Para ser compatível com os padrões da *Web*, a sintaxe SHOE é definida como uma aplicação de SGML (*Standard Generalized Markup Language*). No entanto, existe uma versão do SHOE que utiliza XML ao invés do original SGML [28].

SHOE não possui uma fundamentação lógica formal, e como tal, aspectos fundamentais como a negação e a cardinalidade não estão presentes nesta linguagem.

Actualmente, o projecto SHOE não se encontra em desenvolvimento, mas funcionou como suporte para linguagens posteriores, tais como: DAML+OIL e OWL.

OIL, DAML e DAML+OIL

A linguagem OIL⁵ (*Ontology Inference Layer*) resulta do projecto de investigação On-to-Knowledge financiado pela União Europeia. Surgiu com a necessidade de se obter uma linguagem que permitisse a modelagem de ontologias na *Web*, já que as linguagens antecedentes não possuem formalismos e a semântica necessária de modo a suportarem mecanismos de inferência.

A linguagem DAML⁶ (*DARPA Agent Markup Language*) foi desenvolvido num projecto da organização americana DARPA. O objectivo da linguagem é facilitar a actividade de interacção de agentes de software autónomos na *Web* [29].

DAML permite restringir os valores de uma propriedade, além de contar com um conjunto de novas propriedades tais como: **daml:samePropertyAs**, **daml:equivalentTo**, **daml:inverseOf** e **daml:TransitiveProperty**.

DAML+IOL⁷ *DARPA Agent Markup Language+Ontology Inference Layer* surge nos finais de 2000 e resulta da cooperação entre os investigadores americanos e europeus que desenvolveram o DAML e OIL respectivamente. O papel do DAML no DAML+OIL é fornecer um conjunto de primitivas (classes, subclasses, restrições, etc.) para marcar informações e localizá-las dentro da estrutura conceitual do domínio. Já a função do OIL é fornecer os meios para que possam ser realizadas inferências [20].

OWL

A mais recente linguagem ontológica é a OWL⁸ (*Ontology Web Language*), que permite incluir a descrição das classes, as suas propriedades e respectivos relacionamentos. Esta linguagem é uma evolução das suas antecessoras RDF e DAML+OIL, e é hoje uma recomendação da W3C principalmente devido à sua semântica formal capaz de processar conteúdo semântico na *Web*.

⁵<http://www.ontoknowledge.org/oil/>

⁶<http://www.daml.org/>

⁷<http://www.w3.org/TR/daml+oil-reference>

⁸<http://www.w3.org/TR/owl-ref/>

OWL foi desenvolvida com o intuito de permitir às aplicações processar a informação, em vez de apenas apresentá-la aos utilizadores.

A linguagem OWL permite uma maior interoperabilidade entre máquinas do que a suportada pelos seus antecessores XML, RDF e RDFS, pois fornece um vocabulário adicional para definição de classes e suas propriedades e apresenta uma semântica formal. Esta linguagem fornece três sub-linguagens, cada uma oferece níveis de expressividade diferentes [39]:

- **OWL Lite** - Suporte aos utilizadores que necessitam de uma classificação hierárquica e restrições simples. Em relação às restrições de cardinalidade, só permite valores de cardinalidade 0 ou 1.
- **OWL DL** - Suporte aos utilizadores que querem o máximo de expressividade, mantendo completude computacional (todas as deduções são computáveis). OWL DL inclui todo o vocabulário da linguagem OWL, mas com certas restrições (por exemplo, uma classe não pode ser uma instância de outra classe). É ainda baseado em lógicas de descrição, que são formalismos para a representação de conhecimento.
- **OWL Full** - Suporte aos utilizadores que querem a máxima expressividade e liberdade sintáctica do RDF, sem preocupações computacionais. Com esta sub-linguagem é possível expandir o significado do vocabulário pré-definido do RDF e da OWL.

De acordo com McGuinness et al. [39], **OWL Full** pode ser considerada uma extensão do RDF, enquanto **OWL Lite** e **DL** podem ser vistos como extensões de conjuntos restritos do RDF.

O OWL Lite tem vocabulário para:

- *igualdades*: `equivalentClass`, `equivalentProperty` e `sameAs`;
- *desigualdades*: `differentFrom`, `disjointMembers` e `AllDifferent`;
- *propriedades*: `inverseOf`, `TransitiveProperty`, `SymmetricProperty`, `FunctionalProperty` e `InverseFunctionalProperty`;
- *restrições*: `minCardinality`, `maxCardinality`, `cardinality`;
- *intersecção de classes*: `intersectionOf`.

Para o OWL DL e Full é acrescentado o seguinte vocabulário:

- *axiomas*: `oneOf`, `disjointWith` e `equivalentClass`;
- *expressões booleanas*: `unionOf`, `complementOf` e `intersectionOf`;
- *informação*: `hasValue`.

Uma ontologia em OWL é composta por classes, que podem ser definidas como objectos, e pelas suas instâncias. Assim, uma classe descreve um objecto, que pode ter um conjunto de propriedades e relações com outros objectos.

No código de uma ontologia OWL, as classes são representadas com o prefixo **owl:Class**. Esta linguagem utiliza como predefinição a classe **owl:Thing** que é a classe raiz de qualquer hierarquia, e por isso, todas as classes criadas são membros desta classe.

Em seguida é apresentado um exemplo da linguagem OWL para representação de classes, onde a classe **Carro** é uma subclasse de **Veículo**, onde é visível a adopção de algum vocabulário das suas linguagens antecedentes.

```
<owl:Class rdf:ID="Carro">
  <rdfs:subClassOf rdf:resource="#Veículo"/>
</owl:Class>
```

Em relação às propriedades das classes estas dividem-se em dois tipos:

- **Object Properties** - correspondem às relações entre instâncias de duas classes.
- **Data Type Properties** - correspondem às relações entre uma instância de uma classe e um literal ou tipo de dados. Por outras palavras, correspondem aos atributos de uma classe.

Continuando o exemplo anterior, se acrescentarmos a classe **Pessoa** à hierarquia e as propriedades **proprietário** e **cor**, estas são representadas na linguagem OWL do seguinte modo:

```
<owl:ObjectProperty rdf:ID="proprietário">
  <rdfs:domain rdf:resource="#Pessoa"/>
  <rdfs:range rdf:resource="#Carro">
</owl:ObjectProperty>

<owl:DatatypeProperty rdf:ID="cor">
  <rdfs:domain rdf:resource="#Carro"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>
```

Como se pode verificar, e tal como na definição de classes, é utilizado vocabulário das suas linguagens antecedentes.

Na linguagem OWL é possível criar instâncias de uma classe. Também chamados de *individuals*, as instâncias representam um individuo ou objecto de uma classe.

Seguindo o exemplo anterior, as instâncias na linguagem OWL são representadas do seguinte modo:

```

<owl:Thing rdf:about="#Luis">
  <rdf:type rdf:resource="#Pessoa"/>
  <proprietario rdf:resource="#oMeuCarro"/>
</owl:Thing>

<Carro rdf:about="#oMeuCarro">
  <rdf:type rdf:resource="#owl:Thing"/>
  <cor>preto</cor>
</Carro>

```

O código OWL acima, representa uma instância da classe **Pessoa**, identificada por **Luis**, que é proprietário da instância **oMeuCarro**, que é uma classe de **Carro**, que por sua vez tem uma cor que está descrita com a string **preto**.

Em 2009 surgiu a nova versão do OWL, o OWL2⁹ que ainda está em fase de desenvolvimento, estando nesta altura como candidata a recomendação pela W3C.

A versão OWL2 possui três novos perfis:

- **OWL 2 EL** - Destina-se a aplicações que possuam um grande número de propriedades e/ou classes;
- **OWL 2 QL** - Destina-se a aplicações que utilizam grandes volumes de dados;
- **OWL 2 LR** - Destina-se a aplicações que exigem raciocínio evolutivo, sem sacrificar demais o poder expressivo.

Além dos novos perfis, a nova versão do OWL possui mais funcionalidades tais como:

- chaves (keys);
- redes de propriedades;
- tipos de dados mais ricos, intervalos entre dados;
- restrições de cardinalidade qualificada;
- novas propriedades tais como: asymmetric e reflexive;
- melhores capacidades de anotação.

Pretendeu-se apenas fazer uma pequena introdução a esta nova versão, pois esta versão ainda está em desenvolvimento e actualmente a versão antiga é a linguagem para representar ontologias recomendada pela W3C. Como tal, para uma descrição mais pormenorizada desta nova versão, recomenda-se a leitura do documento referido ao longo do texto.

⁹<http://www.w3.org/TR/owl2-profiles/>

2.2.5 Ferramentas

Actualmente, existem diversas ferramentas que permitem ao utilizador de forma fácil, construir e manipular ontologias. Estas ferramentas permitem definir classes, atributos para classes, axiomas e restrições. Muitas destas ferramentas possuem uma interface gráfica, que vai ao encontro dos padrões de conformidade utilizados na *Web* [37].

Existem muitas ferramentas para o desenvolvimento de ontologias, tais como: Protégé, OntoEdit, Ontolingua, WebODE, SymOntoX, WebONTO, OilEd, DOE, OntoBuidier entre muitas outras. Nesta secção é apresentada a ferramenta Protégé, porque segundo Cardoso [13] é a ferramenta mais utilizada no desenvolvimento de ontologias, e porque foi a ferramenta utilizada no desenvolvimento do trabalho apresentado nesta dissertação.

Protégé

O Protégé¹⁰ é uma ferramenta de desenvolvimento de ontologias, desenvolvido pelo departamento de informática médica da Universidade de Stanford, na Califórnia, Estados Unidos, e está disponível para acesso livre e em vários sistemas operacionais tais como Windows, Linux, Unix, entre outros. Esta ferramenta auxilia os utilizadores na construção de bases de conhecimento e ontologias.

Desenvolvido em Java, possui uma interface, como exibida na figura 2.7, que permite o acesso à barra de menus e barra de ferramentas, além de módulos de navegação e edição de classes, atributos, relações, instâncias e pesquisas na base de conhecimento, propiciando a entrada de dados e a recuperação das informações [48].

Por ser uma ferramenta com código aberto, permite a sua integração com diversas aplicações [46]. Além disso, permite que os utilizadores adicionem módulos (Plug-ins) com funcionalidades que não estão definidas no Protégé. Um dos recursos aproveitados e que é hoje em dia um dos módulos do Protégé, é o Ontoviz, um componente que faz com que o gerador de gráficos Graphviz produza gráficos com as heranças, instâncias e outros tipos de relacionamento.

O Protégé permite ainda importar ou exportar ontologias de/para diversos formatos tais como RDF(S), XML, XML Schema e OWL.

¹⁰<http://protege.stanford.edu/>

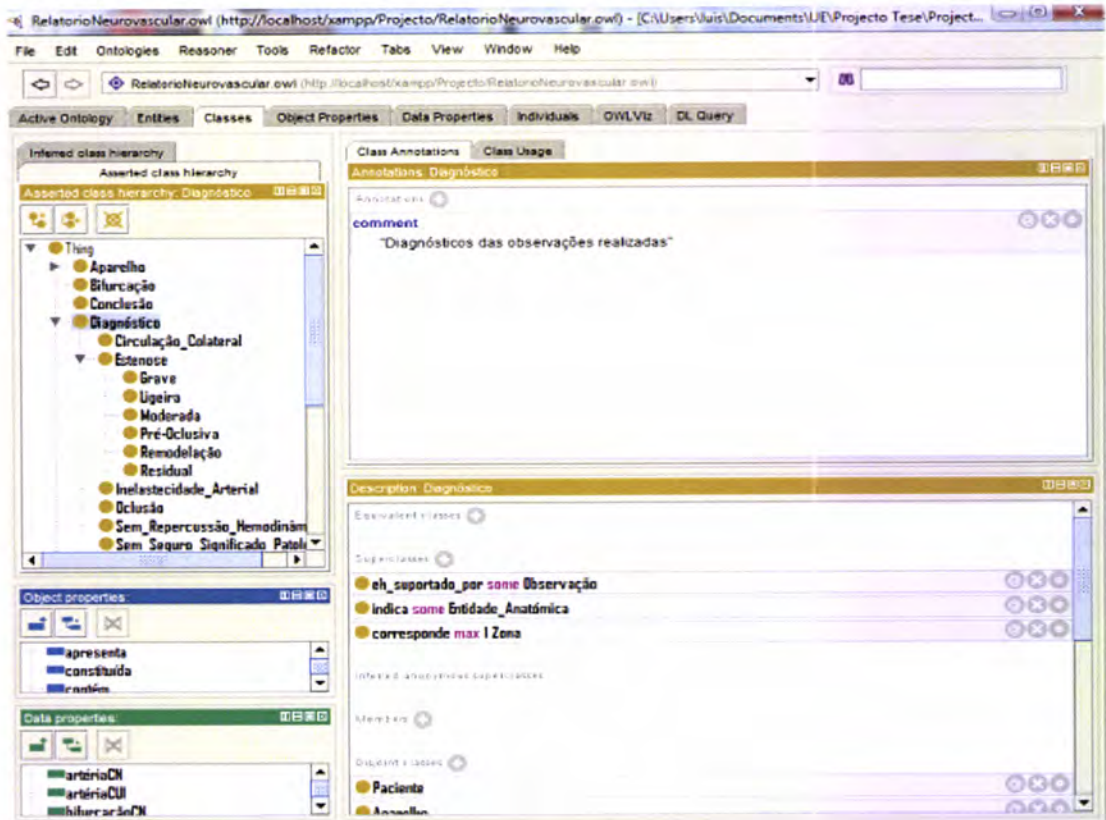


Figura 2.7: Interface da Ferramenta Protégé

2.2.6 Jena

O Jena¹¹ é um *framework* Java que foi desenvolvido pela Hewlett-Packard (HP) e destinado ao desenvolvimento de aplicações para a Web Semântica.

Inicialmente, na sua primeira versão, a API Jena apenas possuía métodos para manipulação de ontologias em DAML+OIL. Mais tarde, foi publicada a segunda versão do API Jena, onde foram acrescentados métodos que vêm permitir a manipulação de ontologias em RDFS e OWL [14].

Esta segunda versão da API Jena fornece um pacote específico denominado API Jena 2 Ontology, que dispõe de duas classes específicas para manipulação de ontologias: a `OntClass` e a `ObjectProperty` que correspondem respectivamente à classe Java que representa a classe OWL e as propriedades de uma classe OWL [32].

Na API Jena, para cada uma das linguagens de ontologias existentes, existem parâmetros com sintaxes diferentes, por exemplo, no parâmetro da linguagem DAML, o URI para uma propriedade do objecto é `daml:ObjectProperty` e na linguagem OWL é `owl:ObjectProperty` [32].

¹¹<http://jena.sourceforge.net/>

Com este *framework*, é assim possível realizar inferências e manipular uma ontologia que possua as especificações da Web Semântica como RDF e a linguagem OWL.

2.2.7 Ontologias Médicas

A constante evolução das tecnologias, tem provocado um grande aumento de informação na área da saúde, originando uma base de conhecimento cada vez maior. Esta informação está naturalmente descentralizada e a sua integração é um dos grandes desafios para a área da informática em saúde. O facto de muitas instituições/médicos utilizarem vocabulários próprios ou simplesmente não codificarem algumas informações, aumenta a complexidade deste problema.

A necessidade de aproveitar a informação desta área, tem feito com que nos últimos anos se venha a procurar soluções para padronizar essa informação. Uma das soluções encontradas é através do uso de ontologias [43]. Além disso, têm também sido desenvolvidos uma enorme variedade de sistemas e vocabulários de modo a agrupar e categorizar termos médicos, tais como: UMLS¹² (*Unified Medical Language System*), MeSH¹³ (*Medical Subject Headings*), SNOMED¹⁴ (*Systematized Nomenclature of Medicine*), entre outros.

Apesar destes sistemas e vocabulários serem classificados por muitos autores como ontologias [43], são utilizados na maior parte das vezes como classificadores de termos, no auxílio à construção de ontologias e à padronização dos termos médicos dentro destas ontologias.

Nesta secção é apresentado o UMLS, por ser um dos classificadores de termos com maior dimensão e dos mais utilizados e referidos na literatura, e porque foi o classificador utilizado no desenvolvimento do trabalho apresentado nesta dissertação.

UMLS

O Unified Medical Language System (UMLS) tem sido desenvolvido pela National Library of Medicine (NLM) desde 1986. É um sistema que resolve os problemas da semântica na área médica, através de milhares de conceitos provenientes de diversos vocabulários biomédicos [5].

Este sistema destina-se a ajudar os profissionais de saúde e investigadores biomédicos a utilizar informações de diferentes fontes através do mapeamento de milhares de terminologias [50].

A NLM, permite o acesso à base de dados do UMLS, através de chamadas ao UMLSKS (*UMLS Knowledge Sources*) que é composto por três bases de conhecimento: o Metathesaurus, uma rede semântica e um especialista léxico. A padronização da informação da área médica ocorre através do Metathesaurus, que

¹²<http://www.nlm.nih.gov/research/umls/>

¹³<http://www.nlm.nih.gov/mesh>

¹⁴<http://www.snomed.org/>

contém informações sobre conceitos biomédicos e relacionados com a saúde, e que contém na sua base de conhecimento vários vocabulários em diversas línguas, tais como o SNOMED e MeSH.

Apesar do UMLS ser utilizado em grande parte para construção de ontologias, um outro exemplo de utilização é o Medical World Search¹⁵. Ao ser realizada uma pesquisa de um conceito médico, o Medical World Search utiliza o UMLS para incluir sinónimos na pergunta realizada na pesquisa, de modo a encontrar mais facilmente a informação pretendida.

2.3 Mineração de Textos

A constante evolução da Internet e a respectiva sobrecarga de informação, provocou dificuldades quando se pretendia analisar, localizar e aceder a essa informação de modo manual. Para ultrapassar estes obstáculos, foram desenvolvidas teorias e ferramentas computacionais para realizarem análises automáticas em texto.

As primeiras técnicas a serem desenvolvidas, foram as técnicas de extracção de informação em Base de Dados, originando assim a área de Descoberta de Conhecimento de Base de Dados (*Knowledge Discovery in Data* - KDD) ou Mineração de Dados. Estas técnicas eram aplicadas a um grande número de dados.

Mais tarde, estas técnicas foram aplicadas a dados unicamente textuais. No entanto, devido a estes dados serem geralmente guardados de modo não estruturado, fez com que se tivessem de desenvolver novas técnicas pois as actuais apenas obtinham resultados em dados estruturados. Estas novas técnicas deram origem a uma nova área, a área de Descoberta de Conhecimento em Texto (*Knowledge Discovery in Text* - KDT) ou Mineração de Texto (*Text Mining*) [30].

Segundo o *Text Mining Research Group*¹⁶, a Mineração de Textos (MT) procura padrões, tendências e regularidades em textos escritos em língua natural, com dados não estruturados, de modo a extrair informação para um propósito em particular.

A MT é considerada uma área emergente de pesquisas [19] e vem sendo utilizada na Inteligência Artificial, envolvendo técnicas, métodos e algoritmos de algumas áreas tais como: Processamento de Língua Natural, Aprendizagem Automática, Recuperação de Informação, Extracção de Informação, entre outras.

A área de MT é bastante complexa e envolve três etapas: Pré-processamento, Análise dos Dados e Extracção do Conhecimento e Pós-processamento. A figura 2.8, representa estas etapas.

Antes de ser aplicada a primeira etapa, são obtidos os documentos sobre o qual se quer utilizar o conhecimento. O conjunto destes documentos formam o *corpus*.

¹⁵<http://www.mwsearch.com/>

¹⁶<http://textmining.zcu.cz/>



Figura 2.8: Etapas da Mineração de Textos [53]

A etapa do Pré-processamento ocorre com o intuito de "tratar" os dados, através da sua normalização, classificação e estruturação, e de modo a facilitar a análise dos dados na etapa seguinte.

A segunda etapa consiste em encontrar padrões nos documentos do *corpus*, que permitam extrair o conhecimento a partir da execução de algoritmos específicos.

Por fim, a etapa do Pós-processamento consiste em interpretar e validar os resultados obtidos na fase anterior, desenvolvendo, como consequência, a descoberta do conhecimento.

2.3.1 Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) é uma subárea da Inteligência Artificial, que tem desempenhado múltiplos papéis nesse contexto, sendo que a sua importância cresce e decresce em consequência das mudanças tecnológicas. As suas ferramentas são bastante utilizadas no pré-processamento e modelagem de textos.

O PLN ocupa-se da análise, codificação e manipulação das informações expressas em língua natural, de modo a permitir que os seres humanos comuniquem da forma mais natural possível com os computadores, e de forma a permitir que os computadores "entendam" e consigam processar informação em linguagem natural.

Na actualidade, a importância do PLN no campo da pesquisa atingiu níveis muito elevados. Existe a preocupação de cada vez mais melhorar a interacção homem-máquina, no entanto, o processamento textual é compreendido como uma tarefa muito complexa, pois possui como finalidade determinar a relação entre as

palavras de modo a conseguir extrair o significado destas.

Uma das formas de tentar ultrapassar a complexidade do PLN, é tentar dividir o problema em vários subproblemas menores que são resolvidos por tarefas específicas. Assim, o *corpus* com os documentos são separados em documentos, documentos "partidos" em parágrafos, parágrafos em frases e frases em palavras, que são etiquetadas de acordo com a sua classe gramatical [31].

A linguagem natural subdivide-se em vários níveis [33]:

- *Nível Morfológico*: estudo da constituição das palavras em elementos básicos;
- *Nível Sintáctico*: determinação da relação (papel) de um conjunto de palavras numa sentença;
- *Nível Semântico*: determinação do significado e inter-relacionamento semântico das palavras;
- *Nível Discursivo*: objectiva-se em determinar o significado de um conjunto de sentenças;
- *Nível Pragmático*: visa determinar o objectivo do uso da língua.

As principais técnicas para PLN são: remoção de *stopwords*, *stemming* e tokenização.

As *stopwords* são termos frequentes nos textos, que não possuem informação que possa ser considerada relevante no processamento dos documentos. A sua remoção resulta numa melhor compreensão dos textos, pois permite uma redução dos termos analisados nos documentos e na diminuição das palavras que são guardadas na base de dados.

Normalmente a remoção de *stopwords* ocorre comparando a palavra que está a ser processada no documento, com as palavras existentes numa *stoplist*, e que correspondem a preposições, conjunções, artigos, etc.

Após a remoção de *stopwords*, ocorre a normalização das palavras através de algoritmos de Radicalização ou *Stemming*, que permitem simplificar as palavras e suas variantes numa forma comum, resultando numa diminuição do número de palavras guardadas na base de dados [47].

A principal aplicação do *stemming* é através do algoritmo de Porter [49], que consiste na remoção de sufixos e/ou prefixos das palavras substituindo-as por radicais comuns.

Por fim, a *tokenização* tem como finalidade separar um documento em secções de *tokens*, mas, de modo a exprimirem a mesma semântica original do texto. As palavras do texto são separadas por *tokens* de controle ou formatação, tais como espaços ou sinais de pontuação, e que em alguns casos se podem chamar de *tokens* delimitadores [19].

A figura 2.9, representa uma metodologia que é proposta em [34] para identificação de *tokens*, e que através do uso de dicionários de dados e regras de formação de palavras, procura manter a semântica apresentada pelos *tokens* antes da execução do processo de tokenização.

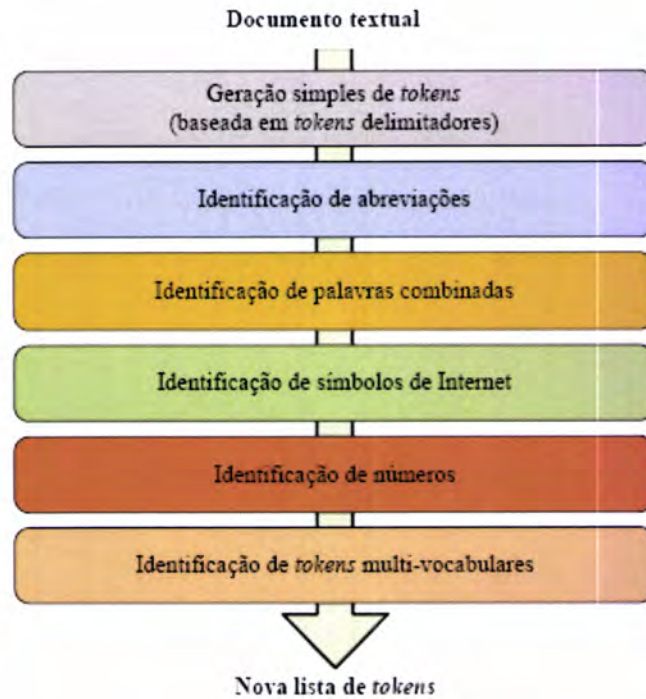


Figura 2.9: Metodologia de Identificação de Tokens [34]

2.3.2 Aprendizagem Automática

A aprendizagem automática ou aprendizagem de máquina (*Machine Learning*) é uma subárea da Inteligência Artificial, que se dedica ao estudo e desenvolvimento de algoritmos e técnicas que permitam ao computador aprender e aperfeiçoar o seu desempenho em determinada tarefa.

Segundo Mitchell [42], um computador aprende a partir da experiência **E**, em relação a um conjunto de tarefas **T**, com medida de desempenho **D**, se seu desempenho em **T**, medido por **D**, melhora com **E**.

Existem essencialmente dois tipos de paradigmas de aprendizagem:

- **Aprendizagem Supervisionada** - Neste tipo de aprendizagem existe um "coordenador", que para um conjunto de dados de entrada indica a sua saída. É construído um classificador que possa determinar correctamente o conjunto de novos exemplos ainda não rotulados. Para rótulos de conjuntos discretos, esse problema é chamado de **classificação** e para valores contínuos como **regressão**;
- **Aprendizagem não Supervisionada** - Neste tipo de aprendizagem não existe um "coordenador". A aprendizagem baseia-se na descoberta de relações entre as características dos dados de entrada, codificando-os na saída.

Têm sido desenvolvidos muitos algoritmos de aprendizagem, cada um deles com características específicas de acordo com determinados tipos de problemas. De entre os mais conhecidos, destacam-se as Árvores de Decisão, as Redes Neurais, os classificadores Bayseanos, entre outros.

No trabalho apresentado nesta dissertação foram utilizadas as Árvores de Decisão, de modo a classificarem os termos médicos de entrada e a atribuírem uma classe de saída. As Árvores de Decisão, têm sido muito utilizadas na área da saúde, na aplicação de diversos trabalhos [12][15][16][23].

