
Modelo de teses e dissertações em LaTeX do ICMC

Edson Francisco Luque Mamani

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Edson Francisco Luque Mamani

Modelo de teses e dissertações em LaTeX do ICMC

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Dilvan de Abreu Moreira

USP – São Carlos
Junho de 2016

Edson Francisco Luque Mamani

Model of theses and dissertations in LaTeX of the ICMC

Doctoral dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Doctorate Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Dilvan de Abreu Moreira

USP – São Carlos
June 2016

*Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.
Em especial, ao pesquisadores do Instituto de Ciências Matemáticas e de Computação (ICMC).*

AGRADECIMENTOS

Os agradecimentos principais são direcionados à Gerald Weber, Miguel Frasson, Leslie H. Watter, Bruno Parente Lima, Flávio de Vasconcellos Corrêa, Otavio Real Salvador, Renato Machnievscz¹ e todos aqueles que contribuíram para que a produção de trabalhos acadêmicos conforme as normas ABNT com L^AT_EX fosse possível.

Agradecimentos especiais são direcionados ao Centro de Pesquisa em Arquitetura da Informação² da Universidade de Brasília (CPAI), ao grupo de usuários *latex-br*³ e aos novos voluntários do grupo *abnT_EX2*⁴ que contribuíram e que ainda contribuirão para a evolução do *abnT_EX2*.

¹ Os nomes dos integrantes do primeiro projeto *abnT_EX* foram extraídos de <http://codigolivre.org.br/projects/abntex/>

² <http://www.cpai.unb.br/>

³ <http://groups.google.com/group/latex-br>

⁴ <http://groups.google.com/group/abntex2> e <http://abntex2.googlecode.com/>

*“As invenções são, sobretudo,
o resultado de um trabalho de teimoso.”
(Santos Dumont)*

Resumo

Informação sobre o estágio de câncer num paciente é fundamental quando os médicos avaliam o progresso individual desse paciente e decidem sobre novas medidas de tratamento. A determinação do estágio de câncer (cancer staging) é um processo que leva em consideração a descrição, localização, características e possíveis metástases dos tumores cancerosos de um paciente. Esse processo deve seguir um padrão de classificação como, por exemplo, o TNM Classification of Malignant Tumors. Porém, na prática clínica, a execução desse processo pode ser tediosa e propensa a erros, gerando assim imprecisão na interpretação das imagens médicas (com possíveis consequências danosas para os pacientes). Com o intuito de amenizar esses problemas, este trabalho tem como objetivo auxiliar os radiologistas fornecendo uma segunda opinião na avaliação do estágio de câncer de um paciente. Para isso, serão usadas tecnologias da Web Semântica, como ontologias e *reasoning* para classificar automaticamente estágios de câncer. Essa classificação vai usar anotações semânticas feitas por radiologistas em imagens médicas no programa ePAD e armazenadas no formato AIM e regras de uma ontologia representando o padrão TNM. As anotações em AIM serão convertidas para OWL e uma máquina de inferência vai usá-las, juntamente com a ontologia e suas regras, para determinar o estágio do câncer num paciente. Todo o processo será validado por meio de uma prova de conceito com os usuários do Rubin Lab (Radiology Dept. – Stanford University).

Palavras Chave: *OWL*, *SWRL*, estágio de câncer, *ePAD*, *TNM*.

Abstract

Informação sobre o estágio de câncer num paciente é fundamental quando os médicos avaliam o progresso individual desse paciente e decidem sobre novas medidas de tratamento. A determinação do estágio de câncer (cancer staging) é um processo que leva em consideração a descrição, localização, características e possíveis metástases dos tumores cancerosos de um paciente. Esse processo deve seguir um padrão de classificação como, por exemplo, o TNM Classification of Malignant Tumors. Porém, na prática clínica, a execução desse processo pode ser tediosa e propensa a erros, gerando assim imprecisão na interpretação das imagens médicas (com possíveis consequências danosas para os pacientes). Com o intuito de amenizar esses problemas, este trabalho tem como objetivo auxiliar os radiologistas fornecendo uma segunda opinião na avaliação do estágio de câncer de um paciente. Para isso, serão usadas tecnologias da Web Semântica, como ontologias e *reasoning* para classificar automaticamente estágios de câncer. Essa classificação vai usar anotações semânticas feitas por radiologistas em imagens médicas no programa ePAD e armazenadas no formato AIM e regras de uma ontologia representando o padrão TNM. As anotações em AIM serão convertidas para OWL e uma máquina de inferência vai usá-las, juntamente com a ontologia e suas regras, para determinar o estágio do câncer num paciente. Todo o processo será validado por meio de uma prova de conceito com os usuários do Rubin Lab (Radiology Dept. – Stanford University).