Árvores de Decisão

As Árvores de Decisão permitem construir um modelo de fácil compreensão e interpretação para qualquer ser humano [54]. Cada nó da árvore representa um teste de algum atributo, e cada nó a partir daquele ramo corresponde a um valor que esse atributo pode ter. Este processo de teste e ramificação é recursivo até cada uma das folhas, que contêm as possíveis classificações para cada exemplo.

Uma árvore de decisão representa uma disjunção de conjunções, onde cada ramo completo representa uma conjunção de atributos, enquanto os nós provêm de disjunções.

Existem diversos algoritmos de aprendizagem em árvores de decisão. Um dos algoritmos mais populares é o ID3 que foi desenvolvido por Quinlan [52]. A construção da árvore de decisão ocorre de cima para baixo e em cada nó é tomada uma decisão de qual atributo deve-se testar.

A figura 2.10 representa uma árvore de decisão, sobre a possibilidade de um ser humano tirar a carta de condução.

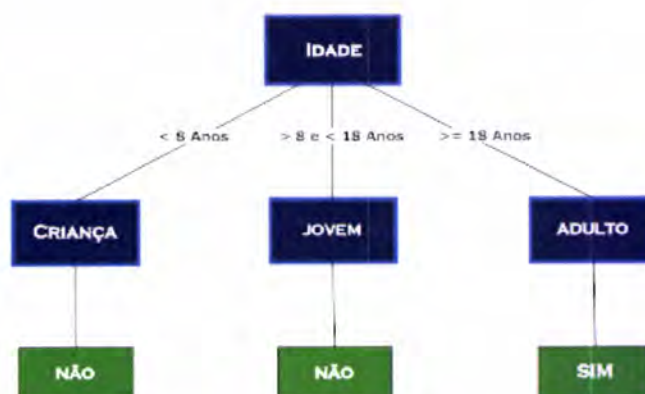


Figura 2.10: Exemplo de Árvore de Decisão

Capítulo 3

Trabalho Relacionado

Este capítulo faz uma breve análise aos trabalhos relacionados com a área do projecto desta dissertação.

Na secção 3.1 é feita uma análise aos trabalhos existentes na área da saúde em geral. Na secção 3.2 é descrito como criar uma ontologia para relatórios médicos.

3.1 Ontologias na Saúde

Nas últimas décadas, existe um claro crescimento no processo de informatização na área da saúde. Foram desenvolvidos diversos sistemas computacionais dirigidos para a captura, armazenamento e disponibilização de informações médicas. O uso de ontologias na área da saúde, permite estruturar essa informação e criar uma base de conhecimento que pode ser usada para investigação, para criação de sistemas de suporte ao apoio diagnóstico, entre outros.

Em 2003, Nardon [44] utiliza ontologias e base de dados dedutivas para partilha de conhecimento da área da saúde. São utilizadas tecnologias da Web Semântica em conjunto com uma linguagem de consulta baseada em base de dados dedutivas, de modo a permitir a partilha da base de conhecimento e a resolução de consultas mais complexas.

Mais tarde, em 2005, Cantais et al. [11] desenvolveram um protótipo que faz parte do projecto Personalized Information Platform (PIPS) que tem como objectivo auxiliar um paciente diabético na escolha dos alimentos. O sistema utiliza ontologias como base de conhecimento e foram utilizados os termos da base de dados de alimentos do instituto Entidade de Investigação e Desenvolvimento da Universidade Politécnica de Valência (ITACA).

No artigo [18], Farias et al. constroem uma ontologia para a gestão do conhecimento das Doenças Sexualmente Transmissíveis, trazendo consigo a conceitualização necessária para a compreensão do processo de desenvolvimento de ontologias.

Já em 2007, Di Maio [36] desenvolve uma ontologia como solução para troca de informações entre os diversos prestadores de serviços em situações de emergência e catástrofes médicas. São também analisadas as questões fundamentais dos sistemas inteligentes e de gestão de emergência médica, com foco nos projectos de código aberto que disponibilizam as respectivas ontologias.

Recentemente, Ferreira et al. (2008) desenvolveram o projecto MedAlert [22]. O MedAlert possui um *corpus* com cerca de 48.229 textos relativos a episódios de internamento ocorridos no Hospital Infante D. Pedro, em Aveiro. Utiliza como suporte ontologias e outras fontes de conhecimento que permitem extrair a informação dos textos médicos, de modo a inferir, de uma forma automática, irregularidades/dúvidas suscitadas pelas decisões tomadas pelos profissionais de saúde.

3.2 Criação de Ontologias em Relatórios Médicos

Actualmente, os relatórios médicos contêm informação para o cuidado de saúde do paciente, num texto livre não estruturado e muitas vezes de difícil compreensão. Têm-se desenvolvido esforços que levem ao aproveitamento da informação resultante dos relatórios médicos.

No trabalho desenvolvido por Annibal et al. (2008), é apresentada uma ontologia para estruturar a informação de Relatórios Radiológicos [1]. O objectivo deste trabalho é investigar o domínio léxico e semântico dos textos presentes em relatórios radiológicos, e desenvolver uma ontologia que permita o armazenamento destes documentos de forma estruturada, sem interferir no modo como os utilizadores interagem com um sistema já existente.

O sistema corresponde ao Sistema de Informação Radiológica (SIR) do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, sistema esse que é responsável pela informatização de vários processos, entre eles o de permitir criar relatórios resultantes da análise de imagens digitais. É nesta base que foi desenvolvido este trabalho, uma vez que o problema actual resulta do não aproveitamento das funcionalidades do sistema, pois a informação dos cerca de 660.000 relatórios radiológicos é guardada em simples campos textuais e em linguagem natural, e não de forma estruturada e consequentemente pouco adequada às ferramentas computacionais voltadas para a extracção e análise do conhecimento.

Como *corpus* para o desenvolvimento do trabalho, foram tidos em conta relatórios a exames de Ressonância Magnética ao Joelho, exames de Tomografia Computadorizada ao Crânio e exames de Ultra-sonografia ao Rim. A figura 3.1, corresponde a um exemplo de um relatório radiológico, neste caso de uma Ressonância Magnética com os campos e respectiva informação.

A primeira parte do trabalho corresponde à análise dos relatórios do *corpus*. A partir desta análise foi possível criar estruturas hierárquicas das entidades anatómicas examinadas, das modalidades de exame, observações, regiões anatómicas entre outros elementos.

Nome da região	Nome do exame	Descrição do exame	Conclusão do laudo do exame
Joelho Direito	Ressonância Magnética	<p>Estruturas ósseas regionais com morfologia e intensidade de sinal preservados.</p> <p>Cartilagem articular com espessura e sinal conservados.</p> <p>Tendões patelar e do quadríceps e demais estruturas do aparelho extensor sem alterações.</p> <p>Ligamentos cruzados anterior e posterior, bem como os colaterais lateral e medial com forma, orientação e intensidade de sinal habituais.</p> <p>Meniscos lateral, com forma, contornos e intensidade de sinal normais.</p> <p>Laceração na transição corpo-corno posterior do menisco medial, com deslocamento da porção central para incisura intercondilar, porção periferia de aspecto irregular.</p> <p>Pequena quantidade de líquido livre intrarticular.</p> <p>Observa-se distensão da bursa gastrocnêmio-semimembranosa medial por coleção líquida.</p>	<p>1. Ruptura do menisco medial (alça de balde)</p> <p>2. Cisto de baker.</p> <p>3. Pequeno derrame articular</p>

Figura 3.1: Exemplo de um Relatório Radiológico [1]

Após a análise dos relatórios foi então possível definir o domínio da ontologia, que foi criada através da ferramenta Protégé (descrita na secção 2.2.5). Para instanciar a ontologia, foram utilizadas ferramentas de mineração, de modo a realizar um pré-processamento do texto contido nos relatórios e de forma a identificar cada palavra e a categorizá-la entre as entidades da ontologia.

A figura 3.2 representa parte da ontologia.

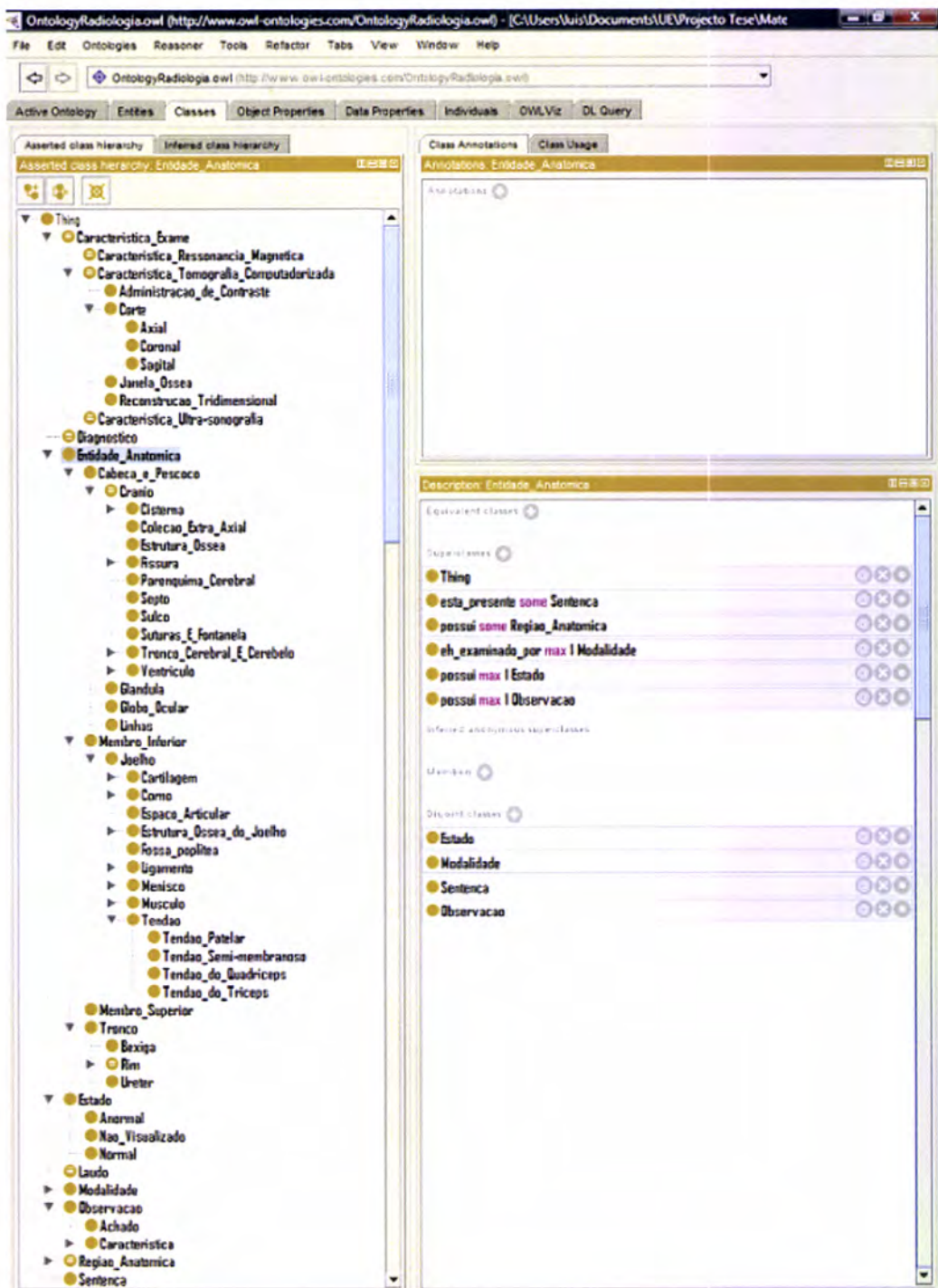


Figura 3.2: Ontologia de Relatórios Radiológicos

A ontologia possui os seguintes relacionamentos entre entidades:

- Característica da Modalidade *eh_caracteristica_de* Modalidade;
- Diagnóstico *refere-se* a um Laudo e *possui* algumas Sentenças;

- Entidade Anatômica *eh_examinada_por* uma Modalidade, *esta_presente* em algumas Sentenças, *possui* um Estado, *possui* algumas Regiões Anatômicas e *possui* uma Observação;
- Estado *esta_presente* em algumas Sentenças e *refere-se* a algumas Entidades Anatômicas;
- Modalidade *eh_caracterizada_por* algumas Características de Modalidade, *examina* algumas Entidades Anatômicas e *gera* um Laudo;
- Laudo *possui* algumas Sentenças e *possui* um Diagnóstico;
- Observação *esta_presente* em algumas Sentenças e *refere-se* a algumas Entidades Anatômicas;
- Região Anatômica *refere-se* a algumas Entidades Anatômicas e *esta_presente* em algumas Sentenças;
- Sentença *refere-se* a um Diagnóstico, *refere-se* a um Laudo, *possui* um Estado, *possui* uma Observação, *possui* uma Entidade Anatômica e *possui* uma Região Anatômica.

Este trabalho permitiu que no futuro o sistema disponibilize um ambiente computacional mais evoluído, no qual o radiologista continuará a preencher os relatórios na forma de texto livre como até aqui tinha feito, no entanto, as informações contidas nos relatórios são agora guardadas de forma estruturada, formando uma base de conhecimento e permitindo consultas de modo automático e perceptível às máquinas.

Capítulo 4

Proposta de Arquitectura

Este capítulo faz uma descrição geral do trabalho proposto.

A secção 4.1 descreve quais os objectivos do trabalho (4.1.1) e qual a metodologia a seguir (4.1.2). A secção 4.2 descreve a arquitectura, e (4.2.1) a plataforma em que o trabalho foi desenvolvido e o que foi utilizado com vista a atingir os objectivos propostos.

Os capítulos seguintes, descrevem detalhadamente os procedimentos que foram adoptados em cada fase do trabalho.

4.1 Apresentação

Cada vez mais existe uma grande aposta na área da EI, sobretudo em criar sistemas dinâmicos capazes de extrair informação proveniente de diversas fontes. Estes sistemas não se têm resumido a extrair a informação para algo específico e fechado a outras pessoas. Existe o interesse e a necessidade de cooperação destes sistemas, de modo a dar a conhecer a sua base de conhecimento, quer para consulta quer para uso no desenvolvimento de novas ferramentas. Para tal, existe a necessidade de publicar a sua informação.

A recente área da Web Semântica, tem provocado um grande interesse por parte das pessoas dos mais variados sectores, entre eles o da medicina, principalmente devido à possibilidade de usar linguagens capazes de estruturar a informação e aplicações capazes de processar essa informação de modo automático.

A grande quantidade de informação existente na área médica, tem levado nos últimos anos à necessidade de aproveitar essa informação. Os procedimentos médicos, mais concretamente os relatórios médicos, são na maior parte das vezes guardados como texto livre não estruturado, e o seu aproveitamento para uso futuro é praticamente nulo.

Uma das formas de estruturar esta informação é através do uso de ontologias, que surgiu como um modo eficaz de publicar informação de modo estruturado, e portanto, são utilizadas neste trabalho como a estrutura que representa a informação.

4.1.1 Objectivos

A partir do estudo das ontologias existentes, verificou-se a quase inexistência de ontologias relacionadas com procedimentos médicos, que possam facilitar a criação de estruturas e de bases de conhecimento. Como tal, o grande objectivo deste trabalho é apresentar uma nova ontologia capaz de representar relatórios médicos, e criar mecanismos automáticos para extrair a sua informação e povoar a referida ontologia.

O uso de ontologias em relatórios médicos é um auxílio importante a médicos no apoio à decisão diagnóstica, para os cuidados de saúde do paciente ou mesmo para fins de investigação.

4.1.2 Metodologia

O trabalho é desenvolvido em quatro fases, como mostra a figura 4.1.

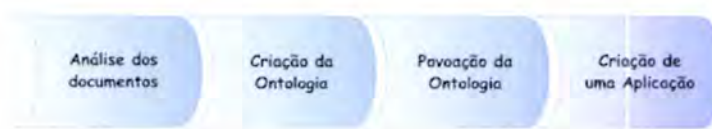


Figura 4.1: Fases de Desenvolvimento do Trabalho

E que corresponde respectivamente a:

1. Análise dos documentos (relatórios médicos)
2. Criação da Ontologia
3. Extracção automática de informação e respectiva povoação da ontologia
4. Criação de uma Aplicação

A primeira fase consiste em analisar o texto dos documentos de modo a extrair-se toda a informação que se vai utilizar na fase seguinte. Objectivamente, é necessário percorrer todos os documentos de forma a ser possível representar todo o conhecimento deste domínio, ou seja, o domínio de todos os relatórios médicos do *corpus*.

A segunda fase consiste em criar a ontologia na linguagem OWL, a partir da base de conhecimento adquirida na fase anterior. São definidas todas as classes, propriedades e restrições da ontologia.

A terceira fase consiste em extrair a informação dos documentos e povoar a ontologia criada na fase anterior, através do Processamento de Linguagem Natural e de técnicas de Machine Learning. É criado automaticamente um novo OWL da

ontologia com as instâncias criadas.

Por fim, é desenvolvida uma aplicação com o intuito de permitir ao utilizador fazer o input dos relatórios médicos que quer transformar na linguagem OWL, partindo sempre da ontologia base desenvolvida na segunda fase.

4.2 Arquitectura

A figura 4.2 representa um esboço da arquitectura do trabalho, que é constituída por vários módulos, onde cada um deles tem uma função específica, contribuindo para o sistema global.

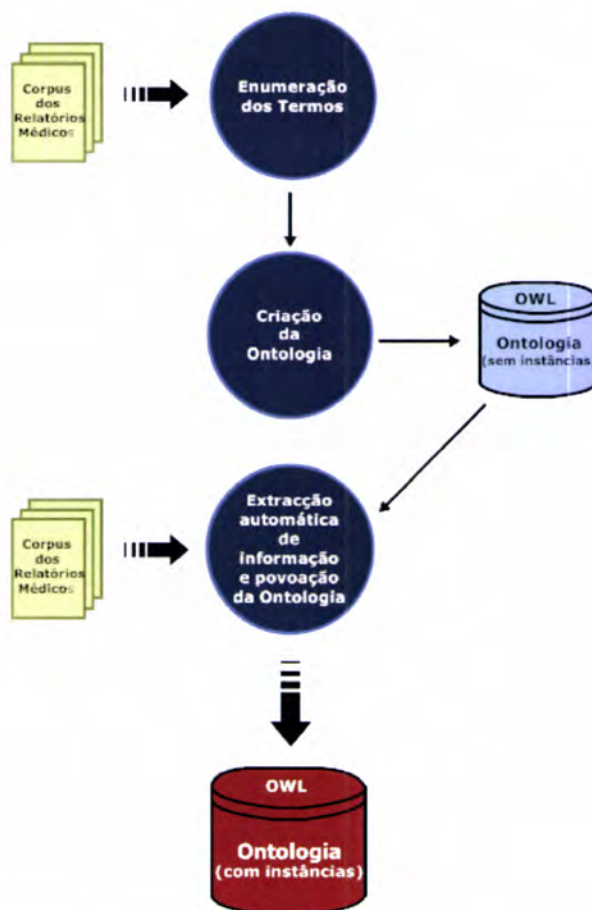


Figura 4.2: Arquitectura

O módulo correspondente à Enumeração dos Termos, tem como função identificar todos os termos dos relatórios médicos. Estes termos são descritos na secção 5.2.

A partir dos termos identificados no módulo anterior, surge o módulo da Criação da Ontologia, onde é criada a ontologia que é constituída por duas partes. A primeira corresponde à criação das classes que são descritas na secção 6.2, e a segunda corresponde à criação dos relacionamentos e das restrições sobre as classes que são descritos na secção 6.3.

Por fim, o último módulo é constituído por vários sub módulos que fornecem um sistema de extracção de informação, que permite numa primeira fase tratar a informação contida nos relatórios preenchendo alguns campos vazios (secção 7.2) e criar uma base de dados com essa informação (secção 7.3), e numa segunda fase processar essa informação através de Árvores de Decisão e de alguns algoritmos que foram desenvolvidos (secção 7.4), permitindo assim instanciar a ontologia e obter uma base de conhecimento com informação estruturada.

4.2.1 Plataforma

Todo o trabalho foi desenvolvido utilizando a plataforma Linux.

A criação da ontologia e consecutivamente a produção do código OWL é feita através de uma aplicação desenvolvida na linguagem Java, o Protégé¹. A ontologia é publicada através de um web server, que neste caso é o Apache².

Para guardar alguma informação dos documentos, é utilizada a base de dados mysql³, que funciona como suporte à fase da extracção de informação (EI).

Na manipulação da ontologia e na criação das instâncias, foi utilizado uma ferramenta já existente e desenvolvida em Java, o Jena⁴ (parser do OWL).

Para o desenvolvimento da interface da aplicação, foi utilizado a linguagem Java, particularmente o package javax.swing.

¹<http://protege.stanford.edu/>

²<http://httpd.apache.org/>

³<http://www.mysql.com/>

⁴<http://jena.sourceforge.net/index.html>

Capítulo 5

Análise dos documentos

Este capítulo descreve a primeira parte do trabalho que corresponde à análise dos documentos utilizados no seu desenvolvimento.

A secção 5.1 explica a proveniência dos documentos fazendo uma descrição dos mesmos (5.1.1). Na secção 5.2 são enumerados todos os termos identificados, podendo esses termos ser específicos dos documentos (5.2.1), ou serem termos médicos (5.2.2).

5.1 Documentos em análise

Antes de se começar a desenvolver o trabalho, é necessário obter algum tipo de documento que represente um procedimento médico. Para tal, entrou-se em contacto com o director clínico do HESE (Hospital do Espírito Santo de Évora), que desde logo se prontificou a ajudar e disponibilizou um conjunto de relatórios médicos.

5.1.1 Descrição dos documentos

Este conjunto de relatórios médicos, mais concretamente relatórios neurovasculares (ver exemplo no Anexo A), encontram-se em formato digital, e são digitalizados com OCR (*Optical Character Recognition*) e guardados no formato TXT.

Os relatórios possuem cinco zonas bem distintas:

1. Identificação do Paciente
2. Informação Clínica
3. Exames realizados¹ (tipo de exame, aparelho e janelas ultra-sónicas caso existam) e respectiva descrição

¹cada relatório pode conter um ou vários exames

4. Conclusão do relatório
5. Identificação do médico(s) e técnico(s)

Observa-se que na zona da identificação do paciente como mostra a figura 5.1, os campos não se encontram preenchidos de forma a proteger a identidade dos pacientes.

IDENTIFICAÇÃO DO PACIENTE:

Nº de Processo

Data do exame

Nome

Idade

Serviço

Figura 5.1: Identificação do Paciente nos relatórios

O conjunto destes relatórios neurovasculares serviu de *corpus* do trabalho apresentado. No entanto, qualquer relatório neurovascular com uma base de conhecimento igual aos relatórios do *corpus*, pode ser usado desde que possua as mesmas zonas e o mesmo formato.

5.2 Enumeração dos termos

O *corpus* do trabalho, formado por um conjunto de relatórios neurovasculares, possui uma base de conhecimento que corresponde ao domínio que queremos representar. Assim, é necessário percorrer todo o domínio e extrair todos os termos existentes, de modo a conhecermos a base de conhecimento e de ser possível construir uma ontologia para este domínio (capítulo 6).

5.2.1 Termos específicos dos documentos

Os termos específicos dos documentos, são aqueles que se encontram em todos os relatórios do *corpus* independentemente da informação médica que estes possuem, e da conclusão que possam ter.

A tabela 5.1 contém os termos existentes no cabeçalho dos relatórios, e que correspondem ao tipo de relatório e ao local onde o relatório é realizado.

Termo
Laboratório de Ultrassons Cardio e Neurovascular
Relatório Neurovascular

Tabela 5.1: Termos do cabeçalho dos relatórios

A tabela 5.2 possui os termos correspondentes à zona da identificação do paciente, e que correspondem ao número de processo, nome, idade e informação clínica do paciente. Possui também um campo com o termo *Serviço* que tem a identificação do serviço hospitalar caso o paciente se encontre internado. Além disso, possui ainda a data cujo os exames pertencentes ao relatório foram efectuados.

Termo
Número de Processo
Nome
Idade
Serviço
Data de exame
Informação Clínica

Tabela 5.2: Termos da Identificação do Paciente

A tabela 5.3 possui os termos correspondentes ao cabeçalho dos exames incluídos nos relatórios do domínio. Para qualquer exame, é referido o tipo de exame realizado e o aparelho usado tal como a sua marca e transdutor. Dependendo do tipo de exame, são ainda identificados três termos referentes a janelas ultra-sônicas utilizadas no exame (*transtemporal*, *transorbital* e *suboccipital*).

Termo
Triplex Scan Cervical
Triplex Scan Transcraneano
Ecografo
Marca
Transdutor
Transtemporal
Transorbital
Suboccipital

Tabela 5.3: Termos do cabeçalho dos exames

A tabela 5.4 refere-se aos termos existentes no rodapé dos relatórios e que corresponde ao médico que realizou o relatório e à técnica que executou o(s) exame(s).

Termo
Médico
Técnica

Tabela 5.4: Termos do rodapé dos relatórios

5.2.2 Termos médicos

Os termos médicos são aqueles que são usados quer na descrição dos conteúdos do(s) exame(s) incluídos nos relatórios, quer na descrição da conclusão. Entende-se por termos médicos, as entidades anatómicas e tudo o que possa estar relacionado com estas, e ainda observações e diagnósticos.

A tabela 5.5 possui os termos das entidades anatómicas identificadas no domínio dos relatórios. Algumas destas entidades anatómicas aparecem no texto também como abreviaturas, e como tal na tabela, cada termo referente a uma entidade anatómica possui ainda uma coluna com a respectiva abreviatura.

Termo			Abreviatura
Artéria Subclávia			—
Artéria Vertebral	V0		AV0
	V1		AV1
	V2		AV2
	V3		AV3
	V4		AV4
Artéria Carótida ou Carótida Comum	Externa		ACE
	Interna		ACI
Artéria Cerebral	Anterior		ACA
	Comunicante Anterior		ACoA
	Posterior	P1	ACP1
		P2	ACP2
	Média		ACM
Artéria Basilar			AB
Artéria Central da Retina/Oftálmica			—

Tabela 5.5: Termos das Entidades Anatômicas

A tabela 5.6 possui os termos que podem estar relacionados com as entidades anatómicas e que se encontram representados nos relatórios. Para uma mais fácil compreensão, existe também uma coluna com a descrição do relacionamento de cada um destes termos com as entidades anatómicas.