Palavras Chave: *OWL, SWRL, estagio de câncer, ePAD, TNM.*

Lista de ilustrações

Figure 1 – CT do peito, (a) anatomia do mediastino: Átrio direito (RA), Ventrículo direito (RV), Válvula aórtica (AV), Aorta (A), Átrio esquerdo (LA). (b) Anatomia do pulmão.	40
Figura 2 – Estudo de paciente com metástase de cólon, pode ser visto que o aumento da captação de glicose no tecido canceroso é facilmente visível.	41
Figura 3 – A imagem CT do fígado anotada com termos semânticos.	42
Figura 4 – Um extrato da ontologia RadLex que é uma terminologia controlada em radiologia.	44
Figura 5 – Arquitetura em camadas da Web Semântica,	50
Figura 6 – Mecanismos de <i>Reasoners</i> , baseados em lógica de descrição	54
Figura 7 – AIM. Anotações de imagem que descreve, localizações anatômicas, medidas e tipos de lesões	58
Figura 8 – Ontologia do Hakan Bulu (BULU; RUBIN, 2015) baseada no modelo AIM 3.0	70
Figura 9 – O diagrama representa as entidades como ovais e retângulos. As propriedades são mostrados em cor vermelho.	72
Figura 10 – Ilustração gráfica do exemplo 5.1 onde as 9 classes base do modelo e seus indivíduos são mostrados. Ovais e retângulos representam entidades e as propriedades são mostrados em cor vermelho.	75
Figura 11 – Neste diagrama descreve-se as classes abstratas como ovais roxas envolvendo sua instância (imageAnnotationCollection, ImageAnnotation , geometricShape). Pode-se ver que o nome de algumas classes mudaram em comparação com o modelo basic e agora eles tem um sufixo "Entity" o que significa que essas classes são uma instância da classe abstrata principal “ Entity ”.	77
Figura 12 – ilustração gráfica do exemplo 5.2 onde as classes de tipo Extended e seus indivíduos são mostrados. Ovais e retângulos representam entidades e as propriedades são mostrados em cor vermelho. A figura mostra entidades ovais cercado por linhas pontilhadas, o que significa que estas entidades são classes abstratas de tipo Entity	80

Figura 13 – A figura mostra as Object Properties adicionadas à ontologia do Hakan Bulu e fazem parte da nova ontologia AIM4-O , podemos ver na figura o ObjectProperty hasAnnotations como propriedade funcional inversa da propriedade hasPerson	81
Figura 14 – Instâncias de anotações(individuos) AIM são mostrados no painel esquerdo, e os detalhes e propriedades de uma instância são mostrados na direita. .	82
Figura 15 – O diagrama da ontologia geral, seus componentes TNM (regras e axiomas) e importações necessárias a fim de ter os mecanismos para classificar imagens de câncer de fígado.	84
Figura 16 – mostra as relações entre as classes e outros conceitos de ONLIRA utilizados neste trabalho.	85
Figura 17 – módulo da ontologia Radlex e as classes que representam grupos de órgãos, tais como adjacentOrganGroup e noadjacentOrganGroup	86
Figura 18 – é apresentada a definição da estrutura de dados gerado na etapa de <i>parsing</i>	89

Lista de tabelas

Tabela 1 – Os critérios de T, N e M para câncer de pulmão.	46
Tabela 2 – Definições para os diferentes estágios de tumores.	47
Tabela 3 – Comparativo entre os trabalhos relacionados.	63
Tabela 4 – Prefijos e Namespaces usados na ontologia AIM4-O	71
Tabela 5 – American Joint Committee on Cancer/International Union against cancer TNM classification system and formal representation	92
Tabela 6 – American Joint Committee on Cancer/International Union against cancer TNM classification system and formal representation	93
Tabela 7 – American Joint Committee on Cancer/International Union against cancer TNM classification system and formal representation	94

Lista de Algoritmos

Lista De Códigos-Fonte

5.1	Representação em sintaxe Manchester do individuo "1.2.288.3.." que representa uma anotação semântica em uma imagem	74
5.2	Representação em sintaxe Manchester de um individuo "Sixmtfqbiqujoz.." da ontologia AIM4-O , que representa uma anotação semântica em uma imagem	78