Termo		Descrição
Eixos Carotídeos		Conjunto das Artérias Carótidas
Segmentos		Corresponde às Artérias Vertebrais
Paredes Arteriais		Paredes das artérias
Bifurcações Carotídeas		Onde se dividem as Artérias Carótidas
Stent		Mecanismo que é implantado nas artérias e é usado para desentupi-las
Lado	Direito Esquerdo	As artérias podem ser constituídas por lado direito e lado esquerdo
Zona	Montante Jusante	Referente a uma zona de uma entidade anatómica

Tabela 5.6: Termos relacionados com as Entidades Anatômicas

A tabela 5.7 possui os termos com as observações identificadas no domínio. Na grande maioria das vezes, verifica-se que estas ocorrem a partir de uma ou várias entidades anatómicas. O termo *IPs* corresponde aos Índices de Pulsatilidade.

Termo	
Permeável/Permeabilidade	
Fluxo	Normal
	Ausência
	Aumento da velocidade
	Reabitação
	Inversão do sentido
Fenómeno Roubo da Subclávia	
Ateromatose Carotídea	
Espessamento Ateromatoso Difuso	
Impossível Visualizar	
Aplanamento do Espectro de Doppler	
Hipoplasia	
Tortuosidade	
IPs aumentados	

Tabela 5.7: Termos das Observações

A tabela 5.8 possui os termos dos diagnósticos identificados no domínio. Verifica-se que na grande maioria das vezes, os diagnósticos são indicados a partir das observações realizadas.

Termo	
Sem Repercussão Hemodinâmica	
Estenose	Ligeira
	Moderada
	Grave
	Residual
	Remodelação
	Pré-oclusiva
Inelastecidade Arterial	
Circulação Colateral	
Oclusão	
Sem seguro significado patológico	

Tabela 5.8: Termos dos Diagnósticos

Capítulo 6

Criação da Ontologia

Este capítulo descreve a segunda parte do trabalho que corresponde à criação da ontologia relacionada com os documentos.

Na secção 6.1 indica-se qual a linguagem *Web* escolhida (6.1.1) descrevendo alguma da sua sintaxe. A criação das classes da ontologia é descrita na secção 6.2. A secção 6.3 corresponde à criação dos relacionamentos e das restrições entre as classes criadas na secção 6.2.

6.1 Ontologia e Web Semântica

Para publicar os relatórios médicos do nosso *corpus* na Web Semântica, é necessário adicionar anotações semânticas que representem a informação contida nos relatórios, de modo a que as máquinas consigam processar automaticamente essa informação. Uma das formas de representar esta informação, é através do uso de Ontologias.

O uso de Ontologias e das suas diversas linguagens, além de permitirem estruturar a informação e representar todo o conhecimento dos relatórios, permitem seguir as regras e as normas de publicação dos dados na Web Semântica, de modo a que o formato dos dados seja sempre o mesmo.

6.1.1 Escolha da Linguagem

Entre as várias linguagens existentes para representar Ontologias, a linguagem OWL foi a escolhida, pois é mais eficiente na *Web* do que XML, RDF e RDFS, e é hoje em dia uma recomendação da W3C, principalmente devido à sua semântica formal capaz de processar conteúdo semântico na *Web*.

A linguagem OWL, que é uma extensão de RDF, permite incluir a descrição das classes, as suas propriedades e respectivos relacionamentos.

Antes de mais, é necessário descrever alguma sintaxe do OWL, de modo a se conseguir perceber o código OWL que irá ser exposto nas secções seguintes. A tabela 6.1 descreve essa sintaxe.

Sintaxe	Descrição
Class	Define uma classe
DatatypeProperty	Atributos de uma classe
disjointWith	Classe disjunta a outra classe
Expressões entre aspas iniciadas por #	URIs que referem objectos no próprio ficheiro OWL em que a classe é definida
Expressões entre aspas iniciadas por &	URI abreviado em que se usa a XML ENTITY
maxQualifiedCardinality	Máximo de cardinalidade de uma propriedade em relação a uma classe
minQualifiedCardinality	Mínimo de cardinalidade de uma propriedade em relação a uma classe
ObjectProperty	Relações entre classes
onClass	Classe a que uma restrição se refere
onProperty	Refere a propriedade da restrição
Restriction	Restrição aplicada a uma classe
someValuesFrom	Refere que uma propriedade tem pelo menos um ou mais valores em relação a uma classe

Tabela 6.1: Descrição de alguma sintaxe do OWL

Os *headers* com a respectiva definição podem ser consultados na versão integral da ontologia, que consta no Apêndice A deste documento.

6.2 Criação das Classes

O primeiro passo na criação da ontologia do trabalho, corresponde à criação das classes a partir dos termos identificados na análise dos documentos do domínio. Estes termos foram enumerados no capítulo 5, mais concretamente na secção 5.2.

Nota: Na linguagem OWL, qualquer atributo ou relação de uma classe, é também atributo ou relação das suas subclasses.

Alguns dos excertos de código OWL que são apresentados nas secções seguintes, não se encontram completos de modo a não ficarem muito extensos nem muito repetitivos, pois grande parte do código é muito semelhante.

6.2.1 Classes dos Termos Específicos

A partir dos termos do cabeçalho dos relatórios (tabela 5.1), define-se a classe *Relatório* que tem como subclasse a classe *Neurovascular*, e que possui como atri-

butos o *Local* onde o relatório é realizado, a sua *Data* e o *Número de Relatório*. A figura 6.1 representa a hierarquia das classes para este conjunto de termos.

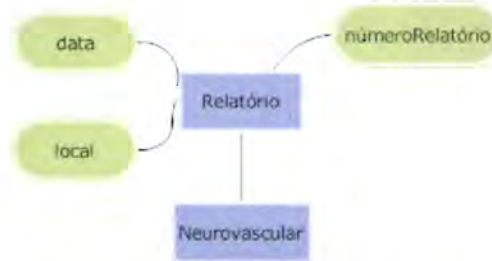


Figura 6.1: Hierarquia das classes para os termos do cabeçalho dos relatórios

Na linguagem OWL, o código correspondente a esta hierarquia é o seguinte:

```

<owl:Class rdf:about="#Relatório">
  <rdfs:comment
    >Relatório Neurovascular</rdfs:comment>
</owl:Class>

    <owl:Class rdf:about="#Neurovascular">
      <rdfs:subClassOf rdf:resource="#Relatório"/>
    </owl:Class>

  <owl:DatatypeProperty rdf:about="#data">
    <rdfs:comment
      >Data do Relatório</rdfs:comment>
    <rdfs:domain rdf:resource="#Relatório"/>
    <rdfs:range rdf:resource="&xsd:date"/>
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="#local">
    <rdfs:comment
      >Local onde se realiza o Relatório</rdfs:comment>
    <rdfs:domain rdf:resource="#Relatório"/>
    <rdfs:range rdf:resource="&xsd:string"/>
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="#númeroRelatório">
    <rdfs:comment
      >Número do Relatório</rdfs:comment>
    <rdfs:domain rdf:resource="#Relatório"/>
    <rdfs:range rdf:resource="&xsd:integer"/>
  </owl:DatatypeProperty>

```

Aos termos correspondentes à identificação do paciente (tabela 5.2), são criadas três hierarquias diferentes, e que estão representadas na figura 6.2.



Figura 6.2: Hierarquia das classes para os termos da identificação do paciente

Assim, é criada a classe *Paciente* que tem como atributos o *Nome do Paciente*, a *Idade do Paciente* e o *Número de Processo do Paciente*. É criada também a classe *Serviço* que tem como atributo o *Nome do Serviço*. Por fim, é criada a classe *Informação Clínica* que tem como atributo a sua *Descrição*.

Na linguagem OWL, o código correspondente a estas hierarquias é o seguinte:

```

<owl:Class rdf:about="#Paciente">
  <rdfs:comment
    >Paciente sobre o qual se realiza o Relatório</rdfs:comment>
</owl:Class>

  <owl:DatatypeProperty rdf:about="#nomePaciente">
    <rdfs:comment
      >Nome do Paciente</rdfs:comment>
    <rdfs:domain rdf:resource="#Paciente"/>
    <rdfs:range rdf:resource="&xsd:string"/>
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="#idadePaciente">
    <rdfs:comment
      >Idade do Paciente</rdfs:comment>
    <rdfs:domain rdf:resource="#Paciente"/>
    <rdfs:range rdf:resource="&xsd:integer"/>
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="#númeroProcessoPaciente">
    <rdfs:comment
      >Número do Processo do Paciente</rdfs:comment>
    <rdfs:domain rdf:resource="#Paciente"/>
    <rdfs:range rdf:resource="&xsd:integer"/>
  </owl:DatatypeProperty>

<owl:Class rdf:about="#Serviço">
  <rdfs:comment
    >Serviço onde está internado o Paciente</rdfs:comment>
</owl:Class>

```



```

<owl:DatatypeProperty rdf:about="#nomeServiço">
  <rdfs:comment
    >Nome do Serviço</rdfs:comment>
  <rdfs:domain rdf:resource="#Serviço"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

<owl:Class rdf:about="#Informação_Clínica">
  <rdfs:comment
    >Informação Clínica do Paciente</rdfs:comment>
</owl:Class>

<owl:DatatypeProperty rdf:about="#descriçãoInformaçãoClínica">
  <rdfs:comment
    >Descrição da Informação Clínica do Paciente</rdfs:comment>
  <rdfs:domain rdf:resource="#Informação_Clínica"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

```

Em relação aos termos do cabeçalho dos exames (tabela 5.3), são criadas também três hierarquias de classes, que estão representadas na figura 6.3.



Figura 6.3: Hierarquia das classes para os termos do cabeçalho dos exames

Deste modo, é criada uma classe *Exame* que tem como subclasses a classe *Ultra-Sonografia*, classe esta que tem como subclasse a classe *Eco-Doppler colorido*, que por fim tem as subclasses *Triplex Scan Cervical* e *Triplex Scan Transcraneano*. Também é criada a classe *Aparelho* que tem como subclasse a classe *Ecografo*, que contém como atributos a *Marca* e o *Transdutor*. Por fim é criada a classe *Janela Ultra-Sônica*, que tem como subclasses as classes *Transtemporal*, *Suboccipital* e *Transorbital*.

Na linguagem OWL, o código correspondente a estas hierarquias é o seguinte:

```

<owl:Class rdf:about="#Exame">
  <rdfs:comment
    >Exames do Relatório</rdfs:comment>
</owl:Class>

  <owl:Class rdf:about="#Ultra-Sonografia">
    <rdfs:subClassOf rdf:resource="#Exame"/>
  </owl:Class>

  <owl:Class rdf:about="#Eco-Doppler_colorido">
    <rdfs:subClassOf rdf:resource="#Ultra-Sonografia"/>
  </owl:Class>

  <owl:Class rdf:about="#Triplex_Scan_Cervical">
    <rdfs:subClassOf rdf:resource="#Eco-Doppler_colorido"/>
    <owl:disjointWith rdf:resource="#Triplex_Scan_Transcraneano"/>
  </owl:Class>

  <owl:Class rdf:about="#Triplex_Scan_Transcraneano">
    <rdfs:subClassOf rdf:resource="#Eco-Doppler_colorido"/>
  </owl:Class>

<owl:Class rdf:about="#Aparelho">
  <rdfs:comment
    >Aparelho usado para realizar os exames</rdfs:comment>
</owl:Class>

  <owl:Class rdf:about="#Ecografo">
    <rdfs:subClassOf rdf:resource="#Aparelho"/>
  </owl:Class>

  <owl:DatatypeProperty rdf:about="#marca">
    <rdfs:comment
      >Marca do Aparelho usado no Exame</rdfs:comment>
    <rdfs:domain rdf:resource="#Aparelho"/>
    <rdfs:range rdf:resource="&xsd:string"/>
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="#transdutor">
    <rdfs:comment
      >Transdutor usado no Aparelho</rdfs:comment>
    <rdfs:domain rdf:resource="#Aparelho"/>
    <rdfs:range rdf:resource="&xsd:string"/>
  </owl:DatatypeProperty>

<owl:Class rdf:about="#Janela_Ultra-Sônica">
  <rdfs:comment
    >Janelas Ultra-Sônicas usadas no Exame Transcraneano</rdfs:comment>
</owl:Class>

```

```

<owl:Class rdf:about="#Transtemporal">
  <rdfs:subClassOf rdf:resource="#Janela_Ultra-Sônica"/>
</owl:Class>

<owl:Class rdf:about="#Suboccipital">
  <rdfs:subClassOf rdf:resource="#Janela_Ultra-Sônica"/>
  <owl:disjointWith rdf:resource="#Transorbital"/>
  <owl:disjointWith rdf:resource="#Transtemporal"/>
</owl:Class>

<owl:Class rdf:about="#Transorbital">
  <rdfs:subClassOf rdf:resource="#Janela_Ultra-Sônica"/>
  <owl:disjointWith rdf:resource="#Transtemporal"/>
</owl:Class>

```

Por fim, em relação aos termos do rodapé dos relatórios (tabela 5.4), foram criadas duas hierarquias de classes, que estão representadas na figura 6.4.



Figura 6.4: Hierarquia das classes para os termos do rodapé dos relatórios

É então criada a classe *Médico* que possui como atributo o *Nome do Médico*. E é ainda criada a classe *Técnico* que possui como atributo o *Nome do Técnico*.

Na linguagem OWL, o código correspondente a estas hierarquias é o seguinte:

```

<owl:Class rdf:about="#Médico">
  <rdfs:comment
    >Médico que realiza o Relatório</rdfs:comment>
</owl:Class>

  <owl:DatatypeProperty rdf:about="#nomeMédico">
    <rdfs:comment>Nome do Médico</rdfs:comment>
    <rdfs:domain rdf:resource="#Médico"/>
    <rdfs:range rdf:resource="&xsd:string"/>
  </owl:DatatypeProperty>

<owl:Class rdf:about="#Técnico">
  <rdfs:comment
    >Técnico que executa os Exames</rdfs:comment>
</owl:Class>

```



```

<owl:DatatypeProperty rdf:about="#nomeTécnico">
  <rdfs:comment>Nome do Técnico</rdfs:comment>
  <rdfs:domain rdf:resource="#Técnico"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

```

6.2.2 Classes dos Termos Médicos

A partir dos termos das Entidades Anatômicas (tabela 5.5), é criada a hierarquia de classes representada na figura 6.5.



Figura 6.5: Hierarquia das classes para os termos das Entidades Anatômicas

A classe *Entidade Anatômica* está no topo da hierarquia e tem como subclasses a classe *Cabeça e Pescoço*, que tem como subclasses a classe *Vaso Sanguíneo* e que tem como subclasse a classe *Artéria* que tem como atributos os respectivos *CUI* (*Concept Unique Identifier*) e *CN* (*Concept Name*). Esta classe subdivide-se em várias subclasses que são a classe *Basilar*, *Carótida ou Carótida Comum*, *Central da Retina/Oftálmica*, *Cerebral*, *Subclávia* e *Vertebral*. Algumas destas classes ainda se subdividem em outras subclasses, como é o caso da classe *Carótida ou Carótida Comum* que se subdivide nas subclasses *Externa* e *Interna*. A classe *Cerebral* subdivide-se nas subclasses *Anterior*, *Comunicante Anterior*, *Posterior* que tem como subclasses a classe *P1* e *P2*, e ainda a subclasse *Média*. Por fim, a classe *Vertebral* subdivide-se nas subclasses *V0*, *V1*, *V2*, *V3* e *V4*.

Na linguagem OWL, o código correspondente a esta hierarquia é o seguinte:

```

<owl:Class rdf:about="#Entidade_Anatómica">
  <rdfs:comment
    >Entidades Anatómicas analisadas em Relatórios Neurovasculares
  </rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Cabeça_e_Pescoço">
  <rdfs:subClassOf rdf:resource="#Entidade_Anatómica"/>
</owl:Class>

<owl:Class rdf:about="#Vaso_Sanguíneo">
  <rdfs:subClassOf rdf:resource="#Cabeça_e_Pescoço"/>
</owl:Class>

<owl:Class rdf:about="#Artéria">
  <rdfs:subClassOf rdf:resource="#Vaso_Sanguíneo"/>
</owl:Class>

  <owl:DatatypeProperty rdf:about="#arteriaCUI">
    <rdfs:comment
      >Código CUI do UMLS para a Artéria</rdfs:comment>
    <rdfs:domain rdf:resource="#Artéria"/>
    <rdfs:range rdf:resource="&xsd:string"/>
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="#arteriaCN">
    <rdfs:comment
      >Nome CN do UMLS para a Artéria</rdfs:comment>
    <rdfs:domain rdf:resource="#Artéria"/>
    <rdfs:range rdf:resource="&xsd:string"/>
  </owl:DatatypeProperty>

<owl:Class rdf:about="#Central_da_Retina/Oftálmica">
  <rdfs:subClassOf rdf:resource="#Artéria"/>
  <owl:disjointWith rdf:resource="#Cerebral"/>
  <owl:disjointWith rdf:resource="#Subclávia"/>
  <owl:disjointWith rdf:resource="#Vertebral"/>
</owl:Class>

(...)

<owl:Class rdf:about="#Subclávia">
  <rdfs:subClassOf rdf:resource="#Artéria"/>
</owl:Class>

<owl:Class rdf:about="#Vertebral">
  <rdfs:subClassOf rdf:resource="#Artéria"/>
</owl:Class>

  (...)

```


Em relação ao conjunto de termos relacionados com as Entidades Anatômicas (tabela 5.6), foram criadas as hierarquias de classes que estão representadas na figura 6.6.



Figura 6.6: Hierarquia das classes para os termos relacionados com as Entidades Anatômicas

Esta hierarquia de classes é criada pelas classes *Parede Arterial*, *Bifurcação* e *Stent* que possuem como atributos os respectivos *CUI* e *CN*. É também criada a classe *Lado* que é subdividida em duas subclasses, a classe *Esquerda* e *Direita*. Por fim é criada a classe *Zona* que também é subdividida em duas subclasses, a classe *Montante* e *Jusante*.

Na linguagem OWL, o código correspondente a estas hierarquias é o seguinte:

```

<owl:Class rdf:about="#Parede_Arterial">
  <rdfs:comment
    >Parede Arterial das artérias</rdfs:comment>
</owl:Class>

<owl:DatatypeProperty rdf:about="#paredeArterialCN">
  <rdfs:comment
    >Nome CN do UMLS para a Parede Arterial</rdfs:comment>
  <rdfs:domain rdf:resource="#Parede_Arterial"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#paredeArterialCUI">
  <rdfs:comment
    >Código CUI do UMLS para a Parede Arterial</rdfs:comment>
  <rdfs:domain rdf:resource="#Parede_Arterial"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>
  
```

```

<owl:Class rdf:about="#Bifurcação">
  <rdfs:comment
    >Bifurcações das Artérias</rdfs:comment>
</owl:Class>

  (...)

<owl:Class rdf:about="#Stent">
  <rdfs:comment
    >Mecanismo implantado nas artérias</rdfs:comment>
</owl:Class>

  (...)

<owl:Class rdf:about="#Lado">
  <rdfs:comment
    >Lado da artéria observada</rdfs:comment>
</owl:Class>

  <owl:Class rdf:about="#Esquerdo">
    <rdfs:subClassOf rdf:resource="#Lado"/>
  </owl:Class>

  <owl:Class rdf:about="#Direito">
    <rdfs:subClassOf rdf:resource="#Lado"/>
    <owl:disjointWith rdf:resource="#Esquerdo"/>
  </owl:Class>

<owl:Class rdf:about="#Zona">
  <rdfs:comment
    >Zona correspondente a uma Observação</rdfs:comment>
</owl:Class>

  <owl:Class rdf:about="#Montante">
    <rdfs:subClassOf rdf:resource="#Zona"/>
  </owl:Class>

  <owl:Class rdf:about="#Jusante">
    <rdfs:subClassOf rdf:resource="#Zona"/>
    <owl:disjointWith rdf:resource="#Montante"/>
  </owl:Class>

```

Em relação aos termos das Observações (tabela 5.7), é criada uma hierarquia de classes como está representada na figura 6.7.



Figura 6.7: Hierarquia das classes para os termos das Observações

A classe *Observação* está no topo da hierarquia tem como atributos os respectivos *CUI* e *CN*. Esta classe tem como subclasses todas as observações identificadas no domínio e que são *Permeável/Permeabilidade*, *Fluxo*, *Fenómeno*, *Placa Ateromatosa*, *Ateromatose Carotídea*, *Espessamento Ateromatoso Difuso*, *Impossível Visualizar*, *Aplanamento do espectro de Doppler*, *Hipoplasia*, *Tortuosidade* e *IP*. As classes *Fluxo*, *Fenómeno*, *Placa Ateromatosa* e *IP* subdividem-se em subclasses que representam os seus tipos.

Na linguagem OWL, o código correspondente a esta hierarquia é o seguinte:

```
<owl:Class rdf:about="#Observação">
  <rdfs:comment
    >Observações dos Exames</rdfs:comment>
</owl:Class>

<owl:DatatypeProperty rdf:about="#observaçãoCN">
  <rdfs:comment
    >Nome CN do UMLS para a Observação</rdfs:comment>
  <rdfs:domain rdf:resource="#Observação"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#observaçãoCUI">
  <rdfs:comment
    >Código CUI do UMLS para a Observação</rdfs:comment>
  <rdfs:domain rdf:resource="#Observação"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

<owl:Class rdf:about="#Permeável">
  <rdfs:subClassOf rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Placa_Ateromatosa"/>
  <owl:disjointWith rdf:resource="#Tortuosidade"/>
</owl:Class>
```



```

<owl:Class rdf:about="#Fluxo">
  <rdfs:subClassOf rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Hipoplasia"/>
  <owl:disjointWith rdf:resource="#IP"/>
  <owl:disjointWith rdf:resource="#Impossível_Visualizar"/>
  <owl:disjointWith rdf:resource="#Permeável"/>
  <owl:disjointWith rdf:resource="#Placa_Ateromatosa"/>
  <owl:disjointWith rdf:resource="#Tortuosidade"/>
</owl:Class>

(...)

<owl:Class rdf:about="#IP">
  <rdfs:subClassOf rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Impossível_Visualizar"/>
  <owl:disjointWith rdf:resource="#Permeável"/>
  <owl:disjointWith rdf:resource="#Placa_Ateromatosa"/>
  <owl:disjointWith rdf:resource="#Tortuosidade"/>
  <rdfs:comment
    >Índice de Pulsatilidade</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Aumentado">
  <rdfs:subClassOf rdf:resource="#IP"/>
</owl:Class>

```

Por fim, em relação aos termos dos Diagnósticos (tabela 5.8), é criada a hierarquia de classes representada na figura 6.8.

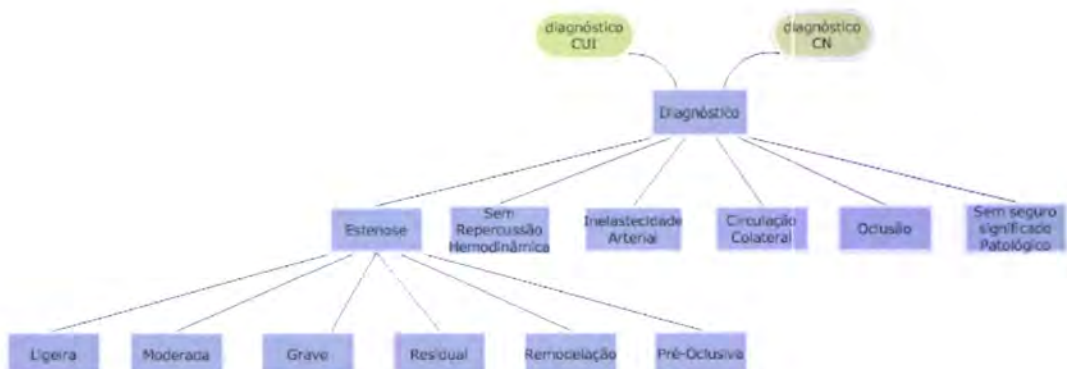


Figura 6.8: Hierarquia das classes para os termos dos Diagnósticos

A classe *Diagnóstico* está no topo da hierarquia tem como atributos os respectivos *CUI* e *CN*. Esta classe tem como subclasses todas os diagnósticos identificados

no domínio e que são *Estenose*, *Sem Repercussão Hemodinâmica*, *Inelastecidade Arterial*, *Circulação Colateral*, *Oclusão* e *Sem seguro significado patológico*. A classe *Estenose* possui algumas subclasses que representam os tipos de estenose existentes.

Na linguagem OWL, o código correspondente a esta hierarquia é o seguinte:

```
<owl:Class rdf:about="#Diagnóstico">
  <rdfs:comment
    >Diagnósticos das observações realizadas</rdfs:comment>
</owl:Class>

  <owl:DatatypeProperty rdf:about="#diagnósticoCN">
    <rdfs:comment
      >Nome CN do UMLS para o Diagnóstico</rdfs:comment>
    <rdfs:domain rdf:resource="#Diagnóstico"/>
    <rdfs:range rdf:resource="&xsd:string"/>
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="#diagnósticoCUI">
    <rdfs:comment
      >Código CUI do UMLS para o Diagnóstico</rdfs:comment>
    <rdfs:domain rdf:resource="#Diagnóstico"/>
    <rdfs:range rdf:resource="&xsd:string"/>
  </owl:DatatypeProperty>

<owl:Class rdf:about="#Estenose">
  <rdfs:subClassOf rdf:resource="#Diagnóstico"/>
  <owl:disjointWith rdf:resource="#Inelastecidade_Arterial"/>
  <owl:disjointWith rdf:resource="#Oclusão"/>
  <owl:disjointWith rdf:resource="#Sem_Repercussão_
    Hemodinâmica"/>
  <owl:disjointWith rdf:resource="#Sem_Seguro_Significado_
    Patológico"/>
</owl:Class>

  <owl:Class rdf:about="#Ligeira">
    <rdfs:subClassOf rdf:resource="#Estenose"/>
    <owl:disjointWith rdf:resource="#Moderada"/>
    <owl:disjointWith rdf:resource="#Pré-Oclusiva"/>
    <owl:disjointWith rdf:resource="#Remodelação"/>
    <owl:disjointWith rdf:resource="#Residual"/>
  </owl:Class>

  (...)

```



```

<owl:Class rdf:about="#Oclusão">
  <rdfs:subClassOf rdf:resource="#Diagnóstico"/>
  <owl:disjointWith rdf:resource="#Sem_Repercussão_
    Hemodinâmica"/>
  <owl:disjointWith rdf:resource="#Sem_Seguro_Significado_
    Patológico"/>
</owl:Class>

(...)

```

6.3 Criação dos Relacionamentos e das Restrições entre Classes

O segundo passo na criação da ontologia, corresponde à criação das relações entre classes e das respectivas restrições, a partir das classes criadas no passo anterior.

No trabalho, verifica-se dois tipos de relações.

- Relações entre classes que possuem objectos que estão relacionados com classes que possuem como objectos as *Artérias*;
- Relações entre classes que representam o esquema global do domínio.

A figura 6.9, representa o primeiro tipo de relações.