Lista De Abreviaturas e Siglas

AAS Ann Arbor Staging

ACR American College of Radiology

AIM Annotation and Image Markup Project

AIM Annotation and Image Markup

caBIG cancer Biomedical Informatics Grid

CT Computed Tomography

DICOM Digital Imaging and Communications in Medicine

FMA Foundational Model of Anatomy

FMA Foundational Model of Anatomy

NCI National Cancer Institute

NGS Nottingham Grading System

OWL Ontology Web Language

OWL Web Ontology Language

PET Positron Emission Tomography

RadLex Unified Language of Radiology Terms

RDF Resource Description Framework

SWRL Semantic Web Rule Language

TI Tecnologia da Informação

TNM TNM Classification of Malignant Tumors

URIs Uniform Resources Identifiers

WFML Wide Field Markup Language

WHO World Health Organization

XML eXtensible Markup Language

Sumário

Lista de ilustrações	17
Lista de tabelas	19
Sumário	27
1	INTRODUÇÃO 31
1.1	Contextualização 32
1.2	Motivação 33
1.3	Objetivo 34
1.4	Organização 34
2	INFORMÁTICA BIOMÉDICA 37
2.1	Informática Biomédica para sistemas de suporte à decisão assistido por computador 38
2.2	Tecnologias para Imagens Médicas 39
2.2.1	<i>Tomografia Computadorizada</i> 39
2.2.2	<i>Positron Emission Tomography</i> 40
2.3	Anotação de Imagens 41
2.3.1	<i>Anotação Semântica</i> 42
2.3.2	<i>Anotação de Imagens Médicas com Vocabulários Estruturados</i> 43
2.4	Medical Background Knowledge 43
2.4.1	<i>Estágio do Câncer (Staging System)</i> 44
2.4.2	<i>O Padrão TNM</i> 45
2.5	Considerações Finais 45
3	WEB SEMÂNTICA 49
3.1	Web Semântica 50
3.2	Ontologias 51
3.3	Web Ontology Language – OWL 52
3.4	Semantic Web Rule Language - SWRL 53

3.5	Reasoners	53
3.6	Bio Ontologies	55
3.6.1	<i>Foundational Model of Anatomy – FMA</i>	55
3.6.2	<i>RadLex</i>	56
3.7	Annotation Imaging Markup(AIM)	57
3.8	AIM e ePad	58
3.9	Considerações Finais	59
4	TRABALHOS RELACIONADOS	61
4.1	Planeamento e Execução do Mapeamento Sistemático	61
4.2	Apresentação e Análise dos Resultados Obtidos no Mapeamento	62
4.3	Considerações Finais	68
5	METODOLOGIA	69
5.1	Representação ontológica do modelo de informação AIM	69
5.1.1	<i>Ontologia AIM4-O</i>	70
5.1.1.1	<i>Namespaces</i>	71
5.1.1.2	<i>Basic Terms</i>	71
5.1.1.3	<i>Expanded Terms</i>	76
5.2	Desenvolver um mecanismo para transformar as anotações no formato AIM-XML para instâncias da ontologia AIM4-O.	82
5.3	Gerar as condições para implementar o reasoning baseado em regras TNM que utilizam instâncias OWL.	83
5.3.0.1	<i>Ontologia geral</i>	83
5.3.0.2	<i>Criando classes do conceitos do sistema TNM</i>	84
5.3.0.3	<i>ONLIRA (Ontologia de figado para Radiologia)</i>	85
5.3.0.4	<i>Terminologia Radlex</i>	85
5.4	Reasoning sobre lesões de câncer	86
5.4.1	<i>Representação Formal do staging do câncer</i>	87
5.4.1.1	<i>Representação formal das condições usando OWL</i>	87
5.5	Considerações Finais	91
6	EXPERIMENTOS	95
6.1	Experimentos	95
6.2	a	95
6.3	b	95

REFERÊNCIAS	97
-----------------------	----

Introdução

Em Biomedicina, especificamente em radiologia e oncologia, avaliar a resposta ao tratamento do câncer depende criticamente dos resultados da análise de imagens pelos especialistas. No entanto, as informações obtidas dessa análise não são facilmente interpretadas por máquinas.

Pode-se ver que há uma falta de conexão entre a informação visual da imagem e sua interpretação. É por essa razão que as tecnologias, como a web semântica, geram um interesse crescente para aplicação em biomedicina. Essas tecnologias podem tornar dados em biomedicina explícitos e computáveis. A comunidade biomédica está em busca de ferramentas para ajudar com o acesso, consulta e análise da vasta quantidade de dados gerados pelos avanços no uso da tecnologia na medicina ([MARQUET et al., 2007](#)).

O uso de imagens na medicina, especificamente no tratamento de pacientes com câncer, gera enormes quantidades de informações do tipo não-texto. No entanto, o uso dessas imagens médicas nas tarefas clínicas é importante, pois permite aos especialistas, diagnosticar, planejar e acompanhar os pacientes ([LEVY; O'CONNOR; RUBIN, 2009](#)). Desse modo, um número considerável de aplicações informáticas, voltadas para área médica, têm sido desenvolvidas. Grande parte dessas aplicações estão focadas na extração de características visuais com a ajuda de algoritmos de processamento de imagem.

Embora esses algoritmos possam auxiliar a comunidade biomédica, quanto ao uso de imagens no tratamento de câncer, eles apresentam problemas quando uma consulta abstrata e ambígua é feita no contexto da classificação de pacientes com câncer. Por exemplo, quando um oncologista quer saber se um tumor já se encontra em estado avançado, perto de se espalhar para alguma região próxima da origem do câncer, mas não para outras partes do corpo ([WENNERBERG; SCHULZ; BUITELAAR, 2011](#)). Portanto, podem haver algu-

mas dificuldades em manipular interpretações abstratas de imagens, porque a informação semântica dos laudos sobre a imagem não é considerada nesses algoritmos.

Contextualização

Embora imagens médicas forneçam uma quantidade relevante de informações para os médicos, estas informações não podem ser integradas facilmente em aplicações médicas avançadas tais como, sistemas de apoio à decisão clínica para tratar pacientes com câncer. Especificamente quando os médicos vão avaliar o progresso individual de um paciente para decidir sobre novas medidas de tratamento (ZILLNER; BRETSCHNEIDER, 2013).

No fluxo de trabalho para avaliar o progresso individual de um paciente com câncer, o radiologista identifica as lesões cancerígenas por meio de imagens e grava as medições detalhadas sobre as lesões utilizando anotações. Em seguida, o oncologista analisa e extrai as informações sobre a localização e o tamanho das lesões tumorais a partir das anotações feitas pelo radiologista e registra as informações em uma folha de fluxo.

As informações, contidas nesta folha de fluxo, são então utilizadas para os cálculos da taxa de resposta individual de um paciente. A execução dessa análise final em um único paciente não representa um problema, mas quando o número de pacientes aumenta essa tarefa se torna laboriosa e propensa a erros.

A análise dessa folha de fluxo baseia-se no conhecimento do estágio do câncer (*cancer staging*) do paciente. O estágio do tumor é um processo de classificação baseado em características como localização e tamanho do tumor sobre o corpo. Portanto, obter informações sobre o estágio do câncer é importante para identificar as possíveis opções de tratamento adequadas. Esse processo de classificação poderia ser automatizado, a fim de otimizar o trabalho dos médicos que pode se tornar pesado e propenso a erros, quando o número de pacientes é consideravelmente grande (ZILLNER, 2010).

Há poucas ferramentas que permitem que radiologistas capturem facilmente informações semânticas estruturadas como parte do seu fluxo de trabalho de investigação de rotina (SERIQUE, 2012). O projeto Annotation and Image Markup (AIM) desenvolvido pelo cancer Biomedical Informatics Grid (caBIG), uma iniciativa do National Cancer Institute (NCI) (RUBIN et al., 2008), fornece um esquema XML para descrever a estrutura anatômica e observações visuais em imagens utilizando a ontologia RadLex (KUNDU et al., 2009). Ele permite uma representação, armazenamento e transferência consistentes de significados semânticos sobre imagens. Ferramentas, como por exemplo o ePAD (RUBIN et al., 2014), usam o formato AIM. O ePAD permite aos pesquisadores e clínicos criar anotações semânticas em imagens radiológicas. Ferramentas, como o ePAD, podem ajudar a reduzir o

esforço de se recolher informação semântica estruturada sobre imagens. No entanto é necessário fazer inferências sobre essas informações (lesões de câncer) usando as relações biológicas e fisiológicas entre as anotações.

Motivação

O processo de classificação dos pacientes com câncer, através da análise de imagens, é uma tarefa executada por especialistas, como oncologistas e radiologistas, com base na inspeção de imagens e pode ser, muitas vezes, um trabalho intensivo que exige precisão na interpretação das lesões de câncer. A precisão do especialista é obtida através de formação e experiência (DEPEURSINGE et al., 2014), mas mesmo com boa formação e experiência podem ocorrer variações na interpretação de imagens entre especialistas. Nesse contexto, o desenvolvimento de um sistema de classificação automática representa uma forte necessidade médica que pode ajudar a se obter uma maior taxa de precisão na interpretação.