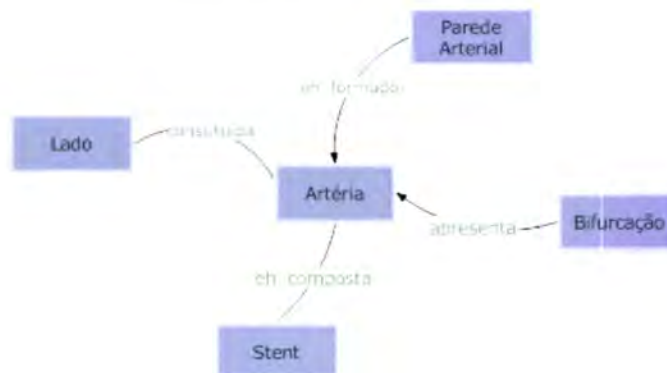


Figura 6.9: Relações com a classe que possui o objecto *Artéria*

A tabela 6.2 especifica este tipo de relações, acrescentado as restrições de cada uma delas.

Classe Domínio	Classe Intersecção	Relação	Restrição
Artéria	Parede Arterial	eh_formado	Artéria eh_formado por algumas Paredes Arteriais
Artéria	Bifurcação	apresenta	Artéria apresenta algumas Bifurcações
Artéria	Stent	eh_composta	Artéria eh_composta no máximo por um Stent
Artéria	Lado	constituída	Artéria constituída no máximo por dois Lados

Tabela 6.2: Especificação das Relações e Restrições da classe com o objecto *Artéria*

Na linguagem OWL, o código correspondente a estas relações é o seguinte:

```

<owl:ObjectProperty rdf:about="#eh_formado">
  <rdfs:domain rdf:resource="#Artéria"/>
  <rdfs:range rdf:resource="#Parede_Arterial"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#apresenta">
  <rdfs:domain rdf:resource="#Artéria"/>
  <rdfs:range rdf:resource="#Bifurcação"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#eh_composta">
  <rdfs:domain rdf:resource="#Artéria"/>
  <rdfs:range rdf:resource="#Stent"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#constituída">
  <rdfs:domain rdf:resource="#Artéria"/>
  <rdfs:range rdf:resource="#Lado"/>
</owl:ObjectProperty>

```

O código correspondente às restrições, pertence todo ele à classe que possui o objecto *Artéria*. Assim, a esta classe é adicionado o seguinte código OWL:

```

<rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#constituída"/>
    <owl:onClass rdf:resource="#Lado"/>
    <owl:maxQualifiedCardinality rdf:datatype=
      "&xsd;nonNegativeInteger">2</owl:maxQualifiedCardinality>
  </owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>

```

```

<owl:Restriction>
  <owl:onProperty rdf:resource="#eh_formado"/>
  <owl:someValuesFrom rdf:resource="#Parede_Arterial"/>
</owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#apresenta"/>
    <owl:someValuesFrom rdf:resource="#Bifurcação"/>
  </owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#eh_composta"/>
    <owl:onClass rdf:resource="#Stent"/>
    <owl:maxQualifiedCardinality rdf:datatype=
      "&xsd;nonNegativeInteger">1</owl:maxQualifiedCardinality>
  </owl:Restriction>
</rdfs:subClassOf>

```

O segundo tipo de relações, corresponde ao esquema global da ontologia onde o *Relatório* é o centro do esquema, como está representado na figura 6.10.

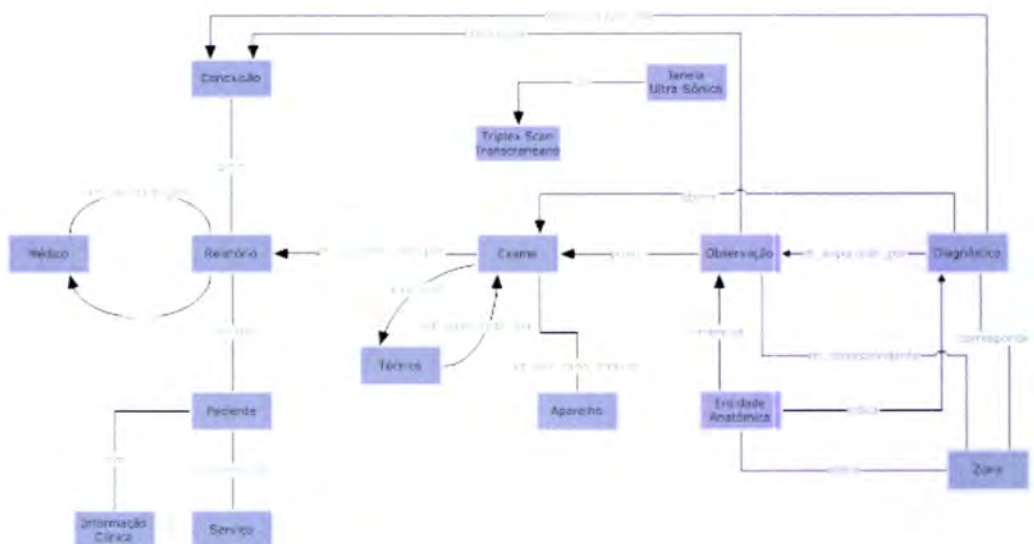


Figura 6.10: Relações entre classes que representam o esquema global do domínio

A tabela 6.3 especifica este tipo de relações, acrescentado as restrições de cada uma delas.

Classe Domínio	Classe Intersecção	Relação	Restrição
Relatório	Médico	eh_realizado_por	Relatório eh_realizado_por um Médico
Relatório	Paciente	contém	Relatório contém um Paciente
Relatório	Conclusão	gera	Relatório gera uma Conclusão
Relatório	Exame	eh_caracterizado_por	Relatório eh_caracterizado_por alguns Exames
Médico	Relatório	realiza	Médico realiza alguns Relatórios
Paciente	Informação Clínica	tem	Paciente tem uma Informação Clínica
Paciente	Serviço	encontra-se	Paciente encontra-se num Serviço
Conclusão	Observação	obtida_por	Conclusão obtida_por algumas Observações
Conclusão	Diagnóstico	obtida_atraves_de	Conclusão obtida_atraves_de alguns Diagnósticos
Exame	Técnico	eh_executado_por	Exame eh_executado_por alguns Técnicos
Exame	Aparelho	eh_realizado_atraves	Exame eh_realizado_atraves de um Aparelho
Exame	Observação	possui	Exame possui algumas Observações
Exame	Diagnóstico	obtem	Exame obtem alguns Diagnósticos
Triplex Scan Transcraneano	Janela Ultra-Sônica	usa	Triplex Scan Transcraneano usa algumas Janelas Ultra-Sônicas
Técnico	Exame	executa	Técnico executa alguns Exames
Observação	Zona	eh_correspondente	Observação eh_correspondente a uma Zona
Observação	Entidade Anatômica	refere-se	Observação refere-se a algumas Entidades Anatômicas
Diagnóstico	Observação	eh_suportado_por	Diagnóstico eh_suportado_por algumas Observações
Diagnóstico	Entidade Anatômica	indica	Diagnóstico indica algumas Entidades Anatômicas
Diagnóstico	Zona	corresponde	Diagnóstico corresponde a uma Zona
Zona	Entidade Anatômica	refere	Zona refere uma Entidade Anatômica

Tabela 6.3: Especificação das Relações e Restrições do esquema global do domínio

Na linguagem OWL, o código correspondente a estas relações é o seguinte:

```

<owl:ObjectProperty rdf:about="#eh_realizado_por">
  <rdfs:range rdf:resource="#Médico"/>
  <rdfs:domain rdf:resource="#Relatório"/>
</owl:ObjectProperty>

(...)

<owl:ObjectProperty rdf:about="#eh_caracterizado_por">
  <rdfs:range rdf:resource="#Exame"/>
  <rdfs:domain rdf:resource="#Relatório"/>
</owl:ObjectProperty>

(...)

<owl:ObjectProperty rdf:about="#indica">
  <rdfs:domain rdf:resource="#Diagnóstico"/>
  <rdfs:range rdf:resource="#Entidade_Anatómica"/>
</owl:ObjectProperty>

(...)

```

Na linguagem OWL, o código correspondente às restrições é acrescentado sempre na classe do domínio, e corresponde por exemplo:

- À classe *Relatórios* acrescenta-se

```

<rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#contém"/>
    <owl:onClass rdf:resource="#Paciente"/>
    <owl:maxQualifiedCardinality rdf:datatype=
      "&xsd;nonNegativeInteger">1</owl:maxQualifiedCardinality>
  </owl:Restriction>
</rdfs:subClassOf>

(...)

```

```

<rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#eh_caracterizado_por"/>
    <owl:someValuesFrom rdf:resource="#Exame"/>
  </owl:Restriction>
</rdfs:subClassOf>

```

- À classe *Exame* acrescenta-se


```

<rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#obtem"/>
    <owl:someValuesFrom rdf:resource="#Diagnóstico"/>
  </owl:Restriction>
</rdfs:subClassOf>

```

(...)

```

<rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#eh_realizado_atraves"/>
    <owl:onClass rdf:resource="#Aparelho"/>
    <owl:maxQualifiedCardinality rdf:datatype=
      "&xsd;nonNegativeInteger">1</owl:maxQualifiedCardinality>
  </owl:Restriction>
</rdfs:subClassOf>

```

- À classe *Diagnóstico* acrescenta-se

```

<rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#indica"/>
    <owl:someValuesFrom rdf:resource="#Entidade_Anatómica"/>
  </owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#eh_suportado_por"/>
    <owl:someValuesFrom rdf:resource="#Observação"/>
  </owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#corresponde"/>
    <owl:onClass rdf:resource="#Zona"/>
    <owl:maxQualifiedCardinality rdf:datatype=
      "&xsd;nonNegativeInteger">1</owl:maxQualifiedCardinality>
  </owl:Restriction>
</rdfs:subClassOf>

```

Capítulo 7

Extracção da Informação e Povoamento da Ontologia

Este capítulo descreve a terceira parte do trabalho, que corresponde à extracção da informação dos relatórios e o respectivo povoamento da ontologia.

A secção 7.1 faz uma descrição do sistema implementado para a extracção da informação. O preenchimento de alguns campos vazios nos relatórios é descrito na secção 7.2. A grande parte da extracção da informação ocorre na secção 7.3 com a criação de uma base de dados. Por fim, na secção 7.4 é descrito o modo como a informação contida nos relatórios foi processada e como foram criadas as respectivas instâncias na ontologia.

7.1 O Sistema de Extracção de Informação

Para extrair-se a informação contida nos relatórios do *corpus* e povoar automaticamente a ontologia, foi desenvolvido um sistema que é constituído pelos seguintes módulos:

- Preenchimento dos campos da identificação do paciente que se encontram vazios;
- Preenchimento dos campos de uma tabela da base de dados com os dados contidos nos relatórios;
- Criação de estruturas e aplicação de técnicas que permitem processar a informação e povoar a ontologia.

7.2 Preenchimento dos Campos Vazios

Tal como foi referido no capítulo 5, existem campos vazios nos relatórios que correspondem à identificação do paciente. Estes campos devem ser preenchidos,

pois o seu preenchimento permite inferir a ontologia sobre os cuidados de saúde do paciente, para fins de investigação ou mesmo no apoio à decisão diagnóstica, embora os dados que são preenchidos não correspondem à realidade.

Os campos a preencher são os seguintes:

- N° de Processo
- Nome
- Serviço
- Idade
- Data de Exame
- Informação Clínica

Ao *N° de Processo* que corresponde ao número de processo do paciente na unidade hospitalar, é atribuído um valor que é incrementado automaticamente em cada relatório.

Para preencher o campo *Nome* do paciente, criou-se um conjunto de nomes, e é a partir desse conjunto, que aleatoriamente o sistema atribui um nome para o paciente.

O campo *Serviço* corresponde ao serviço onde o paciente está internado quando realiza os exames que constituem o relatório. Este campo é sempre preenchido com o valor UAVC (Unidade de Acidente Vascular Cerebral).

O campo *Idade* é um campo extremamente importante quando se pretende inferir a ontologia principalmente para fins de investigação, de modo a saber por exemplo, qual a média de idades para uma certa patologia. Como tal, houve um especial cuidado na informação que iria ser preenchida neste campo, e tratando-se de um relatório de pacientes que têm ou tiveram um AVC, este campo é preenchido por um número aleatório entre 30 e 90, pois é normalmente entre estas idades que ocorre este tipo de problema.

Em relação à *Data de Exame*, esta é atribuída aleatoriamente. Para o ano é atribuído sempre o valor 2009. Em seguida, o sistema atribui aleatoriamente um mês entre 1 e 12, e consoante o mês que é atribuído, é atribuído também aleatoriamente um dia entre 1 e 31 completando assim a data do exame.

Por fim, o campo da *Informação Clínica* do paciente, foi preenchido com a descrição AVC, pois como já foi referido, quem realiza este tipo de exames são habitualmente doentes com AVC.

A figura 7.1, representa o esquema deste módulo. O sistema processa relatório a relatório do *corpus* do projecto, e para cada um deles preenche os campos vazios como descrito acima, criando novos relatórios com os campos da identificação do paciente actualizados.



Figura 7.1: Esquema do Módulo do Sistema para o Preenchimento dos Campos Vazios

O apêndice B representa o relatório do anexo A, mas no formato TXT e com os campos vazios preenchidos.

7.3 Criação e Preenchimento da Base de Dados

A grande parte da extracção da informação, ocorreu com a criação de uma base de dados que permite separar toda a informação que cada relatório possui. Isto vem facilitar o processamento dos dados de cada relatório e o respectivo povoamento da ontologia.

Deste modo, foi criada uma tabela na base de dados com o seguinte código SQL (*Structured Query Language*):

```

CREATE TABLE projecto.relatorio(

    ficheiro          VARCHAR(20) NOT NULL,
    local             VARCHAR(100) NOT NULL,
    tipoDocumento     VARCHAR(50) NOT NULL,
    numeroProcesso    INT NOT NULL,
    nomePaciente      VARCHAR(50) NOT NULL,
    idadePaciente     INT NOT NULL,
    data              DATE NOT NULL,
    servico            VARCHAR(20) NOT NULL,
    informacaoClinica VARCHAR(20) NOT NULL,
    nomeExame1        VARCHAR(50),
    conteudoExame1    TEXT,
    aparelhoExame1    VARCHAR(100),
    nomeExame2        VARCHAR(50),
    propriedadesExame2 VARCHAR(50),
    conteudoExame2    TEXT,
    aparelhoExame2    VARCHAR(100),
    conclusao         TEXT NOT NULL,
    tecnica           VARCHAR(50) NOT NULL,
    medica            VARCHAR(50) NOT NULL,

    PRIMARY KEY(ficheiro)

);
  
```

O campo *ficheiro*, corresponde ao nome do ficheiro de cada relatório.

Os restantes campos da tabela, representam os dados contidos nos relatórios e foram extraídos da seguinte forma:

- *local* - O local onde o relatório é realizado corresponde à primeira linha do ficheiro. Além disso, e sabendo que o relatório é realizado num Laboratório, procura-se pela linha cuja primeira palavra corresponde à palavra "Laboratório";
- *tipoDocumento* - O tipo de documento corresponde ao tipo de relatório que estamos a processar. Para extrair a informação dos relatórios e preencher este campo da tabela, procura-se a linha cuja primeira palavra corresponde à palavra "Relatório", e preenche-se este campo com a informação contida nessa mesma linha;
- *numeroProcesso, nomePaciente, idadePaciente, data, serviço, informacaoClinica* - Todos estes campos da tabela, estão identificados nos relatórios com tags que correspondem respectivamente a "Nº de Processo:", "Nome:", "Idade:", "Serviço:" e "Informação Clínica:". O sistema para preencher estes campos procura no relatório a respectiva tag e preenche o campo com a informação que se encontra à frente dessa tag;
- *nomeExame1* - O exame que corresponde a este campo, é o "Triplex Scan Cervical". Como tal, procura-se a linha que começa por este conjunto de palavras, no entanto, é de realçar que pode ou não existir este tipo de exame no relatório. Esta linha corresponde ao cabeçalho do exame1;
- *nomeExame2* - O exame que corresponde a este campo, é o "Triplex Scan Transcraneano". Como tal, procura-se a linha que começa por este conjunto de palavras, no entanto, é de realçar mais uma vez que pode ou não existir este tipo de exame no relatório. Esta linha corresponde ao cabeçalho do exame2;
- *aparelhoExame1 e aparelhoExame2* - Caso exista o tipo de exame no relatório, a informação que preenche estes campos corresponde à informação que vem a seguir ao nome do exame, e que começa na palavra "Ecografo" (aparelho utilizado neste tipo de exames) e termina na palavra a seguir à palavra "transdutor";
- *propriedadesExame2* - Caso exista este tipo de exame no relatório, este campo é preenchido com a informação que se encontra dentro de parênteses e que se encontra imediatamente a seguir ao nome do exame2;
- *conteudoExame1* - Caso exista este tipo de exame no relatório, a informação que preenche este campo, corresponde às linhas que aparecem imediatamente a seguir à linha do cabeçalho do exame1, até encontrar uma linha que corresponda ao cabeçalho do exame2 ou até encontrar uma linha com a tag "CONCLUSÃO:";
- *conteudoExame2* - Caso exista este tipo de exame no relatório, a informação que preenche este campo, corresponde às linhas que aparecem imediatamente

a seguir à linha do cabeçalho do exame2, até encontrar uma linha com a tag "CONCLUSÃO:";

- *conclusao* - Este campo é preenchido com as linhas que estão entre a tag "CONCLUSÃO:" e a tag "Técnica";
- *tecnica e medica* - A informação a preencher nestes campos, corresponde à informação que aparece respectivamente a seguir às tags "Técnica" e "Médica".

A figura 7.2, representa o esquema deste módulo. O sistema processa relatório a relatório, e para cada um deles preenche os campos da base de dados com a informação de cada relatório.



Figura 7.2: Esquema do Módulo do Sistema para o Preenchimento da Base de Dados

7.4 Processamento da Informação e Criação das Instâncias

Com a informação dos relatórios separada na base de dados, cabe ao sistema neste último módulo processar essa informação de modo a criar as instâncias e a povoar a ontologia.

Facilmente se verifica, que exceptuando os campos com o conteúdo dos exames e da conclusão, toda a informação contida nos outros campos da base de dados permite quase directamente criar as instâncias na ontologia.

Tendo sempre por base o relatório do apêndice B e da utilização do Jena como parser do OWL, a informação de cada campo da base de dados é processada da seguinte forma.

Para a informação contida nos campos *local* e *tipoDocumento*, nos campos da identificação do paciente e no campo *medica*, apenas é necessário criar instâncias pois a informação não necessita de ser processada. Assim, para criar as respectivas instâncias é gerado o seguinte código OWL:

```

<owl:Thing rdf:about="#Relatorio1">
  <rdf:type rdf:resource="#Neurovascular"/>
  <númeroRelatório rdf:datatype="&xsd;int">1</númeroRelatório>
  <local rdf:datatype="&xsd;string"
    >Laboratório de Ultrassons Cardio e Neurovascular</local>
  <data>2009-05-17</data>
  <gera rdf:resource="#Conclusao1"/>
  <eh_caracterizado_por rdf:resource="#Exame1"/>
  <eh_caracterizado_por rdf:resource="#Exame2"/>
  <eh_realizado_por rdf:resource="#Medico1"/>
  <contém rdf:resource="#Paciente1"/>
</owl:Thing>

<owl:Thing rdf:about="#Paciente1">
  <rdf:type rdf:resource="#Paciente"/>
  <númeroProcessoPaciente rdf:datatype=
    "&xsd;int">111</númeroProcessoPaciente>
  <idadePaciente rdf:datatype="&xsd;int">66</idadePaciente>
  <nomePacienterdf:datatype=
    "&xsd;string">LuísClérigo</nomePaciente>
  <tem rdf:resource="#InformacaoClinica1"/>
  <encontra-se rdf:resource="#Servico1"/>
</owl:Thing>

<Serviço rdf:about="#Servico1">
  <rdf:type rdf:resource="&owl;Thing"/>
  <nomeServiço rdf:datatype="&xsd;string">UAVC</nomeServiço>
</Serviço>

<Informação_Clínica rdf:about="#InformacaoClinica1">
  <rdf:type rdf:resource="&owl;Thing"/>
  <descriçãoInformaçãoClínica rdf:datatype=
    "&xsd;string">AVC</descriçãoInformaçãoClínica>
</Informação_Clínica>

<owl:Thing rdf:about="#Medico1">
  <rdf:type rdf:resource="#Médico"/>
  <nomeMédico rdf:datatype=
    "&xsd;string">Irene Mendes</nomeMédico>
  <realiza rdf:resource="#Relatorio1"/>
</owl:Thing>

```

Em relação ao cabeçalho dos exames, a informação do campo correspondente ao *Nome* do exame não precisa de ser processada.

O campo do aparelho onde são realizados os exames, possui a informação do seguinte modo: *"Ecografo IE33/HD11 XE da Philips transdutor L3-12Mhz"*. A informação deste campo tem de ser processada, de forma a extrair desta frase a marca e o transdutor do aparelho. Para se extrair a marca do aparelho, o sistema considera

a informação que aparece a seguir à palavra "Ecografo" até à palavra "transdutor". Já em relação ao transdutor, é considerada a informação que se segue à palavra "transdutor".

Já o campo *propriedadesExame*, possui a informação do seguinte modo: "(Via Transtemporal e Suboccipital)". A informação deste campo tem então de ser processada, de modo a detectarmos quais as Janelas Ultra-Sónicas usadas no exame. Para isso, o sistema elimina da frase os parênteses, artigos e a palavra "Via" caso esta exista na frase. As restantes palavras da frase são comparadas com as Janelas Ultra-Sónicas que existem na ontologia de forma a detectá-las com maior exactidão.

Para o cabeçalho dos exames, o código OWL gerado para criar as instâncias é o seguinte:

```
<owl:Thing rdf:about="#Exame1">
  <rdf:type rdf:resource="#Triplex_Scan_Cervical"/>
  <eh_realizado_atraves rdf:resource="#Ecografo1"/>
  <possui rdf:resource="#Observacao1"/>
  <possui rdf:resource="#Observacao2"/>
  <eh_executado_por rdf:resource="#Tecnica1"/>
  <eh_executado_por rdf:resource="#Tecnica2"/>
</owl:Thing>

<Ecografo rdf:about="#Ecografo1">
  <rdf:type rdf:resource="#owl:Thing"/>
  <marca rdf:datatype="#xsd:string">
    >IE33/HD11 XE da Philips</marca>
  <transdutor rdf:datatype="#xsd:string">L3-12Mhz</transdutor>
</Ecografo>

<Triplex_Scan_Transcraneano rdf:about="#Exame2">
  <rdf:type rdf:resource="#owl:Thing"/>
  <eh_realizado_atraves rdf:resource="#Ecografo2"/>
  <usa rdf:resource="#JanelaUltraSonica1"/>
  <possui rdf:resource="#Observacao3"/>
  <possui rdf:resource="#Observacao4"/>
  <eh_executado_por rdf:resource="#Tecnica1"/>
  <eh_executado_por rdf:resource="#Tecnica2"/>
</Triplex_Scan_Transcraneano>

<Ecografo rdf:about="#Ecografo2">
  <rdf:type rdf:resource="#owl:Thing"/>
  <marca rdf:datatype="#xsd:string">
    >IE33/HD11 XE da Philips</marca>
  <transdutor rdf:datatype="#xsd:string">S1-5MHz</transdutor>
</Ecografo>
```

```

<Transtemporal rdf:about="#JanelaUltraSonica1">
  <rdf:type rdf:resource="#Suboccipital"/>
  <rdf:type rdf:resource="#owl:Thing"/>
</Transtemporal>

```

O campo *tecnica* possui a informação do seguinte modo: "*Sónia Mateus/Vanda Pós-de-Mina*". Facilmente se conclui que o exame pode ser executado por uma ou várias técnicas. Como tal, esta informação tem de ser processada de modo a detectar o número de técnicas, e a separar a informação caso seja necessário. Caso exista mais do que um nome, estes encontram-se separados por "/". Assim, o sistema procura por "/" e substitui por "*". Em seguida vai processando palavra a palavra, e sempre que encontra um "*" sabe que termina um nome e irá começar outro, detectando assim o nome das técnicas que executam o exame. Para criar estas instâncias é gerado o seguinte código OWL:

```

<Técnico rdf:about="#Tecnica1">
  <rdf:type rdf:resource="#owl:Thing"/>
  <nomeTécnico rdf:datatype="#xsd:string">
    >Sónia Mateus</nomeTécnico>
  <executa rdf:resource="#Exame1"/>
  <executa rdf:resource="#Exame2"/>
</Técnico>

<Técnico rdf:about="#Tecnica2">
  <rdf:type rdf:resource="#owl:Thing"/>
  <nomeTécnico rdf:datatype="#xsd:string">
    >Vanda Pós-de-Mina</nomeTécnico>
  <executa rdf:resource="#Exame1"/>
  <executa rdf:resource="#Exame2"/>
</Técnico>

```

Por fim, para os campos *conteudoExame* e *conclusao* a informação necessita de ser processada, porque corresponde a um texto com um conjunto de diagnósticos, observações, entidades anatómicas, etc, relacionadas entre si.

De modo a processar o texto do conteúdo dos exames e da conclusão, recorreu-se a uma outra subárea da Inteligência Artificial, a área da Aprendizagem Automática, utilizando neste caso as árvores de decisão. Esta foi a forma encontrada para ultrapassar a especificidade do vocabulário usado no texto, onde por exemplo, na construção das frases a utilização de verbos e adjectivos é praticamente inexistente.

Antes de mais, e a partir da definição da ontologia realizada no capítulo 6, foram criadas várias classes, para que cada diagnóstico, cada observação, cada entidade anatómica, etc, seja identificado por uma classe, cuja estrutura está representada na figura 7.3.



Figura 7.3: Estrutura de classes

De referir sobre esta estrutura, que as classes *Entidade Anatômica*, *Artéria*, *Diagnóstico* e *Observação* são classes abstractas de modo a que as suas subclasses implementem todos os seus métodos.

Foi criada uma árvore de decisão (apêndice C) a partir da estrutura de classes e dos termos identificados nos relatórios, de forma a conseguir-se processar o texto do conteúdo dos exames e da conclusão. Para não haver repetições na árvore, foram ainda desenvolvidos três métodos que detectam se uma palavra é singular/plural, se tem género ou iniciais de uma palavra da raiz ou do nó, onde o sistema se encontra no processamento do texto na árvore.

O processamento do texto é executado parágrafo a parágrafo. Deste modo, cada parágrafo:

- É transformado em minúsculas;
- São eliminados os acentos, cedilhados e caracteres especiais;
- É processado por uma lista de *stopwords* que possui palavras a eliminar e que são dispensáveis, tais como os determinantes.

Depois de passar por este processo, o parágrafo está pronto a ser analisado! O sistema processa palavra a palavra do parágrafo utilizando a árvore de decisão, e sempre que chega a uma folha da árvore (decisão encontrada) cria uma instância dessa decisão e insere na última posição de um vector que é criado para cada parágrafo, e que é definido de acordo com a estrutura de classes criada anteriormente. Quando se chega ao fim do parágrafo, se forem detectadas instâncias (decisões encontradas na árvore), este vector irá ter preenchidas várias posições com diagnósticos, observações, entidades anatómicas, etc.

Por exemplo, o parágrafo com a frase "*Eixos Carotídeos, Artérias Subclávias e Artérias Vertebrales (V0 a V3) permeáveis, com paredes arteriais e fluxos normais.*" depois de ser processado, iria obter o vector da figura 7.4.

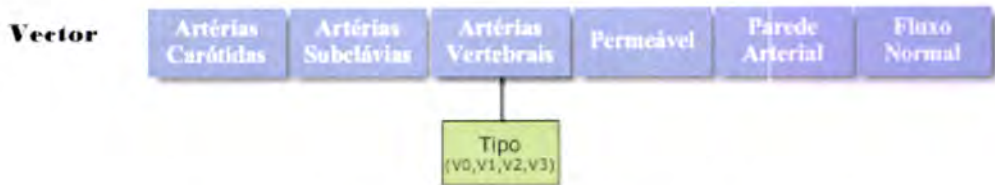


Figura 7.4: Vector com as Instâncias do Parágrafo

Depois do processamento do parágrafo e da obtenção do vector com as instâncias, este vector irá ter preenchidas várias posições com diagnósticos, observações, entidades anatómicas, etc, como mostra a figura 7.4. No entanto, podem existir relações entre instâncias, mais concretamente, seis tipo de relações:

1. Lados relacionados com Entidades Anatômicas (Exemplo: "Artéria Subclávia **direita**");
2. Zonas relacionadas com Entidades Anatômicas (Exemplo: "Aumento da velocidade de fluxo a **montante** da AV4");
3. Stent relacionado com Entidade Anatômica (Exemplo: "**Stent** permeável na Artéria Basilar");
4. Entidades Anatômicas relacionadas com Diagnósticos (Exemplo: "Oclusão da **Artéria Carótida Interna**");
5. Entidades Anatômicas relacionadas com Observações (Exemplo: "**Artérias Vertebrais (V0 a V3)** permeáveis."
6. Observações relacionadas com Diagnósticos (Exemplo: "**IPs aumentados** traduzindo Inelastecidade Arterial").

Como tal, foram desenvolvidos um conjunto de algoritmos que processam este vector e que de acordo com a ordem com que as instâncias estão no vector, atribui as instâncias que pertencem a outras instâncias, a essas mesmas instâncias.

Para o primeiro caso, o algoritmo desenvolvido é o seguinte:

Algorithm 1 Relaciona as instâncias Lados e Entidades Anatômicas

```
1: procedure PROCESSALADOS(Vector < Arvore > instancias)
2:   Arvore[] inst ← instancias
3:   for i ← 0 to i < tamanho(inst) do
4:     if inst[i] instance of Lado then
5:       inst[i - 1].constituente ← inst[i]
6:       remove inst[i]
7:       i ← 0
8:     end if
9:   end for
10:  return inst
11: end procedure
```

A figura 7.5, representa a aplicação deste algoritmo a partir do exemplo dado. Neste caso, a instância *Lado* passa a ser um constituinte da instância *Artéria Subclávia*.

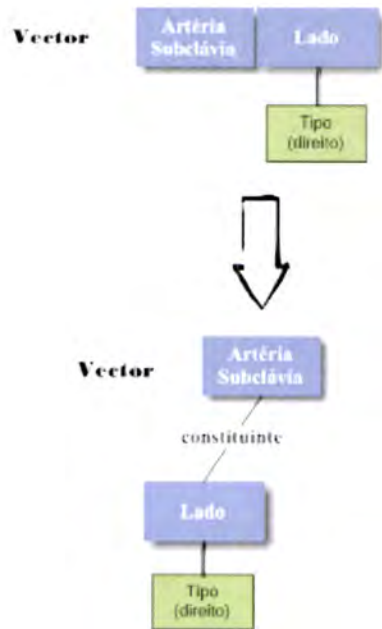


Figura 7.5: Algoritmo que relaciona as instâncias Lado e Entidades Anatômicas

Para o segundo caso, o algoritmo é o seguinte:

Algorithm 2 Relaciona as instâncias Zonas e Entidades Anatômicas	
1:	procedure PROCESSAZONAS(<i>Vector</i> < <i>Arvore</i> > <i>instancias</i>)
2:	<i>Arvore</i> [] <i>inst</i> ← <i>instancias</i>
3:	for <i>i</i> ← 0 to <i>i</i> < tamanho(<i>inst</i>) do
4:	if <i>inst</i> [<i>i</i>] <i>instance of Zona</i> then
5:	<i>k</i> ← <i>i</i> + 1
6:	if <i>inst</i> [<i>k</i>] <i>instance of Entidade Anatomica</i> then
7:	<i>inst</i> [<i>i</i>]. <i>constituente</i> ← <i>inst</i> [<i>k</i>]
8:	remove <i>inst</i> [<i>k</i>]
9:	<i>i</i> ← 0
10:	end if
11:	end if
12:	end for
13:	return <i>inst</i>
14:	end procedure

A figura 7.6, representa a aplicação deste algoritmo a partir do exemplo dado. Neste caso, a instância *Artéria Vertebral V4* passa a ser um constituinte da instância *Zona*.

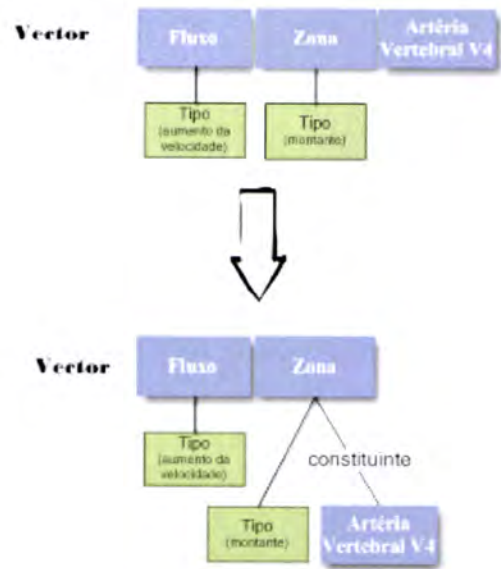


Figura 7.6: Algoritmo que relaciona as instâncias Zona e Entidades Anatômicas

Para o terceiro caso, o algoritmo é o seguinte:

Algorithm 3 Relaciona as instâncias Stent e Entidades Anatômicas

```
1: procedure PROCESSASTENT(Vector < Arvore > instancias)
2:   Arvore[] inst ← instancias
3:   for i ← 0 to i < tamanho(inst) do
4:     if inst[i] instance of Stent then
5:       k ← i + 1
6:       while inst[k] not instance of Entidade Anatomica do
7:         k ← k + 1
8:       end while
9:       if inst[k] instance of Entidade Anatomica then
10:        inst[k].constituente ← inst[i]
11:        remove inst[i]
12:        i ← 0
13:      end if
14:    end if
15:  end for
16:  return inst
17: end procedure
```

A figura 7.7, representa a aplicação deste algoritmo a partir do exemplo dado. Neste caso, a instância *Stent* passa a ser um constituinte da instância *Artéria Basilar*.

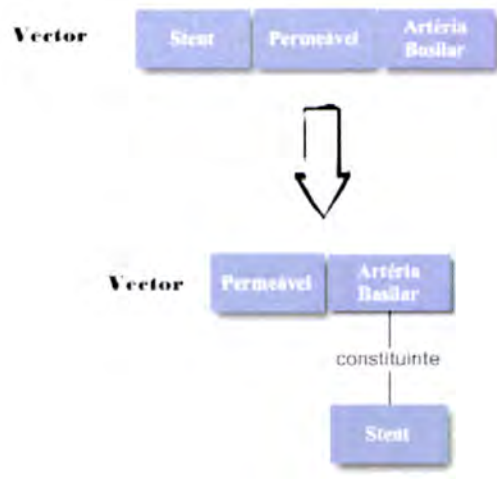


Figura 7.7: Algoritmo que relaciona as instâncias Stent e Entidades Anatômicas

Para o quarto caso, o algoritmo é o seguinte:

Algorithm 4 Relaciona as instâncias Entidades Anatômicas e Diagnósticos

```
1: procedure  PROCESSAENTIDADESDIAGNOSTICO(Vector  <  Arvore  >
   instancias)
2:   Arvore[] inst ← instancias
3:   i ← 0
4:   while i < tamanho(inst) do
5:     if inst[i] instance of Diagnostico then
6:       k ← i + 1
7:       if k < tamanho(inst) and inst[k] instance of Zona ou EA then
8:         while k < tamanho(inst) do
9:           if inst[k] not instance of Diagnostico ou Observacao then
10:            inst[i].constituente ← inst[k]
11:            remove inst[k]
12:          else
13:            break
14:          end if
15:        end while
16:      end if
17:    end if
18:    i ← i + 1
19:  end while
20:  return inst
21: end procedure
```

A figura 7.8, representa a aplicação deste algoritmo a partir do exemplo dado. Neste caso, a instância *Artéria Carótida* passa a ser constituinte do diagnóstico com a instância *Oclusão*.

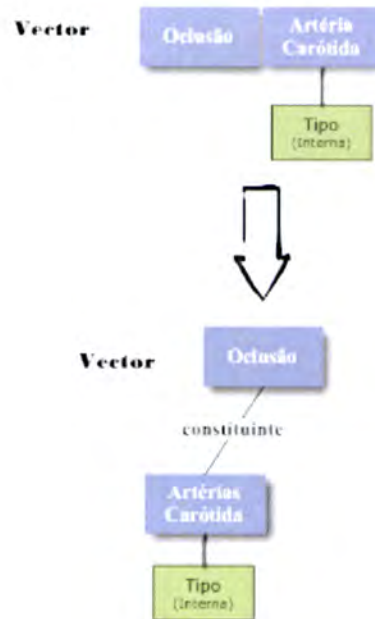


Figura 7.8: Algoritmo que relaciona as instâncias Entidades Anatômicas e Diagnóstico

Para o quinto caso, a figura 7.9 representa a aplicação deste algoritmo a partir do exemplo dado. Neste caso, a instância *Artéria Vertebral* passa a ser constituinte da observação com a instância *Permeável*.

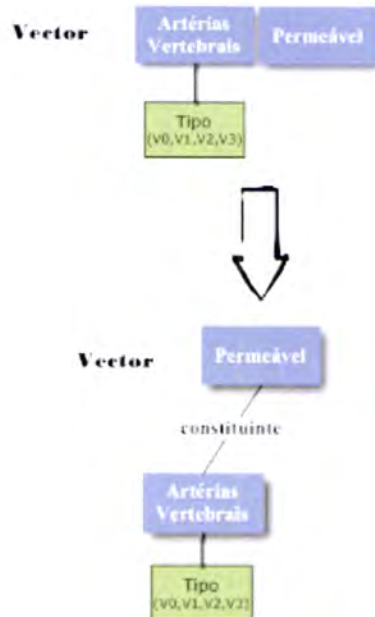


Figura 7.9: Algoritmo que relaciona as instâncias Entidades Anatômicas e Observações

O algoritmo para este caso é o seguinte:

Algorithm 5 Relaciona as instâncias Entidades Anatômicas e Observações

```

1: procedure  PROCESSAENTIDADESOBSERVACAO(Vector  <  Arvore  >
   instancias)
2:   Arvore[] inst  $\leftarrow$  instancias
3:   i  $\leftarrow$  0
4:   while i < tamanho(inst) do
5:     if inst[i] instance of Observacao then
6:       if i > 0 and inst[i - 1] instance of EA ou Zona then
7:         k  $\leftarrow$  i - 1
8:         while k  $\geq$  0 do
9:           if inst[k] not instance of Observacao ou Diagnostico then
10:            inst[i].constituente  $\leftarrow$  inst[k]
11:            remove inst[k]
12:            i  $\leftarrow$  i - 1
13:            k  $\leftarrow$  k - 1
14:           else
15:             break
16:           end if
17:         end while
18:       else if i + 1 < tamanho(inst) and inst[i + 1] instance of EA ou
   Zona then
19:         k  $\leftarrow$  i + 1
20:         while k < tamanho(inst) do
21:           if inst[k] not instance of Observacao ou Diagnostico then
22:             inst[i].constituente  $\leftarrow$  inst[k]
23:             remove inst[k]
24:           else
25:             break
26:           end if
27:         end while
28:       else
29:         k  $\leftarrow$  i - 1
30:         while k  $\geq$  0 do
31:           if inst[k] instance of Observacao ou Diagnostico then
32:             existe  $\leftarrow$  true
33:             break
34:           end if
35:           existe  $\leftarrow$  false
36:           k  $\leftarrow$  k - 1
37:         end while
38:         if existe == true and inst[k] instance of Observacao then
39:           inst[i].constituente  $\leftarrow$  inst[k].constituente
40:         else if existe == true and inst[k] instance of Diagnostico then
41:           inst[i].constituente  $\leftarrow$  inst[k].constituente
42:           remove inst[k].constituente
43:         end if
44:       end if
45:     end if
46:     i  $\leftarrow$  i + 1
47:   end while
48:   return inst
49: end procedure

```

Por fim para o sexto caso, o algoritmo é o seguinte:

Algorithm 6 Relaciona as instâncias Observações e Diagnósticos

```

1: procedure PROCESSAOBSERVACOESDIAGNOSTICO(Vector < Arvore >
   instancias)
2:   Arvore[] inst  $\leftarrow$  instancias
3:   i  $\leftarrow$  0
4:   while i < tamanho(inst) do
5:     if inst[i] instance of Diagnostico then
6:       if i > 0 and inst[i - 1] instance of Observacao then
7:         k  $\leftarrow$  i - 1
8:         while k  $\geq$  0 do
9:           if inst[k] not instance of Diagnostico then
10:            inst[i].constituente  $\leftarrow$  inst[k]
11:            remove inst[k]
12:            i  $\leftarrow$  i - 1
13:            k  $\leftarrow$  k - 1
14:           else
15:             break
16:           end if
17:         end while
18:       else if i == 0 and tamanho(inst) > 1 then
19:         if tamanho(inst[i].constituientes) == 0 and inst[i + 1] instance
           of Observacao then
20:           k  $\leftarrow$  i + 1
21:           while k < tamanho(inst) do
22:             if inst[k] not instance of Diagnostico then
23:               inst[i].constituente  $\leftarrow$  inst[k]
24:               remove inst[k]
25:             else
26:               break
27:             end if
28:           end while
29:         end if
30:       end if
31:     end if
32:     i  $\leftarrow$  i + 1
33:   end while
34:   return inst
35: end procedure

```

A figura 7.10, representa a aplicação deste algoritmo a partir do exemplo dado. Neste caso, a observação com a instância *IP* passa a ser constituinte do diagnóstico com a instância *Inelastecidade Arterial*.

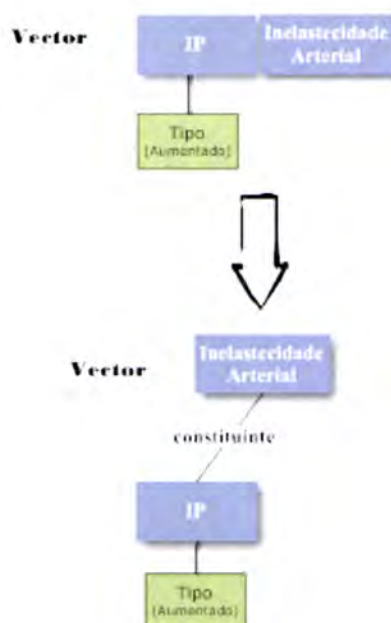


Figura 7.10: Algoritmo que relaciona as instâncias Observação e Diagnóstico

No conteúdo da conclusão, verifica-se por vezes que:

- Depois de um parágrafo ser processado, o vector das instâncias está vazio;
- No fim da frase, depois de uma observação ou diagnóstico, existe informação relevante que se deve instanciar na ontologia. Um desses exemplos é por exemplo a frase: *"Stent no Eixo Carotídeo Esquerdo/Direito permeável, compatível com Estenose Residual/Remodelação de cerca de (segundo Cavatas 2005), a controlar aos 6 meses após a intervenção."*

O sistema tem em conta estes dois casos, e para tal foi criado na ontologia o atributo *descricao* na classe *conclusao*.

No primeiro caso, o parágrafo é imediatamente instanciado no atributo. Um exemplo para este caso, é o do relatório do apêndice B que possui na conclusão o parágrafo *"Exames normais."*

No segundo caso, e no fim do processamento do parágrafo, é extraída a última parte da frase pois no caso de existir informação relevante, esta estará a seguir à última vírgula. Esta frase extraída é então processada pela árvore de decisão, e caso o vector das instâncias esteja vazio, esta frase extraída do parágrafo original, é instanciada no atributo *descricao*.

No fim de cada parágrafo do conteúdo dos exames e da conclusão ser processado pela árvore de decisão, e do vector das instâncias resultante do processamento anterior ser processado pelos algoritmos acima mencionados, cria-se então as instâncias na ontologia. Para tal foi criado em todas as classes da estrutura de classes, um método que é invocado recursivamente e que cria as instâncias na ontologia.

Para todas as instâncias que possuem termos médicos, tais como as entidades anatómicas, observações e diagnósticos, existem atributos (*CUI* e *CN*) que são

preenchidos a partir das invocações realizadas ao *WebService* do UMLS com o respectivo termo.

Assim, após processamento do primeiro parágrafo do primeiro exame do apêndice B, o código OWL gerado para criar as instâncias é o seguinte:

```
<owl:Thing rdf:about="#Observacao1">
  <rdf:type rdf:resource="#Permeável"/>
  <observaçãoCUI rdf:datatype=
    "&xsd:string">C0031164</observaçãoCUI>
  <observaçãoCN rdf:datatype=
    "&xsd:string">Permeability</observaçãoCN>
  <refere-se rdf:resource="#Arterial1"/>
  <refere-se rdf:resource="#Arteria2"/>
  <refere-se rdf:resource="#Arteria3"/>
</owl:Thing>

<Interna rdf:about="#Arteria3">
  <rdf:type rdf:resource="#Externa"/>
  <rdf:type rdf:resource="#owl:Thing"/>
  <artériaCUI rdf:datatype=
    "&xsd:string">C0007272</artériaCUI>
  <artériaCN rdf:datatype="&xsd:string"
    >Carotid Arteries</artériaCN>
</Interna>

<owl:Thing rdf:about="#Arteria2">
  <rdf:type rdf:resource="#Subclávia"/>
  <artériaCUI rdf:datatype=
    "&xsd:string">C0038530</artériaCUI>
  <artériaCN rdf:datatype="&xsd:string"
    >Structure of subclavian artery</artériaCN>
</owl:Thing>

<owl:Thing rdf:about="#Arterial1">
  <rdf:type rdf:resource="#V0"/>
  <rdf:type rdf:resource="#V1"/>
  <rdf:type rdf:resource="#V2"/>
  <rdf:type rdf:resource="#V3"/>
  <artériaCUI rdf:datatype=
    "&xsd:string">C0042559</artériaCUI>
  <artériaCN rdf:datatype="&xsd:string"
    >Structure of vertebral artery</artériaCN>
</owl:Thing>

<owl:Thing rdf:about="#Observacao2">
  <rdf:type rdf:resource="#Normal"/>
  <observaçãoCUI rdf:datatype=
    "&xsd:string">C0806140</observaçãoCUI>
```



```

<observaçãoCN rdf:datatype=
  "&xsd:string">Flow</observaçãoCN>
<refere-se rdf:resource="#Parede_Arterial1"/>
</owl:Thing>

<owl:Thing rdf:about="#Parede_Arterial1">
  <rdf:type rdf:resource="#Parede_Arterial"/>
  <paredeArterialCUI rdf:datatype=
    "&xsd:string">C0507850</paredeArterialCUI>
  <paredeArterialCN rdf:datatype=
    "&xsd:string">Wall of artery</paredeArterialCN>
</owl:Thing>

```

A figura 7.11, representa o esquema deste módulo. O sistema recebe como input a ontologia base (sem instâncias), e vai recebendo a informação que está contida nos relatórios que se encontram na base de dados. Se essa informação não precisar de ser processada, cria logo automaticamente as instâncias na ontologia. Caso contrário, se a informação corresponder ao conteúdo dos exames e da conclusão, essa informação é processada pela árvore de decisão e pelos algoritmos criados para o efeito, são realizadas invocações ao *Webservice* do UMLS, e são criadas as instâncias de acordo com a informação processada.

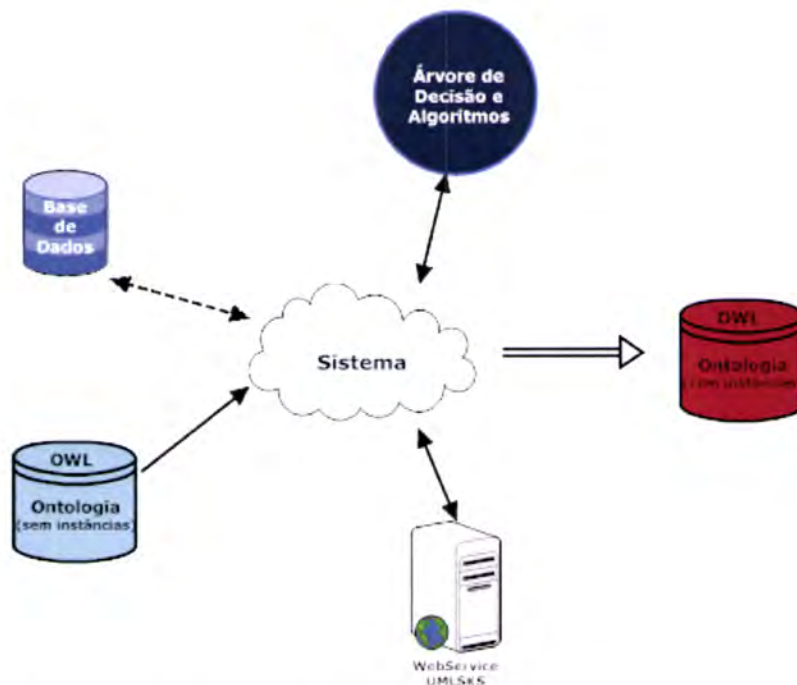


Figura 7.11: Esquema do Módulo do Sistema para o Processamento da Informação e Criação de Instâncias

Capítulo 8

Utilização do Sistema

Este capítulo descreve a última parte do trabalho que corresponde à utilização do sistema desenvolvido nos capítulos anteriores.

A secção 8.1 corresponde ao desenvolvimento de uma aplicação que facilita o uso do sistema. Na secção 8.2 é realizado um teste ao resultado final obtido após o processamento do sistema.

8.1 Interface

Para facilitar todo o processo de inserção dos relatórios no *corpus*, e a obtenção do respectivo ficheiro com o código OWL da ontologia, foi desenvolvida uma aplicação, que corresponde a uma interface desenvolvida em Java.

O utilizador tem assim apenas de introduzir os relatórios que quer converter, e escolher qual o local onde o sistema irá guardar o ficheiro OWL. A figura 8.1, corresponde a um exemplo da utilização da aplicação.



Figura 8.1: Exemplo da Utilização da Aplicação

Deste modo, a legenda da figura 8.1 corresponde respectivamente a:

1. Relatórios a converter
2. Local onde o sistema cria o ficheiro OWL
3. Ícone para converter os relatórios

A interface exibe ainda três tipos de mensagens, como mostra a figura 8.2.



Figura 8.2: Mensagens Exibidas pela Interface

A figura 8.2a informa o utilizador que é necessário haver relatórios como input e que é necessário escolher o local do output do ficheiro OWL. Em relação à figura 8.2b, esta mensagem é exibida quando houve um erro a converter os relatórios devido ao seu formato. Por fim a figura 8.2c, informa o utilizador que o sistema converteu com sucesso os relatórios e criou o ficheiro OWL no local escolhido pelo utilizador.

8.2 Testes

Depois da conversão dos relatórios para OWL, cabe agora nesta fase fazer alguns testes à ontologia de forma a avaliar as instâncias criadas pelo sistema. Para tal, foi utilizado como software de apoio o Protégé, que permite manipular facilmente as instâncias da ontologia.

No teste aqui apresentado, foi tomado como exemplo o relatório do apêndice B.

Para o cabeçalho deste relatório, e a partir da definição da ontologia representada na figura 6.1, o sistema cria a instância da figura 8.3.

As relações (object property) deste relatório e representadas nesta figura, estão de acordo com as relações do esquema da ontologia representado na figura 6.10. O conjunto destas relações representa a restante informação contida no relatório.

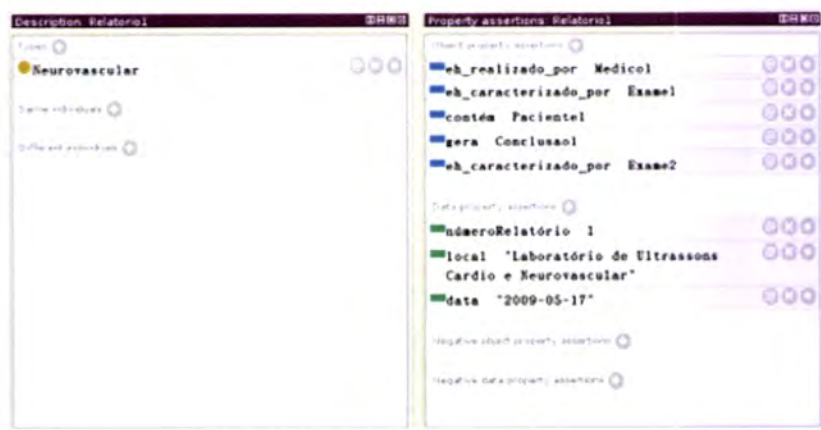


Figura 8.3: Instância do Relatório

A informação respeitante à "Identificação do Paciente" e "Informação Clínica" está representada na instância *Paciente1* como mostra a figura 8.4, e que corresponde à definição da ontologia representada na figura 6.2.



Figura 8.4: Instâncias do Paciente

Para a relação médico, que está definido na ontologia como representado na figura 6.4, o sistema criou a instância da figura 8.5.