Por outro lado, embora sistemas como o ePAD permitam a criação de anotações de imagem no formato AIM, facilitando assim a manipulação e consulta de metadados de imagens em AIM, eles não permitem representar anotações em um formato que seja diretamente adequado para o raciocínio (*reasoning*). O AIM fornece apenas um formato para transferência e armazenamento de dados. Um fato importante é que, atualmente, não há muitos métodos de *reasoning* semântico para se fazer inferências sobre lesões cancerígenas utilizando anotações sobre imagens médicas codificadas pelo AIM (LEVY; O'CONNOR; RUBIN, 2009). Outros sistemas, tais como mint Lesion (MintMedical GmbH, Dossenheim, Germany) and syngo.via (Siemens Healthcare, Malvern) são softwares proprietários, ou seja, todos os dados sobre imagens que estão armazenados internamente estão em um formato proprietário que não pode ser acessado por terceiros (RUBIN et al., 2014).

Podemos ver então que existe uma carência de métodos de *reasoning* semântico para fazer inferências sobre lesões cancerígenas a partir de anotações semânticas sobre imagens, baseadas no meio padrão (como AIM) de adição de informação e conhecimento para uma imagem. Sendo assim, a principal motivação deste trabalho é a possibilidade de desenvolver esses métodos de *reasoning* baseados em sistemas de notação, como *TNM Classification of Malignant Tumors* (TNM). A fim de classificar automaticamente pacientes com câncer, a partir de anotações AIM sobre imagens desses pacientes. E finalmente incorporar esses métodos aos requisitos atendidos por sistemas de anotação de imagens, como ePAD.

Objetivo

Este trabalho tem como objetivo gerar automaticamente o *Cancer Staging* (estágios do câncer) de lesões cancerígenas presentes em imagens médicas, utilizando tecnologias de *reasoning* e web semântica para processar metadados sobre anotações dessas imagens gerados por ferramentas que utilizam um esquema XML como AIM, para descrever a estrutura anatômica e observações visuais em imagens. A fim de fornecer aos médicos uma segunda opinião sobre a classificação dos pacientes com um determinado tipo de câncer e obter uma maior taxa de precisão na interpretação das lesões. Vale a pena mencionar que o trabalho atual concentra-se em *staging* do câncer de fígado, devido à disponibilidade de dados.

Para atingir o objetivo proposto, são necessários os seguintes objetivos específicos:

1. A expansão da representação ontológica do modelo de informação AIM 3.0 para a sua versão 4.0 que é usada por ferramentas como ePAD.
2. Representar apropriadamente o critério de classificação(*staging*) do câncer TNM. Esse conhecimento tem que ser codificado em OWL e regras SWRL.
3. Procura e Análise de dados com informações do *staging* TNM, como relatórios de diagnóstico clínico com as suas respectivas imagens CT (foram usados os repositórios *The NCI's Genomic Data Commons* (GDC) e *The Cancer Imaging Archive*).
4. Disponibilização de um protótipo do sistema de classificação automática do câncer no ePAD.

Com os objetivos acima traçados, demonstramos que a utilização de tecnologias da Web Semântica pode obter uma boa taxa de precisão na classificação das lesões e, por tanto pode aumentar a qualidade e uniformidade na interpretação de imagens por especialistas.

Organização

Este trabalho está estruturado da seguinte forma:

Capítulo 2: Apresenta tópicos sobre Informática Biomédica e são abordadas tecnologias de Tecnologia da Informação (TI) usadas nessa área que são pertinentes para infraestrutura necessária para este trabalho (e em outras áreas de pesquisa médica).

Capítulo 3: Apresenta a fundamentação teórica sobre Web Semântica, evidenciando os principais conceitos que serão utilizados no desenvolvimento deste trabalho como ontologias e SWRL, entre outros.

Capítulo 4: Esse capítulo apresenta os trabalhos relacionados, evidenciando o uso de ontologias, e as principais investigações quanto ao estado da arte relacionado ao trabalho.

Capítulo 5: Esse capítulo apresenta a metodologia utilizada para a expansão do AIM, a codificação(SWRL e OWL) do critério TNM, para finalmente permitir o *staging* dos pacientes como base em suas imagens.

Capítulo 6: Este capítulo mostra os resultados obtidos

Capítulo 7: Este capítulo conclui a dissertação de mestrado e exhibe as produções científicas obtidas com o desenvolvimento do trabalho, as limitações encontradas e os trabalhos futuros para aperfeiçoamento.