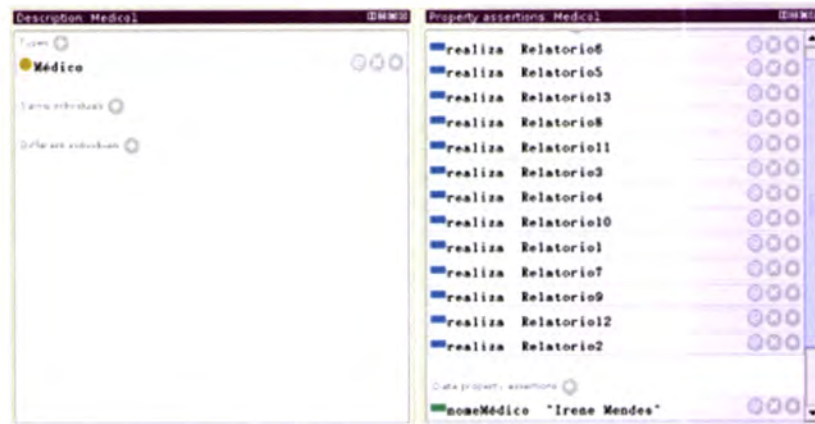


Figura 8.5: Instância do Médico

Este relatório, e tal como revela a instância da figura 8.3, possui dois tipos de exames.

Para o primeiro exame, e de acordo com a definição da ontologia representada na figura 6.3, o sistema criou as instâncias da figura 8.6.



Figura 8.6: Instâncias do Primeiro Exame

Este primeiro exame, tem como conteúdo a seguinte frase: *"Eixos Carotídeos, Artérias Subclávias e Artérias Vertebrais (V0 a V3) permeáveis, com paredes arteriais e fluxos normais."* De acordo com o esquema da ontologia representado na figura 6.10, o sistema criou dois grupos de instâncias, onde no topo estão as instâncias *Observacao1* e *Observacao2*.

As instâncias do primeiro grupo estão representadas na figura 8.7.

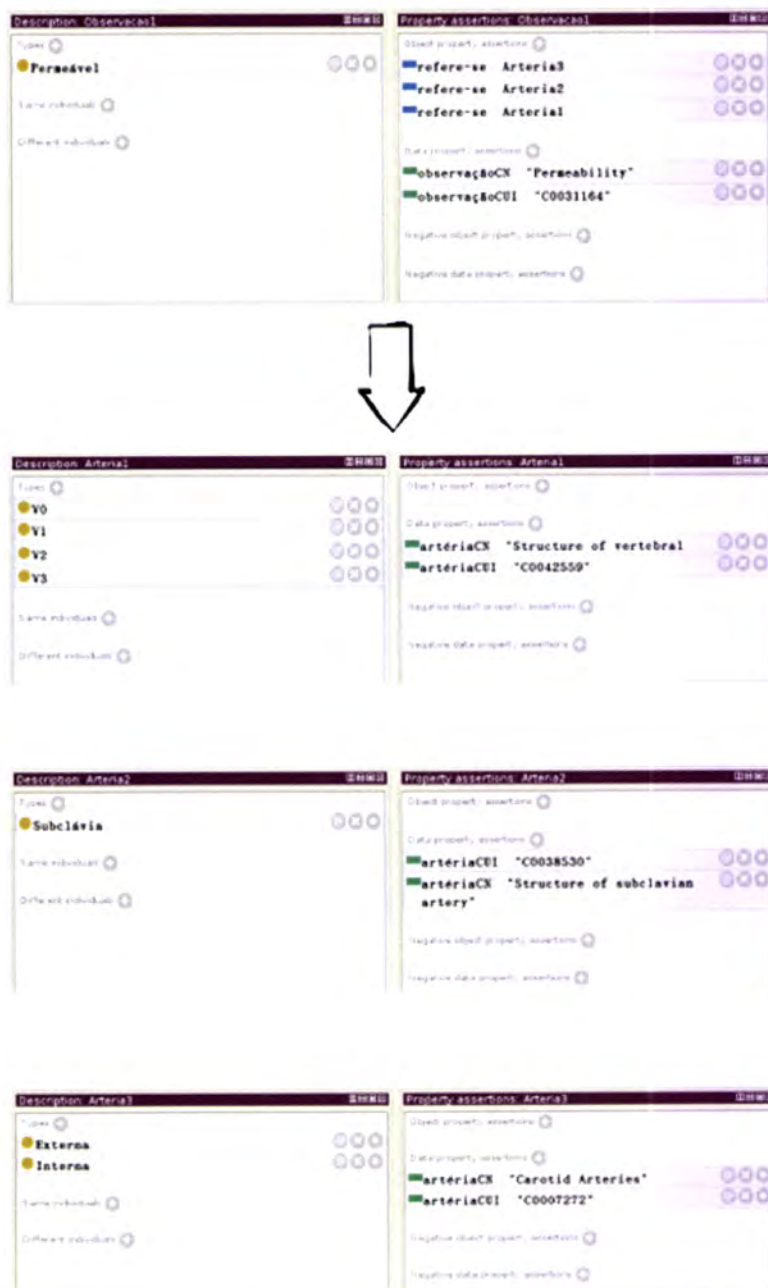


Figura 8.7: Instâncias do Primeiro Grupo do Conteúdo do Primeiro Exame

As instâncias do segundo grupo estão representadas na figura 8.8.

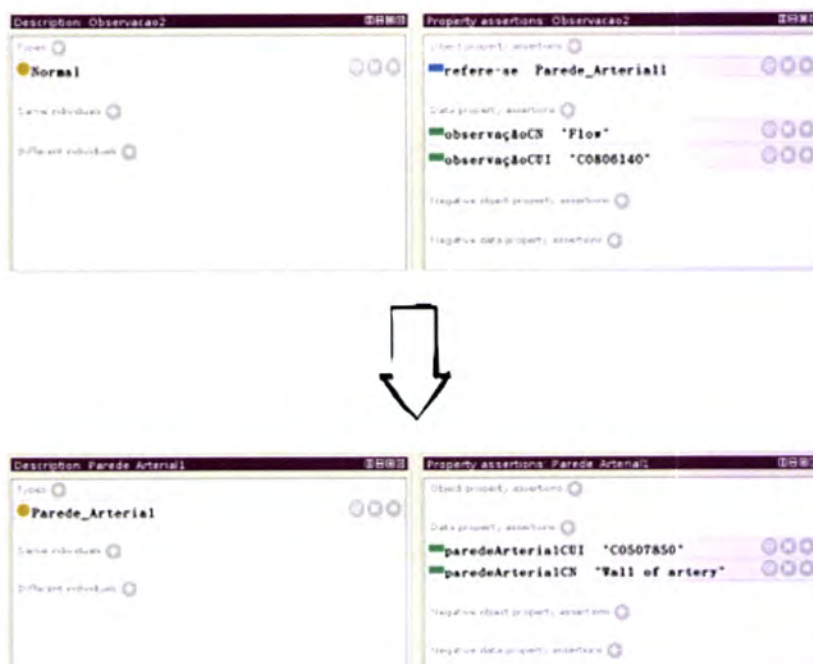


Figura 8.8: Instâncias do Segundo Grupo do Conteúdo do Primeiro Exame

Para as técnicas que executaram os exames, e de acordo com a definição da ontologia representada na figura 6.4, o sistema criou as instâncias da figura 8.9.

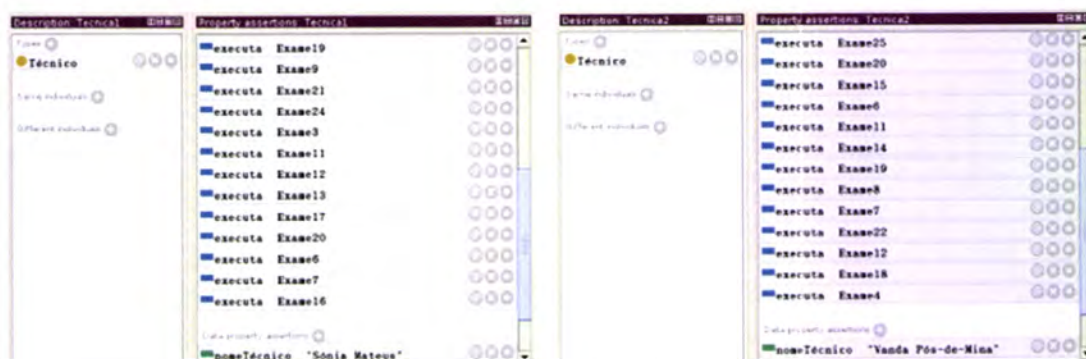


Figura 8.9: Instância do Técnico

Para o segundo exame, e de acordo com a definição da ontologia representada na figura 6.3, o sistema criou as instâncias da figura 8.10.

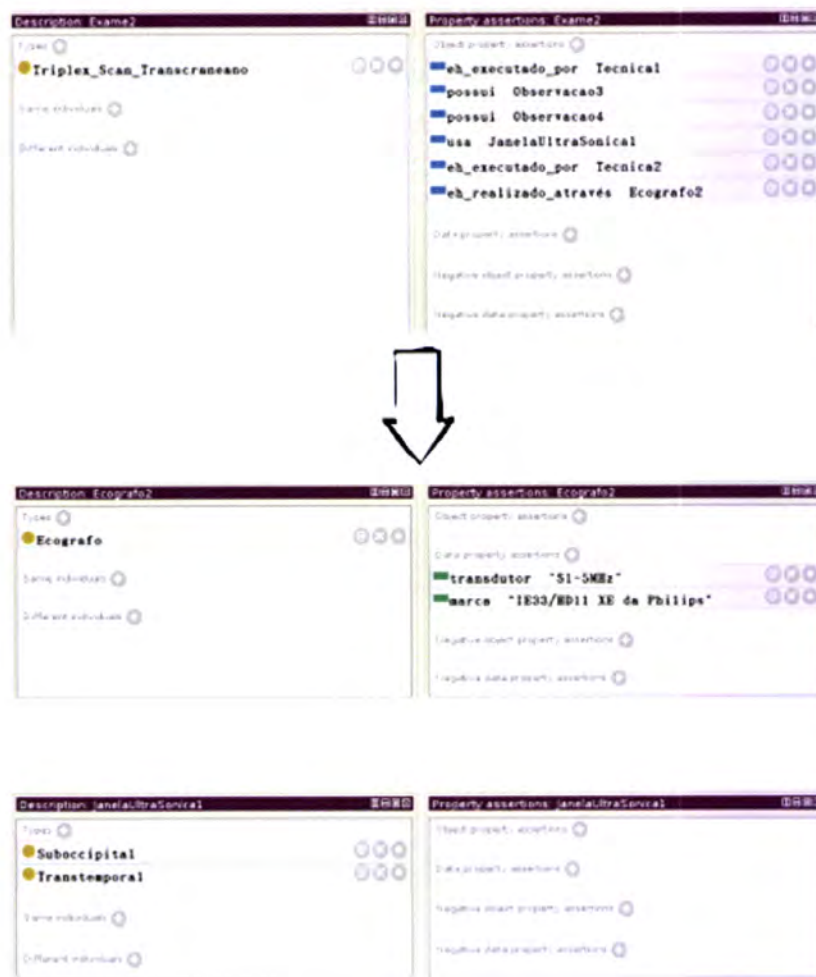


Figura 8.10: Instâncias do Segundo Exame

Este segundo exame, tem como conteúdo a seguinte frase: *"Permeabilidade de todos os segmentos arteriais estudados (ACM, ACI, ACA, ACP1, ACP2, AV4 e AB), com fluxos normais."* De acordo com o esquema da ontologia representado na figura 6.10, o sistema criou dois grupos de instâncias, onde no topo estão as instâncias *Observacao3* e *Observacao4*.

As instâncias estão representadas na figura 8.11.

Por fim, e para a conclusão do relatório que possui como conteúdo a frase "Exames normais.", o sistema gera a instância que está representada na figura 8.12.

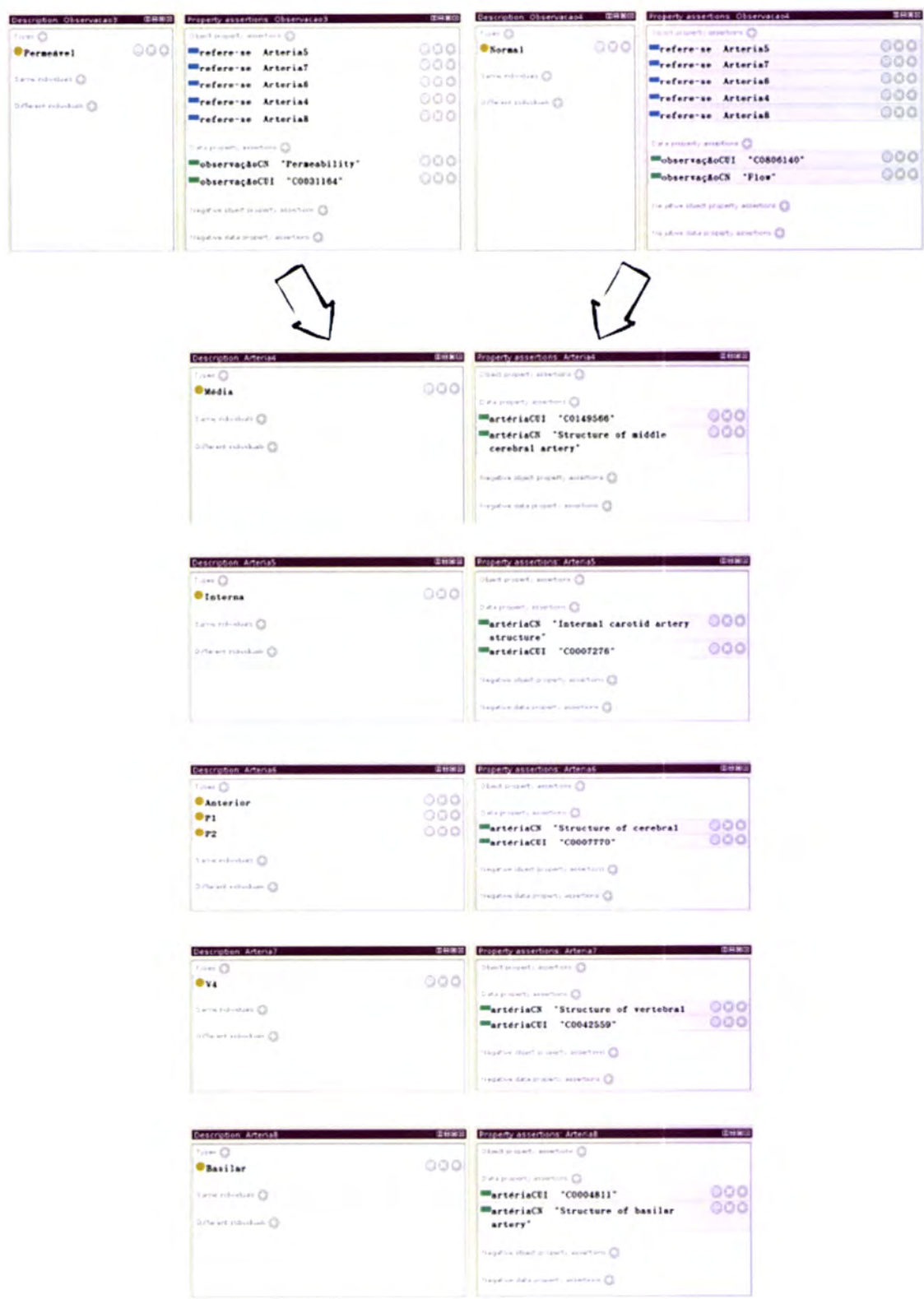


Figura 8.11: Instâncias do Conteúdo do Segundo Exame



Figura 8.12: Instância da Conclusão

Comparando o relatório com as instâncias criadas pelo sistema na ontologia, concluí-se que a base de conhecimento da ontologia está bem estruturada e de acordo a informação contida no relatório.

Para os restantes doze relatórios que formam o *corpus* realizou-se o mesmo tipo de testes, podendo afirmar-se que todas as instâncias criadas na ontologia estão estruturadas de acordo com o texto dos relatórios, e que a informação que se pretendia extrair encontra-se instanciada na ontologia não tendo sido detectados quaisquer erros.

Capítulo 9

Conclusão

Neste capítulo são realizadas algumas conclusões ao trabalho desenvolvido e apresentado nos capítulos anteriores desta dissertação. Na secção 9.1 são enumerados alguns aspectos a melhorar num próximo trabalho.

O trabalho desenvolvido e apresentado nesta dissertação, assenta em duas grandes áreas. A área da Web Semântica e as suas linguagens que possuem vocabulário específico que permite representar ontologias, e a área da mineração de textos.

Este trabalho abordou o problema da criação de ontologias como forma de adquirir um conhecimento estruturado de relatórios médicos, que possuem textos em língua natural. Para tal, foi proposta uma metodologia como modo de atingir os objectivos propostos.

Os relatórios médicos fornecidos pelo Hospital do Espírito Santo de Évora, formaram o *corpus* deste trabalho. A colaboração dos profissionais de saúde foi fundamental na atribuição e definição dos termos e seus relacionamentos, que funcionou como ponto de partida para a criação da ontologia e obtenção do respectivo domínio.

A ontologia gerada envolve um conjunto de classes, propriedades e instâncias, e está representada na linguagem OWL. A ontologia foi construída de raiz, devido à escassez de trabalhos desenvolvidos neste âmbito, muito provavelmente, devido à Web Semântica ser uma área recente.

Para extrair a informação dos relatórios médicos e preencher a ontologia, foram utilizadas técnicas de processamento de linguagem natural e aprendizagem automática. No entanto, devido à especificidade do vocabulário usado na descrição dos exames e da conclusão dos relatórios, foram desenvolvidos um conjunto de algoritmos que permitem relacionar os termos médicos. O desenvolvimento destes algoritmos veio tornar-se fundamental, de modo a conseguir-se representar o texto dos relatórios na ontologia.

Foi desenvolvida uma aplicação que permite converter relatórios neurovasculares do formato digital para o formato OWL. Esta aplicação é de fácil utilização e não requer grandes conhecimentos de informática aos utilizadores, que serão fundamentalmente os técnicos e médicos que executam os exames e realizam os relatórios.

Por fim, foram realizados diversos testes, como forma de verificar que as instâncias criadas na ontologia estão estruturadas de acordo com o texto presente nos relatórios. Os resultados obtidos para o domínio dos relatórios do *corpus* são muito positivos e vão ao encontro dos objectivos propostos.

9.1 Trabalho Futuro

Ao longo do desenvolvimento desta dissertação e do trabalho aqui apresentado, foram identificadas algumas melhorias que se podem aplicar num trabalho futuro. Entre elas destacam-se:

- O sistema de extracção de informação foi pensado para extrair informação apenas para os tipos de relatórios fornecidos, como representado no anexo A, e no formato digital presente no apêndice B. Criar um sistema de extracção de informação que independentemente do modo como a informação está disposta nos relatórios, conseguisse extrair a informação, permitiria que este sistema facilmente se adapta-se a qualquer tipo de relatório neurovascular produzido por uma qualquer instituição;
- A estrutura da hierarquia de classes presente na ontologia, corresponde à informação disponibilizada pelo domínio dos relatórios que compõe o *corpus*. Ou seja, poderão existir outros termos médicos presentes neste tipo de relatório neurovascular que não fazem parte do domínio da ontologia. Para solucionar este problema, podia-se recorrer a fontes externas tais como o UMLS, a Wikipedia, entre outros, que dado um termo do relatório este era identificado como termo médico ou não, permitindo adicionar conhecimento à ontologia e tornar o seu domínio o mais completo possível;
- Apresentar os resultados a especialistas, para que a sua avaliação permita transformar o uso de ontologias em relatórios médicos o mais útil possível.

Apêndice A

Código da Ontologia

Este código na linguagem OWL, corresponde à ontologia do trabalho descrito nesta dissertação antes de se adicionar as instâncias.

```
<!DOCTYPE rdf:RDF [  
  <!ENTITY owl "http://www.w3.org/2002/07/owl#">  
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#">  
  <!ENTITY owl2xml "http://www.w3.org/2006/12/owl2-xml#">  
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#">  
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#">  
>  
  
<rdf:RDF xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"  
  xmlns:owl2xml="http://www.w3.org/2006/12/owl2-xml#"  
  xmlns:owl="http://www.w3.org/2002/07/owl#"  
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"  
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
>  
  
<owl:Ontology rdf:about=>  
  <rdfs:comment  
    >Ontologia de Relatórios Neurovasculares</rdfs:comment>  
</owl:Ontology>  
  
< ===== >  
<!-- Object Properties -->  
< ===== >  
  
<owl:ObjectProperty rdf:about="#apresenta">  
  <rdfs:domain rdf:resource="#Artéria"/>  
  <rdfs:range rdf:resource="#Bifurcação"/>  
</owl:ObjectProperty>
```

```

<owl:ObjectProperty rdf:about="#constituída">
  <rdfs:domain rdf:resource="#Artéria"/>
  <rdfs:range rdf:resource="#Lado"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#contém">
  <rdfs:range rdf:resource="#Paciente"/>
  <rdfs:domain rdf:resource="#Relatório"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#corresponde">
  <rdfs:domain rdf:resource="#Diagnóstico"/>
  <rdfs:range rdf:resource="#Zona"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#eh_caracterizado_por">
  <rdfs:range rdf:resource="#Exame"/>
  <rdfs:domain rdf:resource="#Relatório"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#eh_composta">
  <rdfs:domain rdf:resource="#Artéria"/>
  <rdfs:range rdf:resource="#Stent"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#eh_correspondente">
  <rdfs:domain rdf:resource="#Observação"/>
  <rdfs:range rdf:resource="#Zona"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#eh_executado_por">
  <rdfs:domain rdf:resource="#Exame"/>
  <rdfs:range rdf:resource="#Técnico"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#eh_formado">
  <rdfs:domain rdf:resource="#Artéria"/>
  <rdfs:range rdf:resource="#Parede_Arterial"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#eh_realizado_através">
  <rdfs:range rdf:resource="#Aparelho"/>
  <rdfs:domain rdf:resource="#Exame"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#eh_realizado_por">
  <rdfs:range rdf:resource="#Médico"/>
  <rdfs:domain rdf:resource="#Relatório"/>
</owl:ObjectProperty>

```

```

<owl:ObjectProperty rdf:about="#eh_suportado_por">
  <rdfs:domain rdf:resource="#Diagnóstico"/>
  <rdfs:range rdf:resource="#Observação"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#encontra-se">
  <rdfs:domain rdf:resource="#Paciente"/>
  <rdfs:range rdf:resource="#Serviço"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#executa">
  <rdfs:range rdf:resource="#Exame"/>
  <rdfs:domain rdf:resource="#Técnico"/>
  <owl:inverseOf rdf:resource="#eh_executado_por"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#gera">
  <rdfs:range rdf:resource="#Conclusão"/>
  <rdfs:domain rdf:resource="#Relatório"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#indica">
  <rdfs:domain rdf:resource="#Diagnóstico"/>
  <rdfs:range rdf:resource="#Entidade_Anatômica"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#obtida_através_de">
  <rdfs:domain rdf:resource="#Conclusão"/>
  <rdfs:range rdf:resource="#Diagnóstico"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#obtida_por">
  <rdfs:domain rdf:resource="#Conclusão"/>
  <rdfs:range rdf:resource="#Observação"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#obtém">
  <rdfs:range rdf:resource="#Diagnóstico"/>
  <rdfs:domain rdf:resource="#Exame"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#possui">
  <rdfs:domain rdf:resource="#Exame"/>
  <rdfs:range rdf:resource="#Observação"/>
</owl:ObjectProperty>

```

```

<owl:ObjectProperty rdf:about="#realiza">
  <rdfs:domain rdf:resource="#Médico"/>
  <rdfs:range rdf:resource="#Relatório"/>
  <owl:inverseOf rdf:resource="#eh_realizado_por"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#refere">
  <rdfs:range rdf:resource="#Entidade_Anatómica"/>
  <rdfs:domain rdf:resource="#Zona"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#refere-se">
  <rdfs:range rdf:resource="#Entidade_Anatómica"/>
  <rdfs:domain rdf:resource="#Observação"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#tem">
  <rdfs:range rdf:resource="#Informação_Clínica"/>
  <rdfs:domain rdf:resource="#Paciente"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#usa">
  <rdfs:range rdf:resource="#Janela_Ultra-Sônica"/>
  <rdfs:domain rdf:resource="#Triplex_Scan_Transcraneano"/>
</owl:ObjectProperty>

< ===== >
<!-- Data properties -->
< ===== >

<owl:DatatypeProperty rdf:about="#artériaCN">
  <rdfs:comment
    >Nome CN do UMLS para a Artéria</rdfs:comment>
  <rdfs:domain rdf:resource="#Artéria"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#artériaCUI">
  <rdfs:comment
    >Código CUI do UMLS para a Artéria</rdfs:comment>
  <rdfs:domain rdf:resource="#Artéria"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

```

```

<owl:DatatypeProperty rdf:about="#bifurcaçãoCN">
  <rdfs:comment
    >Nome CN do UMLS para a Bifurcação</rdfs:comment>
  <rdfs:domain rdf:resource="#Bifurcação"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#bifurcaçãoCUI">
  <rdfs:comment
    >Código CUI do UMLS para a Bifurcação</rdfs:comment>
  <rdfs:domain rdf:resource="#Bifurcação"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#data">
  <rdfs:comment
    >Data do Relatório</rdfs:comment>
  <rdfs:domain rdf:resource="#Relatório"/>
  <rdfs:range rdf:resource="&xsd:date"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#descrição">
  <rdfs:comment
    >Descrição da Conclusão</rdfs:comment>
  <rdfs:domain rdf:resource="#Conclusão"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#descriçãoInformaçãoClínica">
  <rdfs:comment
    >Descrição da Informação Clínica do Paciente</rdfs:comment>
  <rdfs:domain rdf:resource="#Informação_Clínica"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#diagnósticoCN">
  <rdfs:comment
    >Nome CN do UMLS para o Diagnóstico</rdfs:comment>
  <rdfs:domain rdf:resource="#Diagnóstico"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#diagnósticoCUI">
  <rdfs:comment
    >Código CUI do UMLS para o Diagnóstico</rdfs:comment>
  <rdfs:domain rdf:resource="#Diagnóstico"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