Informática biomédica

A Informática Biomédica tem sido, ao longo das últimas décadas, um campo emergente graças aos progressos feitos nos campos da computação, microeletrônica e telecomunicações. No entanto, ainda não existe uma definição universal para “Informática” nos campos da medicina. A Informática Biomédica pode ser considerada como a ciência da informação aplicada no estudo do contexto da biomedicina, que surgiu para atender as aplicações da TI (Tecnologia da Informação) na área da saúde ([BULU; RUBIN, 2015](#)). Por sua vez, a Informática Radiológica (sub-área da informática biomédica) é a área da radiologia responsável pela melhoria da eficiência, precisão e confiabilidade dos serviços radiológicos dentro da área médica ([SERIQUE, 2012](#)).

Embora em radiologia muitas coisas, incluindo a criação da imagem, sejam bastante padronizadas, existem muitas maneiras de se descrever o conteúdo de uma imagem, incluindo diferentes sinônimos de termos, vários níveis de descrição e foco em atores diferentes. Nesse contexto, o uso de terminologias visuais e estruturadas para anotações em imagens é uma abordagem promissora, pois permite a definição precisa de imagens médicas utilizando a semântica fornecida pelas anotações ([CHANNIN et al., 2010](#)).

Atualmente, anotações sobre imagens médicas fornecem apenas algumas informações, como, por exemplo, metadados que contém o nome do paciente, idade, médico responsável, etc. As imagens em si não possuem nenhuma informação semântica, por isso, essas anotações não podem ser automaticamente integrados em aplicações médicas avançadas, tais como aquelas para o apoio à decisão clínica. Para facilitar o processamento dessas anotações por máquina, é necessário adicionar informações semânticas que sejam baseadas em fontes de conhecimento, como sistemas de *staging*. Além disso, é preciso que essas informações sejam codificadas de forma compreensíveis para máquinas, utilizando-se ontologias para fa-

cilitar o mapeamento entre conceitos semânticos. Assim, será possível a realização de um processamento mais complexo, como por exemplo, o uso *de reasoning* (HU et al., 2003).

Como o foco deste trabalho é um sistema de classificação automática de pacientes com câncer a partir de imagens usando *reasoning* baseados em sistemas de notação como TNM. Este capítulo abordará os principais conceitos referentes a Sistemas de suporte à decisão assistido por computador, em seguida, na seção 2.2 Informática Radiológica, em seguida, na Seção 2.2, Imagens Medicas. Na seção 2.3, é discutido o processo de Anotações em imagens. Por fim, na seção 2.4, serão descritos os sistemas de *staging* que são a base de conhecimento para a classificação dos pacientes com câncer, a partir de imagens radiológicas.

Informática Biomédica para sistemas de suporte à decisão assistido por computador

Nao é que até os anos 70 que os sistemas de suporte à decisão influenciadas pela informática biomédica foram vistos. Trabalhos como (DOMBAL et al., 1972) com seu sistema chamado INTERNIST-I e (MYERS, 1987) que desenvolveu um sistema baseado em regras para diagnóstico, representam desenvolvimentos iniciais importantes de sistemas de informática biomédica Belle, Kon e Najarian (2013). Mais tarde houve toda uma evolução significativa com grande reconhecimento de seu sucesso na melhoria de desempenho destes sistemas, que é descrito por (PEARSON et al., 2009). Esta evolução permitiu o nascimento de mais subgêneros específicos de métodos de informática biomédica e suas implementações na forma de sistemas de diagnóstico assistido por computador. Nesse contexto o apoio à decisão na prática radiológica clínica faz a sua aparição (STIVAROS et al., 2010).

há uma série de áreas de aplicação na medicina para que os sistemas de apoio à decisão auxiliada por computador tornaram-se implementados. Algumas das principais áreas de aplicação são : Medicina de Emergência e Unidades de Terapia Intensiva, Medicina cardiovascular, Aplicações dentárias, Medicina Pediátrica, Radiologia e Câncer.

No caso de câncer, a informática biomédica começou a desempenhar um papel importante na detecção e tratamento do câncer, muitos estudos aplicam metodos como redes neurales e sistemas híbridos através da combinação de informações heterogêneas. O uso de ferramentas de diagnóstico auxiliado por computador tem o potencial para reduzir a variabilidade entre as opiniões de especialistas, bem como melhorar a precisão diagnóstica para a interpretação do classificação de cancer (JIANG et al., 2001).

No caso de Radiologia, o processamento de imagem baseado em computador é uma área de pesquisa ativa. O processamento de imagem combinado com a visualização e aprendizagem de maquina para a tomada de decisões tem proporcionado uma mais-valia

para aplicações clínicas, e mais recentemente com o surgimento de múltiplas tecnologias de imagens médicas: como a tomografia computadorizada (CT), raios-X, ressonância magnética (MRI), ressonância magnética funcional e (fMRI), numerosos métodos biomédicos foram concebidos como soluções de aplicações clínicas específicas.