```



```

<owl:DatatypeProperty rdf:about="#idadePaciente">
  <rdfs:comment
    >Idade do Paciente</rdfs:comment>
  <rdfs:domain rdf:resource="#Paciente"/>
  <rdfs:range rdf:resource="&xsd;integer"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#local">
  <rdfs:comment
    >Local onde se realiza o Relatório</rdfs:comment>
  <rdfs:domain rdf:resource="#Relatório"/>
  <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#marca">
  <rdfs:comment
    >Marca do Aparelho usado no Exame</rdfs:comment>
  <rdfs:domain rdf:resource="#Aparelho"/>
  <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#nomeMédico">
  <rdfs:comment>Nome do Médico</rdfs:comment>
  <rdfs:domain rdf:resource="#Médico"/>
  <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#nomePaciente">
  <rdfs:comment>Nome do Paciente</rdfs:comment>
  <rdfs:domain rdf:resource="#Paciente"/>
  <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#nomeServiço">
  <rdfs:comment>Nome do Serviço</rdfs:comment>
  <rdfs:domain rdf:resource="#Serviço"/>
  <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#nomeTécnico">
  <rdfs:comment>Nome do Técnico</rdfs:comment>
  <rdfs:domain rdf:resource="#Técnico"/>
  <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#númeroProcessoPaciente">
  <rdfs:comment>Número do Processo do Paciente</rdfs:comment>
  <rdfs:domain rdf:resource="#Paciente"/>
  <rdfs:range rdf:resource="&xsd;integer"/>
</owl:DatatypeProperty>

```

```

<owl:DatatypeProperty rdf:about="#númeroRelatório">
  <rdfs:comment>Número do Relatório</rdfs:comment>
  <rdfs:domain rdf:resource="#Relatório"/>
  <rdfs:range rdf:resource="&xsd;integer"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#observaçãoCN">
  <rdfs:comment
    >Nome CN do UMLS para a Observação</rdfs:comment>
  <rdfs:domain rdf:resource="#Observação"/>
  <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#observaçãoCUI">
  <rdfs:comment
    >Código CUI do UMLS para a Observação</rdfs:comment>
  <rdfs:domain rdf:resource="#Observação"/>
  <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#paredeArterialCN">
  <rdfs:comment
    >Nome CN do UMLS para a Parede Arterial</rdfs:comment>
  <rdfs:domain rdf:resource="#Parede_Arterial"/>
  <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#paredeArterialCUI">
  <rdfs:comment
    >Código CUI do UMLS para a Parede Arterial</rdfs:comment>
  <rdfs:domain rdf:resource="#Parede_Arterial"/>
  <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#stentCN">
  <rdfs:comment>Nome CN do UMLS para o Stent</rdfs:comment>
  <rdfs:domain rdf:resource="#Stent"/>
  <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="#stentCUI">
  <rdfs:comment>Código CUI do UMLS para o Stent</rdfs:comment>
  <rdfs:domain rdf:resource="#Stent"/>
  <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

```

```

<owl:DatatypeProperty rdf:about="#transdutor">
  <rdfs:comment>Transdutor usado no Aparelho</rdfs:comment>
  <rdfs:domain rdf:resource="#Aparelho"/>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

```

```

< ===== >
<!-- Classes -->
< ===== >

```

```

<owl:Class rdf:about="#Anterior">
  <rdfs:subClassOf rdf:resource="#Cerebral"/>
  <owl:disjointWith rdf:resource="#Comunicante_Anterior"/>
  <owl:disjointWith rdf:resource="#Média"/>
  <owl:disjointWith rdf:resource="#Posterior"/>
</owl:Class>

```

```

<owl:Class rdf:about="#Aparelho">
  <owl:disjointWith rdf:resource="#Bifurcação"/>
  <owl:disjointWith rdf:resource="#Conclusão"/>
  <owl:disjointWith rdf:resource="#Diagnóstico"/>
  <owl:disjointWith rdf:resource="#Entidade_Anatômica"/>
  <owl:disjointWith rdf:resource="#Exame"/>
  <owl:disjointWith rdf:resource="#Informação_Clínica"/>
  <owl:disjointWith rdf:resource="#Janela_Ultra-Sônica"/>
  <owl:disjointWith rdf:resource="#Lado"/>
  <owl:disjointWith rdf:resource="#Médico"/>
  <owl:disjointWith rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Paciente"/>
  <owl:disjointWith rdf:resource="#Parede_Arterial"/>
  <owl:disjointWith rdf:resource="#Relatório"/>
  <owl:disjointWith rdf:resource="#Serviço"/>
  <owl:disjointWith rdf:resource="#Stent"/>
  <owl:disjointWith rdf:resource="#Técnico"/>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment
    >Aparelho usado para realizar os exames</rdfs:comment>
</owl:Class>

```

```

<owl:Class rdf:about="#Aplanamento_do_Espectro_de_Doppler">
  <rdfs:subClassOf rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Ateromatose_Carotídea"/>
  <owl:disjointWith rdf:resource="#Espessamento_Ateromatoso_Difuso"/>
  <owl:disjointWith rdf:resource="#Fenómeno"/>
  <owl:disjointWith rdf:resource="#Fluxo"/>
  <owl:disjointWith rdf:resource="#Hipoplasia"/>

```

```

    <owl:disjointWith rdf:resource="#IP"/>
    <owl:disjointWith rdf:resource="#Impossível_Visualizar"/>
    <owl:disjointWith rdf:resource="#Permeável"/>
    <owl:disjointWith rdf:resource="#Placa_Ateromatosa"/>
    <owl:disjointWith rdf:resource="#Tortuosidade"/>
</owl:Class>

<owl:Class rdf:about="#Artéria">
  <rdfs:subClassOf rdf:resource="#Vaso_Sanguíneo"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#constituída"/>
      <owl:onClass rdf:resource="#Lado"/>
      <owl:maxQualifiedCardinality rdf:datatype=
        "&xsd;nonNegativeInteger">2</owl:maxQualifiedCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#eh_formado"/>
      <owl:someValuesFrom rdf:resource="#Parede_Arterial"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#eh_composta"/>
      <owl:onClass rdf:resource="#Stent"/>
      <owl:maxQualifiedCardinality rdf:datatype=
        "&xsd;nonNegativeInteger">1</owl:maxQualifiedCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#apresenta"/>
      <owl:someValuesFrom rdf:resource="#Bifurcação"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:about="#Ateromatose_Carotídea">
  <rdfs:subClassOf rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Espessamento_Ateromatoso_Difuso"/>
  <owl:disjointWith rdf:resource="#Fenómeno"/>
  <owl:disjointWith rdf:resource="#Fluxo"/>
  <owl:disjointWith rdf:resource="#Hipoplasia"/>
  <owl:disjointWith rdf:resource="#IP"/>
  <owl:disjointWith rdf:resource="#Impossível_Visualizar"/>
  <owl:disjointWith rdf:resource="#Permeável"/>
  <owl:disjointWith rdf:resource="#Placa_Ateromatosa"/>

```

```

    <owl:disjointWith rdf:resource="#Tortuosidade"/>
</owl:Class>

<owl:Class rdf:about="#Aumentado">
    <rdfs:subClassOf rdf:resource="#IP"/>
</owl:Class>

<owl:Class rdf:about="#Aumento_da_Velocidade">
    <rdfs:subClassOf rdf:resource="#Fluxo"/>
    <owl:disjointWith rdf:resource="#Ausência"/>
    <owl:disjointWith rdf:resource="#Inversão_do_Sentido"/>
    <owl:disjointWith rdf:resource="#Normal"/>
    <owl:disjointWith rdf:resource="#Reabilitação"/>
</owl:Class>

<owl:Class rdf:about="#Ausência">
    <rdfs:subClassOf rdf:resource="#Fluxo"/>
    <owl:disjointWith rdf:resource="#Inversão_do_Sentido"/>
    <owl:disjointWith rdf:resource="#Normal"/>
    <owl:disjointWith rdf:resource="#Reabilitação"/>
</owl:Class>

<owl:Class rdf:about="#Basilar">
    <rdfs:subClassOf rdf:resource="#Artéria"/>
    <owl:disjointWith rdf:resource="#Carótida_ou_Carótida_Comum"/>
    <owl:disjointWith rdf:resource="#Central_da_Retina/Oftálmica"/>
    <owl:disjointWith rdf:resource="#Cerebral"/>
    <owl:disjointWith rdf:resource="#Subclávia"/>
    <owl:disjointWith rdf:resource="#Vertebral"/>
</owl:Class>

<owl:Class rdf:about="#Bifurcação">
    <owl:disjointWith rdf:resource="#Conclusão"/>
    <owl:disjointWith rdf:resource="#Diagnóstico"/>
    <owl:disjointWith rdf:resource="#Entidade_Anatômica"/>
    <owl:disjointWith rdf:resource="#Exame"/>
    <owl:disjointWith rdf:resource="#Informação_Clínica"/>
    <owl:disjointWith rdf:resource="#Janela_Ultra-Sônica"/>
    <owl:disjointWith rdf:resource="#Lado"/>
    <owl:disjointWith rdf:resource="#Médico"/>
    <owl:disjointWith rdf:resource="#Observação"/>
    <owl:disjointWith rdf:resource="#Paciente"/>
    <owl:disjointWith rdf:resource="#Parede_Arterial"/>
    <owl:disjointWith rdf:resource="#Relatório"/>
    <owl:disjointWith rdf:resource="#Serviço"/>
    <owl:disjointWith rdf:resource="#Stent"/>
    <owl:disjointWith rdf:resource="#Técnico"/>
    <owl:disjointWith rdf:resource="#Zona"/>
    <rdfs:comment

```



```

    >Bifurcações das Artérias</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Cabeça_e_Pescoço">
    <rdfs:subClassOf rdf:resource="#Entidade_Anatômica"/>
</owl:Class>

<owl:Class rdf:about="#Calcificada">
    <rdfs:subClassOf rdf:resource="#Calcificação"/>
    <owl:disjointWith rdf:resource="#Não_Calcificada"/>
</owl:Class>

<owl:Class rdf:about="#Calcificação">
    <rdfs:subClassOf rdf:resource="#Placa_Ateromatosa"/>
    <owl:disjointWith rdf:resource="#Ecogeneidade"/>
    <owl:disjointWith rdf:resource="#Homogeneidade"/>
    <owl:disjointWith rdf:resource="#Regularidade"/>
</owl:Class>

<owl:Class rdf:about="#Carótida_ou_Carótida_Comum">
    <rdfs:subClassOf rdf:resource="#Artéria"/>
    <owl:disjointWith rdf:resource="#Central_da_Retina/Oftálmica"/>
    <owl:disjointWith rdf:resource="#Cerebral"/>
    <owl:disjointWith rdf:resource="#Subclávia"/>
    <owl:disjointWith rdf:resource="#Vertebral"/>
</owl:Class>

<owl:Class rdf:about="#Central_da_Retina/Oftálmica">
    <rdfs:subClassOf rdf:resource="#Artéria"/>
    <owl:disjointWith rdf:resource="#Cerebral"/>
    <owl:disjointWith rdf:resource="#Subclávia"/>
    <owl:disjointWith rdf:resource="#Vertebral"/>
</owl:Class>

<owl:Class rdf:about="#Circulação_Colateral">
    <rdfs:subClassOf rdf:resource="#Diagnóstico"/>
    <owl:disjointWith rdf:resource="#Estenose"/>
    <owl:disjointWith rdf:resource="#Inelastecidade_Arterial"/>
    <owl:disjointWith rdf:resource="#Oclusão"/>
    <owl:disjointWith rdf:resource="#Sem_Repercussão_Hemodinâmica"/>
    <owl:disjointWith rdf:resource="#Sem_Seguro_Significado_Patológico"/>
</owl:Class>

<owl:Class rdf:about="#Comunicante_Anterior">
    <rdfs:subClassOf rdf:resource="#Cerebral"/>
    <owl:disjointWith rdf:resource="#Média"/>
    <owl:disjointWith rdf:resource="#Posterior"/>
</owl:Class>

```

```

<owl:Class rdf:about="#Conclusão">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#obtida_por"/>
      <owl:someValuesFrom rdf:resource="#Observação"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#obtida_atraves_de"/>
      <owl:someValuesFrom rdf:resource="#Diagnóstico"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <owl:disjointWith rdf:resource="#Diagnóstico"/>
  <owl:disjointWith rdf:resource="#Entidade_Anatómica"/>
  <owl:disjointWith rdf:resource="#Exame"/>
  <owl:disjointWith rdf:resource="#Informação_Clínica"/>
  <owl:disjointWith rdf:resource="#Janela_Ultra-Sônica"/>
  <owl:disjointWith rdf:resource="#Lado"/>
  <owl:disjointWith rdf:resource="#Médico"/>
  <owl:disjointWith rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Paciente"/>
  <owl:disjointWith rdf:resource="#Parede_Arterial"/>
  <owl:disjointWith rdf:resource="#Relatório"/>
  <owl:disjointWith rdf:resource="#Serviço"/>
  <owl:disjointWith rdf:resource="#Stent"/>
  <owl:disjointWith rdf:resource="#Técnico"/>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment
    >Conclusão do Relatório</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Diagnóstico">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#indica"/>
      <owl:someValuesFrom rdf:resource="#Entidade_Anatómica"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#corresponde"/>
      <owl:onClass rdf:resource="#Zona"/>
      <owl:maxQualifiedCardinality rdf:datatype=
        "&xsd;nonNegativeInteger">1</owl:maxQualifiedCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>

```

```

    <owl:Restriction>
      <owl:onProperty rdf:resource="#eh_suportado_por"/>
      <owl:someValuesFrom rdf:resource="#Observação"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <owl:disjointWith rdf:resource="#Entidade_Anatômica"/>
  <owl:disjointWith rdf:resource="#Exame"/>
  <owl:disjointWith rdf:resource="#Informação_Clínica"/>
  <owl:disjointWith rdf:resource="#Janela_Ultra-Sônica"/>
  <owl:disjointWith rdf:resource="#Lado"/>
  <owl:disjointWith rdf:resource="#Médico"/>
  <owl:disjointWith rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Paciente"/>
  <owl:disjointWith rdf:resource="#Parede_Arterial"/>
  <owl:disjointWith rdf:resource="#Relatório"/>
  <owl:disjointWith rdf:resource="#Serviço"/>
  <owl:disjointWith rdf:resource="#Stent"/>
  <owl:disjointWith rdf:resource="#Técnico"/>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment
    >Diagnósticos das observações realizadas</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Direito">
  <rdfs:subClassOf rdf:resource="#Lado"/>
  <owl:disjointWith rdf:resource="#Esquerdo"/>
</owl:Class>

<owl:Class rdf:about="#Eco-Doppler_colorido">
  <rdfs:subClassOf rdf:resource="#Ultra-Sonografia"/>
</owl:Class>

<owl:Class rdf:about="#Ecogeneidade">
  <rdfs:subClassOf rdf:resource="#Placa_Ateromatosa"/>
  <owl:disjointWith rdf:resource="#Homogeneidade"/>
  <owl:disjointWith rdf:resource="#Regularidade"/>
</owl:Class>

<owl:Class rdf:about="#Ecografo">
  <rdfs:subClassOf rdf:resource="#Aparelho"/>
</owl:Class>

<owl:Class rdf:about="#Entidade_Anatômica">
  <owl:disjointWith rdf:resource="#Exame"/>
  <owl:disjointWith rdf:resource="#Informação_Clínica"/>
  <owl:disjointWith rdf:resource="#Janela_Ultra-Sônica"/>
  <owl:disjointWith rdf:resource="#Lado"/>
  <owl:disjointWith rdf:resource="#Médico"/>
  <owl:disjointWith rdf:resource="#Observação"/>

```

```

<owl:disjointWith rdf:resource="#Paciente"/>
<owl:disjointWith rdf:resource="#Parede_Arterial"/>
<owl:disjointWith rdf:resource="#Relatório"/>
<owl:disjointWith rdf:resource="#Serviço"/>
<owl:disjointWith rdf:resource="#Stent"/>
<owl:disjointWith rdf:resource="#Técnico"/>
<owl:disjointWith rdf:resource="#Zona"/>
<rdfs:comment
  >Entidades Anatômicas analisadas em Relatórios Neurovasculares</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Espessamento_Ateromatoso_Difuso">
  <rdfs:subClassOf rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Fenômeno"/>
  <owl:disjointWith rdf:resource="#Fluxo"/>
  <owl:disjointWith rdf:resource="#Hipoplasia"/>
  <owl:disjointWith rdf:resource="#IP"/>
  <owl:disjointWith rdf:resource="#Impossível_Visualizar"/>
  <owl:disjointWith rdf:resource="#Permeável"/>
  <owl:disjointWith rdf:resource="#Placa_Ateromatosa"/>
  <owl:disjointWith rdf:resource="#Tortuosidade"/>
</owl:Class>

<owl:Class rdf:about="#Esquerdo">
  <rdfs:subClassOf rdf:resource="#Lado"/>
</owl:Class>

<owl:Class rdf:about="#Estenose">
  <rdfs:subClassOf rdf:resource="#Diagnóstico"/>
  <owl:disjointWith rdf:resource="#Inelastecidade_Arterial"/>
  <owl:disjointWith rdf:resource="#Oclusão"/>
  <owl:disjointWith rdf:resource="#Sem_Repercussão_Hemodinâmica"/>
  <owl:disjointWith rdf:resource="#Sem_Seguro_Significado_Patológico"/>
</owl:Class>

<owl:Class rdf:about="#Exame">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#eh_executado_por"/>
      <owl:someValuesFrom rdf:resource="#Técnico"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#possui"/>
      <owl:someValuesFrom rdf:resource="#Observação"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>

```

```

    <owl:Restriction>
      <owl:onProperty rdf:resource="#eh_realizado_através"/>
      <owl:onClass rdf:resource="#Aparelho"/>
      <owl:maxQualifiedCardinality rdf:datatype="
        &xsd;nonNegativeInteger">1</owl:maxQualifiedCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#obtem"/>
      <owl:someValuesFrom rdf:resource="#Diagnóstico"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <owl:disjointWith rdf:resource="#Informação_Clínica"/>
  <owl:disjointWith rdf:resource="#Janela_Ultra-Sônica"/>
  <owl:disjointWith rdf:resource="#Lado"/>
  <owl:disjointWith rdf:resource="#Médico"/>
  <owl:disjointWith rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Paciente"/>
  <owl:disjointWith rdf:resource="#Parede_Arterial"/>
  <owl:disjointWith rdf:resource="#Relatório"/>
  <owl:disjointWith rdf:resource="#Serviço"/>
  <owl:disjointWith rdf:resource="#Stent"/>
  <owl:disjointWith rdf:resource="#Técnico"/>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment
    >Exames do Relatório</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Externa">
  <rdfs:subClassOf rdf:resource="#Carótida_ou_Carótida_Comum"/>
  <owl:disjointWith rdf:resource="#Interna"/>
</owl:Class>

<owl:Class rdf:about="#Fenómeno">
  <rdfs:subClassOf rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Fluxo"/>
  <owl:disjointWith rdf:resource="#Hipoplasia"/>
  <owl:disjointWith rdf:resource="#IP"/>
  <owl:disjointWith rdf:resource="#Impossível_Visualizar"/>
  <owl:disjointWith rdf:resource="#Permeável"/>
  <owl:disjointWith rdf:resource="#Placa_Ateromatosa"/>
  <owl:disjointWith rdf:resource="#Tortuosidade"/>
</owl:Class>

<owl:Class rdf:about="#Fluxo">
  <rdfs:subClassOf rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Hipoplasia"/>
  <owl:disjointWith rdf:resource="#IP"/>

```



```

    <owl:disjointWith rdf:resource="#Impossível_Visualizar"/>
    <owl:disjointWith rdf:resource="#Permeável"/>
    <owl:disjointWith rdf:resource="#Placa_Ateromatosa"/>
    <owl:disjointWith rdf:resource="#Tortuosidade"/>
</owl:Class>

<owl:Class rdf:about="#Grave">
    <rdfs:subClassOf rdf:resource="#Estenose"/>
    <owl:disjointWith rdf:resource="#Ligeira"/>
    <owl:disjointWith rdf:resource="#Moderada"/>
    <owl:disjointWith rdf:resource="#Pré-Oclusiva"/>
    <owl:disjointWith rdf:resource="#Remodelação"/>
    <owl:disjointWith rdf:resource="#Residual"/>
</owl:Class>

<owl:Class rdf:about="#Heterogónea">
    <rdfs:subClassOf rdf:resource="#Homogeneidade"/>
    <owl:disjointWith rdf:resource="#Homogénea"/>
</owl:Class>

<owl:Class rdf:about="#Hiperecogénica">
    <rdfs:subClassOf rdf:resource="#Ecogeneidade"/>
    <owl:disjointWith rdf:resource="#Hipoecogénica"/>
    <owl:disjointWith rdf:resource="#Isoecogénica"/>
</owl:Class>

<owl:Class rdf:about="#Hipoecogénica">
    <rdfs:subClassOf rdf:resource="#Ecogeneidade"/>
    <owl:disjointWith rdf:resource="#Isoecogénica"/>
</owl:Class>

<owl:Class rdf:about="#Hipoplasia">
    <rdfs:subClassOf rdf:resource="#Observação"/>
    <owl:disjointWith rdf:resource="#IP"/>
    <owl:disjointWith rdf:resource="#Impossível_Visualizar"/>
    <owl:disjointWith rdf:resource="#Permeável"/>
    <owl:disjointWith rdf:resource="#Placa_Ateromatosa"/>
    <owl:disjointWith rdf:resource="#Tortuosidade"/>
</owl:Class>

<owl:Class rdf:about="#Homogeneidade">
    <rdfs:subClassOf rdf:resource="#Placa_Ateromatosa"/>
    <owl:disjointWith rdf:resource="#Regularidade"/>
</owl:Class>

<owl:Class rdf:about="#Homogénea">
    <rdfs:subClassOf rdf:resource="#Homogeneidade"/>
</owl:Class>

```

```

<owl:Class rdf:about="#IP">
  <rdfs:subClassOf rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Impossível_Visualizar"/>
  <owl:disjointWith rdf:resource="#Permeável"/>
  <owl:disjointWith rdf:resource="#Placa_Ateromatosa"/>
  <owl:disjointWith rdf:resource="#Tortuosidade"/>
  <rdfs:comment
    >Índice de Pulsatilidade</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Impossível_Visualizar">
  <rdfs:subClassOf rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Permeável"/>
  <owl:disjointWith rdf:resource="#Placa_Ateromatosa"/>
  <owl:disjointWith rdf:resource="#Tortuosidade"/>
</owl:Class>

<owl:Class rdf:about="#Inelastecidade_Arterial">
  <rdfs:subClassOf rdf:resource="#Diagnóstico"/>
  <owl:disjointWith rdf:resource="#Oclusão"/>
  <owl:disjointWith rdf:resource="#Sem_Repercussão_Hemodinâmica"/>
  <owl:disjointWith rdf:resource="#Sem_Seguro_Significado_Patológico"/>
</owl:Class>

<owl:Class rdf:about="#Informação_Clínica">
  <owl:disjointWith rdf:resource="#Janela_Ultra-Sônica"/>
  <owl:disjointWith rdf:resource="#Lado"/>
  <owl:disjointWith rdf:resource="#Médico"/>
  <owl:disjointWith rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Paciente"/>
  <owl:disjointWith rdf:resource="#Parede_Arterial"/>
  <owl:disjointWith rdf:resource="#Relatório"/>
  <owl:disjointWith rdf:resource="#Serviço"/>
  <owl:disjointWith rdf:resource="#Stent"/>
  <owl:disjointWith rdf:resource="#Técnico"/>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment
    >Informação Clínica do Paciente</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Interna">
  <rdfs:subClassOf rdf:resource="#Carótida_ou_Carótida_Comum"/>
</owl:Class>

<owl:Class rdf:about="#Inversão_do_Sentido">
  <rdfs:subClassOf rdf:resource="#Fluxo"/>
  <owl:disjointWith rdf:resource="#Normal"/>
  <owl:disjointWith rdf:resource="#Reabilitação"/>
</owl:Class>

```

```

<owl:Class rdf:about="#Irregular">
  <rdfs:subClassOf rdf:resource="#Regularidade"/>
  <owl:disjointWith rdf:resource="#Regular"/>
</owl:Class>

<owl:Class rdf:about="#Isoecogénica">
  <rdfs:subClassOf rdf:resource="#Ecogeneidade"/>
</owl:Class>

<owl:Class rdf:about="#Janela_Ultra-Sônica">
  <owl:disjointWith rdf:resource="#Lado"/>
  <owl:disjointWith rdf:resource="#Médico"/>
  <owl:disjointWith rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Paciente"/>
  <owl:disjointWith rdf:resource="#Parede_Arterial"/>
  <owl:disjointWith rdf:resource="#Relatório"/>
  <owl:disjointWith rdf:resource="#Serviço"/>
  <owl:disjointWith rdf:resource="#Stent"/>
  <owl:disjointWith rdf:resource="#Técnico"/>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment
    >Janelas Ultra-Sônicas usadas no Exame Transcraneano</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Jusante">
  <rdfs:subClassOf rdf:resource="#Zona"/>
  <owl:disjointWith rdf:resource="#Montante"/>
</owl:Class>

<owl:Class rdf:about="#Lado">
  <owl:disjointWith rdf:resource="#Médico"/>
  <owl:disjointWith rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Paciente"/>
  <owl:disjointWith rdf:resource="#Parede_Arterial"/>
  <owl:disjointWith rdf:resource="#Relatório"/>
  <owl:disjointWith rdf:resource="#Serviço"/>
  <owl:disjointWith rdf:resource="#Stent"/>
  <owl:disjointWith rdf:resource="#Técnico"/>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment
    >Lado da artéria observada</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Ligeira">
  <rdfs:subClassOf rdf:resource="#Estenose"/>
  <owl:disjointWith rdf:resource="#Moderada"/>
  <owl:disjointWith rdf:resource="#Pré-Oclusiva"/>
  <owl:disjointWith rdf:resource="#Remodelação"/>
  <owl:disjointWith rdf:resource="#Residual"/>

```

```

</owl:Class>

<owl:Class rdf:about="#Moderada">
  <rdfs:subClassOf rdf:resource="#Estenose"/>
  <owl:disjointWith rdf:resource="#Pré-Oclusiva"/>
  <owl:disjointWith rdf:resource="#Remodelação"/>
  <owl:disjointWith rdf:resource="#Residual"/>
</owl:Class>

<owl:Class rdf:about="#Montante">
  <rdfs:subClassOf rdf:resource="#Zona"/>
</owl:Class>

<owl:Class rdf:about="#Média">
  <rdfs:subClassOf rdf:resource="#Cerebral"/>
  <owl:disjointWith rdf:resource="#Posterior"/>
</owl:Class>

<owl:Class rdf:about="#Médico">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#realiza"/>
      <owl:someValuesFrom rdf:resource="#Relatório"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <owl:disjointWith rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Paciente"/>
  <owl:disjointWith rdf:resource="#Parede_Arterial"/>
  <owl:disjointWith rdf:resource="#Relatório"/>
  <owl:disjointWith rdf:resource="#Serviço"/>
  <owl:disjointWith rdf:resource="#Stent"/>
  <owl:disjointWith rdf:resource="#Técnico"/>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment>
    >Médico que realiza o Relatório</rdfs:comment>
  </rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Neurovascular">
  <rdfs:subClassOf rdf:resource="#Relatório"/>
</owl:Class>

<owl:Class rdf:about="#Normal">
  <rdfs:subClassOf rdf:resource="#Fluxo"/>
  <owl:disjointWith rdf:resource="#Reabilitação"/>
</owl:Class>

<owl:Class rdf:about="#Não_Calcificada">
  <rdfs:subClassOf rdf:resource="#Calcificação"/>
</owl:Class>