A seguir, descreveremos as características de algumas dessas tecnologias de imagens e sua aplicação no estudo de pacientes com câncer.

Tecnologias para Imagens Médicas

A pesquisa e tratamento do câncer são ambos criticamente dependentes de tecnologias de imagens. Avanços em imagem já permitem uma precisão notável em detectar se um tumor invadiu tecidos vitais, cresceu ou se espalhou para órgãos distantes; permitindo que os médicos possam monitorar o progresso do paciente sem a necessidade de biópsias ou outros métodos invasivos ([POLIDAIIS, 2006](#)).

Antes de analisar como é que a semântica pode ser adicionada no processo de monitoramento do progresso dos pacientes com câncer, é importante compreender as tecnologias de imagem primárias com que iremos trabalhar em nossa proposta e o que elas fazem. Essas tecnologias de imagem abrangem Computed Tomography (CT) e Positron Emission Tomography (PET).

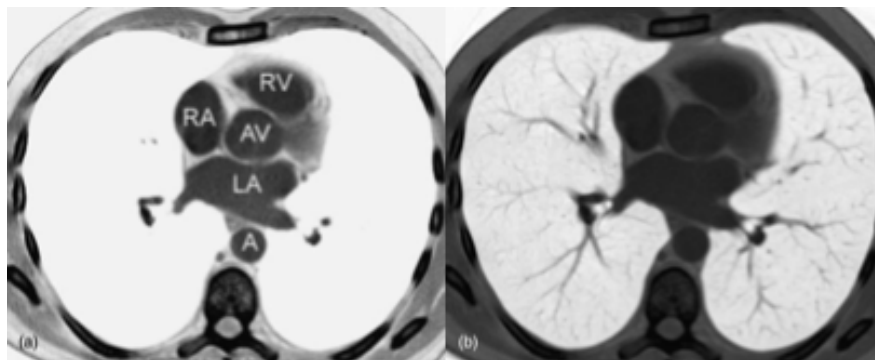
Tomografia Computadorizada

A tomografia computadorizada, ou Computed Tomography (CT), usa técnicas de reconstrução intensivas por computador para criar imagens 3D do corpo a partir de raios-X. Devido ao uso de computadores, uma gama de densidades mais elevada pode ser exibida em comparação com imagens de raios-X convencionais. Esse recurso permite a diferenciação entre órgãos e patologias e detectar a presença de materiais específicos, tais como gordura ou cálcio. Assim como em imagens convencionais de raios-X, objetos de alta densidade causam mais atenuação e, portanto, são exibidos como cinza mais claro que objetos de baixa densidade.

Por exemplo, no CT do peito uma vasta gama de densidades de tecido está presente, uma boa imagem das estruturas do mediastino são mostradas em detalhes do pulmão, mostrados na Figura 1.

Tomografias estão entre as tecnologias de imagem mais comuns utilizados no diagnóstico, bem como no planejamento e monitoramento do tratamento de câncer; especialmente na detecção de câncer de fígado, pâncreas, pulmões e ossos([POLIDAIIS, 2006](#)).

Figura 1 – CT do peito, (a) anatomia do mediastino: Átrio direito (RA), Ventrículo direito (RV), Válvula aórtica (AV), Aorta (A), Átrio esquerdo (LA). (b) Anatomia do pulmão.



Fonte: Adaptada de [Arnold \(2012\)](#).

Positron Emission Tomography

Os exames de Positron Emission Tomography (PET) fornecem uma ferramenta fundamental na gestão e terapia de câncer para muitos tumores comuns. A maioria dos estudos clínicos de PET são realizados para determinar o grau de um tumor, a localização e número de metástases.

Nesses estudos oncológicos PET, o paciente recebe uma injeção do açúcar radioactivo que pode ajudar na localização de um tumor, porque as células cancerosas utilizam o açúcar (glicose) mais avidamente do que outros tecidos no corpo([ZIEGLER, 2005](#)).