```

```

<owl:Class rdf:about="#Observação">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#eh_correspondente"/>
      <owl:onClass rdf:resource="#Zona"/>
      <owl:maxQualifiedCardinality rdf:datatype=
        "&xsd;nonNegativeInteger">1</owl:maxQualifiedCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#refere-se"/>
      <owl:someValuesFrom rdf:resource="#Entidade_Anatômica"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <owl:disjointWith rdf:resource="#Paciente"/>
  <owl:disjointWith rdf:resource="#Parede_Arterial"/>
  <owl:disjointWith rdf:resource="#Relatório"/>
  <owl:disjointWith rdf:resource="#Serviço"/>
  <owl:disjointWith rdf:resource="#Stent"/>
  <owl:disjointWith rdf:resource="#Técnico"/>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment
    >Observações dos Exames</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Oclusão">
  <rdfs:subClassOf rdf:resource="#Diagnóstico"/>
  <owl:disjointWith rdf:resource="#Sem_Repercussão_Hemodinâmica"/>
  <owl:disjointWith rdf:resource="#Sem_Seguro_Significado_Patológico"/>
</owl:Class>

<owl:Class rdf:about="#P1">
  <rdfs:subClassOf rdf:resource="#Posterior"/>
  <owl:disjointWith rdf:resource="#P2"/>
</owl:Class>

<owl:Class rdf:about="#P2">
  <rdfs:subClassOf rdf:resource="#Posterior"/>
</owl:Class>

<owl:Class rdf:about="#Paciente">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#encontra-se"/>
      <owl:onClass rdf:resource="#Serviço"/>
      <owl:maxQualifiedCardinality rdf:datatype=
        "&xsd;nonNegativeInteger">1</owl:maxQualifiedCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>

```



```

</rdfs:subClassOf>
<rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#tem"/>
    <owl:onClass rdf:resource="#Informação_Clínica"/>
    <owl:maxQualifiedCardinality rdf:datatype=
      "&xsd;nonNegativeInteger">1</owl:maxQualifiedCardinality>
  </owl:Restriction>
</rdfs:subClassOf>
<owl:disjointWith rdf:resource="#Parede_Arterial"/>
<owl:disjointWith rdf:resource="#Relatório"/>
<owl:disjointWith rdf:resource="#Serviço"/>
<owl:disjointWith rdf:resource="#Stent"/>
<owl:disjointWith rdf:resource="#Técnico"/>
<owl:disjointWith rdf:resource="#Zona"/>
<rdfs:comment
  >Paciente sobre o qual se realiza o Relatório</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Parede_Arterial">
  <owl:disjointWith rdf:resource="#Relatório"/>
  <owl:disjointWith rdf:resource="#Serviço"/>
  <owl:disjointWith rdf:resource="#Stent"/>
  <owl:disjointWith rdf:resource="#Técnico"/>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment
    >Parede Arterial das artérias</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Permeável">
  <rdfs:subClassOf rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Placa_Ateromatosa"/>
  <owl:disjointWith rdf:resource="#Tortuosidade"/>
</owl:Class>

<owl:Class rdf:about="#Placa_Ateromatosa">
  <rdfs:subClassOf rdf:resource="#Observação"/>
  <owl:disjointWith rdf:resource="#Tortuosidade"/>
</owl:Class>

<owl:Class rdf:about="#Posterior">
  <rdfs:subClassOf rdf:resource="#Cerebral"/>
</owl:Class>

<owl:Class rdf:about="#Pré-Oclusiva">
  <rdfs:subClassOf rdf:resource="#Estenose"/>
  <owl:disjointWith rdf:resource="#Remodelação"/>
  <owl:disjointWith rdf:resource="#Residual"/>
</owl:Class>

```

```

<owl:Class rdf:about="#Reabilitação">
  <rdfs:subClassOf rdf:resource="#Fluxo"/>
</owl:Class>

<owl:Class rdf:about="#Regular">
  <rdfs:subClassOf rdf:resource="#Regularidade"/>
</owl:Class>

<owl:Class rdf:about="#Regularidade">
  <rdfs:subClassOf rdf:resource="#Placa_Ateromatosa"/>
</owl:Class>

<owl:Class rdf:about="#Relatório">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#contém"/>
      <owl:onClass rdf:resource="#Paciente"/>
      <owl:maxQualifiedCardinality rdf:datatype=
        "&xs;nonNegativeInteger">1</owl:maxQualifiedCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#eh_caracterizado_por"/>
      <owl:someValuesFrom rdf:resource="#Exame"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#gera"/>
      <owl:onClass rdf:resource="#Conclusão"/>
      <owl:maxQualifiedCardinality rdf:datatype=
        "&xs;nonNegativeInteger">1</owl:maxQualifiedCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#eh_realizado_por"/>
      <owl:onClass rdf:resource="#Médico"/>
      <owl:maxQualifiedCardinality rdf:datatype=
        "&xs;nonNegativeInteger">1</owl:maxQualifiedCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <owl:disjointWith rdf:resource="#Serviço"/>
  <owl:disjointWith rdf:resource="#Stent"/>
  <owl:disjointWith rdf:resource="#Técnico"/>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment
    >Relatório Neurovascular</rdfs:comment>

```

```

</owl:Class>

<owl:Class rdf:about="#Remodelação">
  <rdfs:subClassOf rdf:resource="#Estenose"/>
  <owl:disjointWith rdf:resource="#Residual"/>
</owl:Class>

<owl:Class rdf:about="#Residual">
  <rdfs:subClassOf rdf:resource="#Estenose"/>
</owl:Class>

<owl:Class rdf:about="#Roubo_da_Subclávia">
  <rdfs:subClassOf rdf:resource="#Fenómeno"/>
</owl:Class>

<owl:Class rdf:about="#Sem_Repercussão_Hemodinâmica">
  <rdfs:subClassOf rdf:resource="#Diagnóstico"/>
  <owl:disjointWith rdf:resource="#Sem_Seguro_Significado_Patológico"/>
</owl:Class>

<owl:Class rdf:about="#Sem_Seguro_Significado_Patológico">
  <rdfs:subClassOf rdf:resource="#Diagnóstico"/>
</owl:Class>

<owl:Class rdf:about="#Serviço">
  <owl:disjointWith rdf:resource="#Stent"/>
  <owl:disjointWith rdf:resource="#Técnico"/>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment
    >Serviço onde está internado o Paciente</rdfs:comment> </owl:Class>

<owl:Class rdf:about="#Stent">
  <owl:disjointWith rdf:resource="#Técnico"/>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment
    >Mecanismo implantado nas artérias</rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Subclávia">
  <rdfs:subClassOf rdf:resource="#Artéria"/>
</owl:Class>

<owl:Class rdf:about="#Suboccipital">
  <rdfs:subClassOf rdf:resource="#Janela_Ultra-Sônica"/>
  <owl:disjointWith rdf:resource="#Transorbital"/>
  <owl:disjointWith rdf:resource="#Transtemporal"/>
</owl:Class>

<owl:Class rdf:about="#Tortuosidade">
  <rdfs:subClassOf rdf:resource="#Observação"/>

```

```

</owl:Class>

<owl:Class rdf:about="#Transorbital">
  <rdfs:subClassOf rdf:resource="#Janela_Ultra-Sônica"/>
  <owl:disjointWith rdf:resource="#Transtemporal"/>
</owl:Class>

<owl:Class rdf:about="#Transtemporal">
  <rdfs:subClassOf rdf:resource="#Janela_Ultra-Sônica"/>
</owl:Class>

<owl:Class rdf:about="#Triplex_Scan_Cervical">
  <rdfs:subClassOf rdf:resource="#Eco-Doppler_colorido"/>
  <owl:disjointWith rdf:resource="#Triplex_Scan_Transcraneano"/>
</owl:Class>

<owl:Class rdf:about="#Triplex_Scan_Transcraneano">
  <rdfs:subClassOf rdf:resource="#Eco-Doppler_colorido"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#usa"/>
      <owl:someValuesFrom rdf:resource="#Janela_Ultra-Sônica"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:about="#Técnico">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#executa"/>
      <owl:someValuesFrom rdf:resource="#Exame"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <owl:disjointWith rdf:resource="#Zona"/>
  <rdfs:comment>
    >Técnico que executa os Exames</rdfs:comment>
  </rdfs:comment>
</owl:Class>

<owl:Class rdf:about="#Ultra-Sonografia">
  <rdfs:subClassOf rdf:resource="#Exame"/>
</owl:Class>

<owl:Class rdf:about="#V0">
  <rdfs:subClassOf rdf:resource="#Vertebral"/>
  <owl:disjointWith rdf:resource="#V1"/>
  <owl:disjointWith rdf:resource="#V2"/>
  <owl:disjointWith rdf:resource="#V3"/>
  <owl:disjointWith rdf:resource="#V4"/>
</owl:Class>

```

```

<owl:Class rdf:about="#V1">
  <rdfs:subClassOf rdf:resource="#Vertebral"/>
  <owl:disjointWith rdf:resource="#V2"/>
  <owl:disjointWith rdf:resource="#V3"/>
  <owl:disjointWith rdf:resource="#V4"/>
</owl:Class>

<owl:Class rdf:about="#V2">
  <rdfs:subClassOf rdf:resource="#Vertebral"/>
  <owl:disjointWith rdf:resource="#V3"/>
  <owl:disjointWith rdf:resource="#V4"/>
</owl:Class>

<owl:Class rdf:about="#V3">
  <rdfs:subClassOf rdf:resource="#Vertebral"/>
  <owl:disjointWith rdf:resource="#V4"/>
</owl:Class>

<owl:Class rdf:about="#V4">
  <rdfs:subClassOf rdf:resource="#Vertebral"/>
</owl:Class>

<owl:Class rdf:about="#Vaso_Sanguíneo">
  <rdfs:subClassOf rdf:resource="#Cabeça_e_Pescoço"/>
</owl:Class>

<owl:Class rdf:about="#Vertebral">
  <rdfs:subClassOf rdf:resource="#Artéria"/>
</owl:Class>

<owl:Class rdf:about="#Zona">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#refere"/>
      <owl:onClass rdf:resource="#Entidade_Anatómica"/>
      <owl:maxQualifiedCardinality rdf:datatype=
        "&xsd;nonNegativeInteger">1</owl:maxQualifiedCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:comment
    >Zona correspondente a uma Observação</rdfs:comment>
</owl:Class>

</rdf:RDF>

```


Apêndice B

Relatório no Formato Digital

O relatório que se segue corresponde ao relatório do anexo A, mas em formato digital (TXT) e com os campos vazios preenchidos.

Laboratório de Ultrassons Cardio e Neurovascular

RELATÓRIO DE NEUROVASCULAR

IDENTIFICAÇÃO DO PACIENTE:

Nº de Processo: 111

Data de exame: 2009-05-17

Nome: Luís Clérigo

Idade: 66

Serviço: UAVC

INFORMAÇÃO CLÍNICA: AVC

TRIPLEX SCAN CERVICAL realizado em Ecografo IE33/HD11 XE da Philips e transdutor L3-12Mhz, revela:

Eixos Carotídeos, Artérias Subclávias e Artérias Vertebrais (V0 a V3) permeáveis, com paredes arteriais e fluxos normais.

TRIPLEX SCAN TRANSCRANEANO (Via Transtemporal e Suboccipital) realizado em Ecografo IE33/HD11 XE da Philips e transdutor S1-5MHz, revela:

Permeabilidade de todos os segmentos arteriais estudados (ACM, ACI, ACA, ACP1, ACP2, AV4 e AB), com fluxos normais.

CONCLUSÃO:

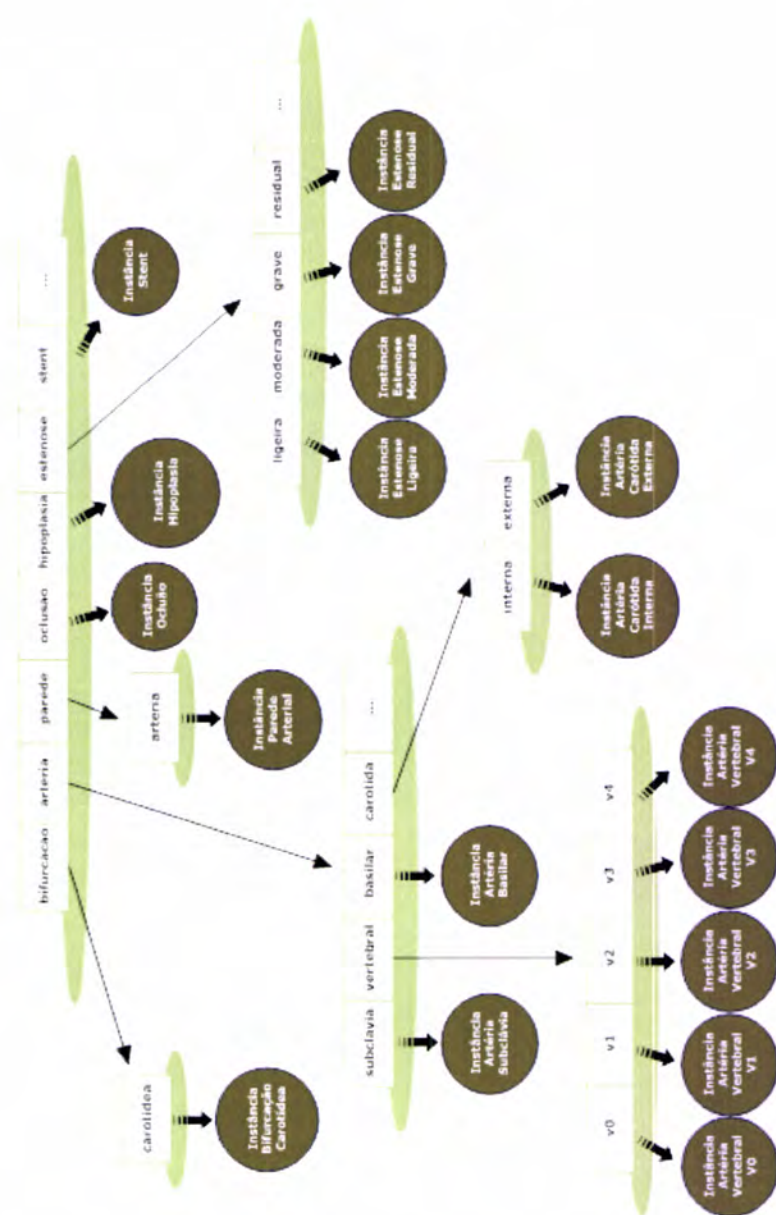
Exames normais.

Técnica: Sónia Mateus/Vanda Pós-de-Mina

Médica: Irene Mendes

Apêndice C

Árvore de Decisão



Anexo A

Relatório Neurovascular



Laboratório de Ultrassons Cardio e Neurovascular

Relatório de Neurovascular

IDENTIFICAÇÃO DO PACIENTE:

Nº de Processo:

Data de exame:

Nome:

Idade: anos

Serviço:

INFORMAÇÃO CLÍNICA: AVC

TRIPLEX SCAN CERVICAL realizado em Ecografo IE33/HD11 XE da Philips e transdutor L3-12Mhz, revela:

Eixos Carotídeos, Artérias Subclávias e Artérias Vertebrais (V0 a V3) permeáveis, com paredes arteriais e fluxos normais.

TRIPLEX SCAN TRANSCRANEANO (Via Transtemporal e Suboccipital) realizado em Ecografo IE33/HD11 XE da Philips e transdutor S1-5MHz, revela:

Permeabilidade de todos os segmentos arteriais estudados (ACM, ACI, ACA, ACP1, ACP2, AV4 e AB), com fluxos normais.

CONCLUSÃO:

Exames normais.

Técnica:

(Cpl: Sônia Mateus/Vanda Pós-de-Mina)

Médico:

(Dra. Irene Mendes)

Bibliografia

- [1] Luana P. Annibal, Natalia A. Rosa, Paulo M. A. Marques, José A. Baranauskas, and Joaquim C. Felipe. Uma ontologia para estruturação da informação contida em laudos radiológicos. In *XI Congresso Brasileiro de Informática em Saúde*, Campos do Jordão, São Paulo, 2008.
- [2] T. Berners-lee, J. Hender, and O. Lassila. The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, May 2001.
- [3] Tim Berners-lee. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*. HarperOne, San Francisco, 1999.
- [4] Tim Berners-lee. *Semantic Web - XML2000*, 2000. Disponível em <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>. Acedido a 20 de Setembro de 2010.
- [5] Olivier Bodenreider. *Integrating Biomedical Terminology*. The Unified Medical Language System (UMLS), 2004. Nucleic Acids Research.
- [6] Karin Breitman. *Web Semântica: A Internet do Futuro*. LTC Editora, Rio de Janeiro, 2005.
- [7] Dan Brickley and R.V. Guha. *Resource Description Framework (RDF) schema specification 1.0*. W3C candidate recommendation, March 2000. Disponível em <http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>. Acedido a 20 de Setembro de 2010.
- [8] Dan Brickley and R.V. Guha. *RDF Vocabulary Description Language 1.0: RDF Schema*. W3C Recommendation, February 2004. Disponível em <http://www.w3.org/TR/rdf-schema/>. Acedido a 20 de Setembro de 2010.
- [9] Allen Brown, Matthew Fuchs, Jonathan Robie, and Philip Wadler. *XML Schema: Formal Description*. Working Draft, September 2001. Disponível em <http://www.w3.org/TR/2001/WD-xmlschema-formal-20010925/>. Acedido a 20 de Setembro de 2010.
- [10] Diego Calvanese, Giuseppe De Giacomo, Ian Horrocks, Carsten Lutz, Boris Motik, Bijan Parsia, and Peter F. Patel-Schneider. *OWL 1.1 Web Ontology Language - Tractable Fragments*. W3C Member Submission, December 2006. Disponível em <http://www.w3.org/Submission/owl11-tractable/>. Acedido a 20 de Setembro de 2010.

- [11] Jaime Cantais, David Dominguez, Valeria Gigante, Loredana Laera, and Valentina Tamma. An example of food ontology for diabetes control. In *Proceedings of the International Semantic Web Conference 2005 workshop on Ontology Patterns for the Semantic Web*, 2005.
- [12] C. S. Cardoso, W. T. Caiaffa, M. Bandeira, A. L. Siqueira, M. N. S. Abreu, and J. O. P. Fonseca. Qualidade de vida e dimensão ocupacional na esquizofrenia: uma comparação por sexo. *Cadernos de Saúde Pública*, 26(6):1303–1314, 2006.
- [13] Jorge Cardoso. The semantic web vision: Where are we? *IEEE Intelligent Systems*, pages 22–26, September/October 2007.
- [14] Jeremy J. Carroll, Ian Dickinson, Chris Dollin, Dave Reynolds, Andy Seaborne, and Kevin Wilkinson. Jena: Implementing the semantic web recommendations. In *Proceedings of 13th International Wouird Wide Web Conference*, New York, USA, May 2004.
- [15] C. Cavaliere and K. C. Chung. Total wrist arthroplasty and total wrist arthrodesis in rheumatoid arthritis: A decision analysis from the hand surgeons perspective. *Journal of Hand Surgery-American Volume*, 33(10):1744–1755, 2008.
- [16] C. L. Chang and C. H. Chen. Applying decision tree and neural network to increase quality of dermatologic diagnosis. *Expert Systems with Applications*, 36(2):4035–4041, 2009.
- [17] M. Cristiani and R. Cuel. A survey on ontology creation methodologies. *International Journal on Semantic Web and Information System*, 1(2):48–68, 2005.
- [18] Renan Figueredo Farias, Merisandra Côrtes de Mattos, and Priscyla Walseska Targino de Azevedo Simões. *Ontologia para a Ggestão do Conhecimento em Saúde por Meio da Metodologia Methontology*. Brasil, 2006.
- [19] Ronen Feldman and James Sanger. *The Text Mining Handbook – Advanced Approaches in Analyzing Unstructured Data*. Cambridge, Cambridge University Press, 2007.
- [20] D. Fensel, I. Horrocks, F. Van Harmelen, S. Decker, M. Erdmann, and M. Klein. Oil in a nutshell. In *Knowledge Acquisition, Modeling, and Management, Proceedings of the European Knowledge Acquisition Conference*, October 2000.
- [21] D. Fensel, W. Wahlster, H. Lieberman, and Jr. Hendler. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. MIT Press, 2002.
- [22] Liliana Ferreira, César Telmo Oliveira, António Teixeira, and João Paulo Silva Cunha. *Extracção de Informação de Relatórios Médicos*. Universidade de Aveiro, 2008.
- [23] C. L. Gillies, P. C. Lambert, K. R. Abrams, A. J. Sutton, N. J. Cooper, R. T. Hsu, M. J. Davies, and K. Khunti. Different strategies for screening and prevention of type 2 diabetes in adults: cost effectiveness analysis. *British Medical Journal*, 336(7654):1–11, 2008.

- [24] Bernardo Cuenca Grau. A possible simplification of the semantic web architecture. In *Proceedings of the 13th international conference on World Wide Web*, New York, NY, USA, 2004.
- [25] Thomas R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:199–220, 1993.
- [26] Nicola Guarino. Formal ontology in information systems. In *Proceedings of FOIS'98*, Trento, Italy, 1998.
- [27] J. Heflin, J. Hendler, and S. Luke. *SHOE: A Knowledge Representation Language for Internet Applications*. Dept. of Computer Science, University of Maryland, 1999.
- [28] Jeff Heflin and James Hendler. Searching the web with shoe. In *AAAI-2000 Workshop on AI for Web Search*, 2000.
- [29] James Hendler. Agents and the semantic web. *IEEE Intelligent Systems*, pages 30–37, Março/Abril 2001.
- [30] Ah hwee Tan. Text mining: The state of the art and the challenges. In *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*, pages 65–70, 1999.
- [31] P. Jackson and I. Mouliner. *Natural Language Processing for Online Applications: Text Retrieval, Extraction & Categorization*. John Benjamins, 2002.
- [32] JENA. *Jena 2 Ontology API*, 2010. Disponível em <http://jena.sourceforge.net/ontology/>. Acedido a 20 de Setembro de 2010.
- [33] Daniel Jurafsky and James H. Martin. *Speech and Language Processing. An Introduction to Natural Language Processing*. Computational Linguistics and Speech Recognition, second edition, 2009.
- [34] Manu Konchady. *Text Mining Application Programming (Programming Series)*. Charles River Media, 2006.
- [35] Ora Lassila and Ralph R. Swick. *Resource Description Framework (RDF) Model and Syntax Specification*. World Wide Web Consortium, 1999. Disponível em <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>. Acedido a 20 de Setembro de 2010.
- [36] Paola Di Maio. *An Open Ontology for Open Source Emergency Response System*. Mae Fa Luang University, Chiang Rai, Thailand, 2007.
- [37] Andreia Malucelli. *Ontology-based Services for Agents Interoperability*. Department of Computing and Electrical Engineering - Faculty of Engineering, University of Porto, 2006.
- [38] Frank Manola and Eric Miller. *RDF Primer*. W3C Recommendation, February 2004. Disponível em <http://www.w3.org/TR/rdf-primer/>. Acedido a 20 de Setembro de 2010.

- [39] Deborah L. McGuinness and Frank van Harmelen. *Web Ontology Language - Overview*. W3C Recommendation, February 2004. Disponível em <http://www.w3.org/TR/owl-features/>. Acedido a 20 de Setembro de 2010.
- [40] Eric Miller. An introduction to the resource description framework. *D-Lib Magazine*, May 1998.
- [41] Eric Miller. Weaving meaning: An overview of the semantic web. In *Presentation at the University of Michigan*, Ann Arbor, Michigan USA, 2004.
- [42] T. Mitchell. *Machine Learning*. [S.l.]: Boston: McGraw-Hill, 1997.
- [43] F. B. Nardon. *Compartilhamento de Conhecimento em Saúde utilizando Ontologias e Banco de Dados Dedutivos*. Escola Politécnica da Universidade de São Paulo, São Paulo, 2003.
- [44] Fabiane Bizinella Nardon. *Compartilhamento de Conhecimento em Saúde Utilizando Ontologias e Banco de Dados Dedutivos*. Escola Politécnica da Universidade de São Paulo, São Paulo, 2003.
- [45] N. Noy and D. McGuinness. *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.
- [46] Natalya Fridman Noy, Ray W. Fergerson, and Mark A. Musen. The knowledge model of protege-2000: Combining interoperability and flexibility. In *2th International Conference on Knowledge Engineering and Knowledge Management*, Juan-les-Pins, France, 2000.
- [47] V. M. Orenço and C. R. Huyck. A stemming algorithm for the portuguese language. In *Proceedings of SPIRES 2001 Symposium on String Processing and Information Retrieval*, Laguna de San Raphael, Chile, 2001.
- [48] A. Paschoal. *Protégé + Ausubel + ITIL: Uma Proposta Ontológica para Compartilhamento de Conhecimento na Secção de Suporte Operacional do Tribunal Superior Eleitoral*. Universidade de Brasília, Brasília, 2005.
- [49] M. Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.
- [50] P. Quaresma, C. Bratsas, P. Bamidis, G. Pangalos, and N. Maglaveras. Knowledge management for medical computational problem solving: An ontological approach. In *2nd Balkan Conference in Informatics*, Ohrid, Macedonia, 2005.
- [51] P. Quaresma, Ch. Bratsas, and N. Maglaveras. A framework to describe problems and algorithms in medical informatics via ontologies. In *4th European Symposium on Biomedical Engineering*, Patras, Greece. University of Patras, 2004.
- [52] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1:81–106, 1986.
- [53] S. O. Rezende, J. B. Pugliesi, E. A. Melanda, and M. F. Paula. Mineração de dados. In *Solange Oliveira Rezende: Sistemas Inteligentes - Fundamentos e Aplicações*, pages 307–335, Manole, 2003.

- [54] José Rocha and Paulo Cortez. *Análise Inteligente de Dados, Algoritmos e Implementação em Java*. FCA, 2008.
- [55] World Wide Web Consortium. *W3C Semantic Web Activity*, 2001. Disponível em <http://www.w3.org/2001/sw/>. Acedido a 20 de Setembro de 2010.