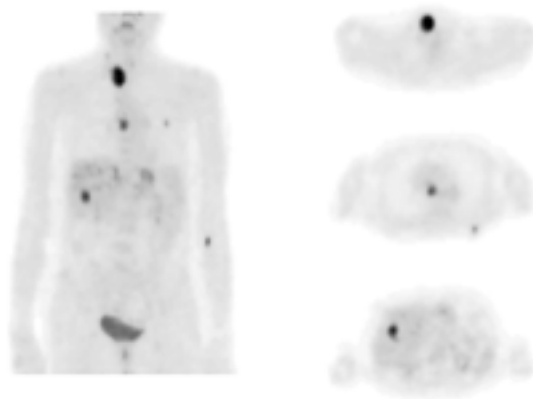
Por exemplo, a Figura 2 mostra um exemplo de um paciente com metástases de câncer de cólon, onde a glicose no tecido canceroso é facilmente visível. Por esse motivo, para muitos tumores comuns, a PET é a técnica mais precisa para visualizar a propagação do tumor ou a sua resposta à terapia([POLIDAI, 2006](#)).

É bom notar que, o sistema ePAD no qual vamos desenvolver nossa proposta, permite a manipulação de imagens de TC e PET([RUBIN et al., 2014](#)).

Todas as modalidades de análise acima mencionadas fornecem uma visão detalhada sobre a anatomia humana, suas funções, associações a doenças. As técnicas avançadas de análise dessas imagens geram parâmetros quantitativos adicionais, abrindo assim o caminho para a melhoria da prática clínica e diagnóstico([ZILLNER; BRETSCHNEIDER, 2013](#)).

Embora imagens não processadas (raw images) possam ser úteis em algumas aplicações de processamento por computador, grande parte do conteúdo semântico nas imagens e dos relatórios radiológicos associados a elas, não é explícito ou disponível para computadores. Isso dificulta a descoberta de novas informações num grande volume de informação. Com

Figura 2 – Estudo de paciente com metástase de cólon, pode ser visto que o aumento da captação de glicose no tecido canceroso é facilmente visível.



Fonte: Adaptada de [Ziegler \(2005\)](#).

base nisso, as aplicações médicas avançadas devem possuir descrições semânticas sobre dados clínicos, dados como as imagens médicas.

Anotação de Imagens

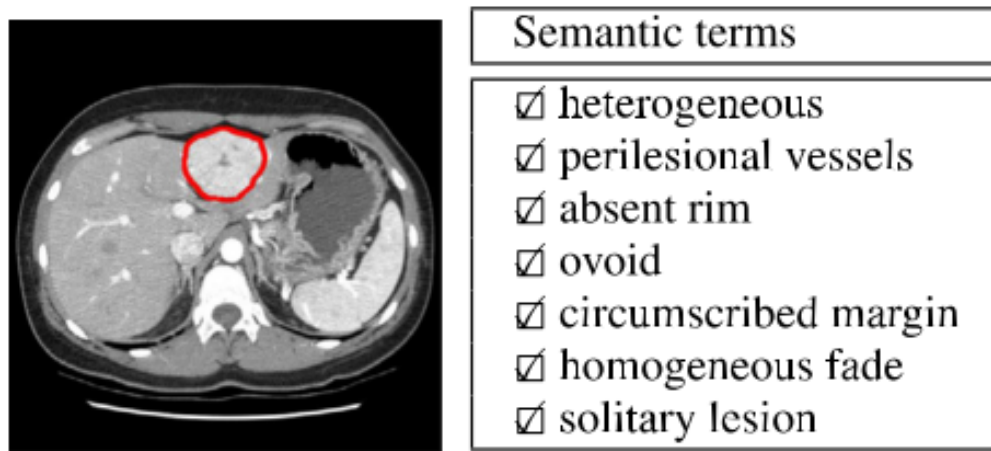
A necessidade de anotar imagens é reconhecida em uma ampla variedade de aplicações diferentes, abrangendo tanto o uso profissional e pessoal de dados sobre imagens. Anotação de imagens é uma tarefa complexa, que tem sido amplamente estudada nos domínios da visão computacional e recuperação de imagens.

No contexto da radiologia, anotar imagens é uma tarefa que pode ser assistido por computador. No entanto, os radiologistas geralmente não gravam suas anotações num formato estruturado, que seja acessível por máquina, impedindo assim o bom desempenho dos sistemas de decisão sobre diagnósticos. O desempenho deste tipo de sistema baseia-se na escolha dos termos que estão sendo usados para descrever o conteúdo das imagens. Essa escolha é altamente dependente da aplicação, as necessidades e a experiência do usuário.

Tais termos podem ser diretamente derivados da terminologia fornecida pelos radiologistas em seus relatórios ou automaticamente previsto a partir das características das imagens ([GIMENEZ et al., 2011](#)). Eles podem ser usados para descrever uma variedade de informações sobre o conteúdo de uma imagem (por exemplo, a forma da lesão de um tumor canceroso). Esses termos podem representar também o conteúdo semântico da imagem, nesse caso eles podem ser chamados de anotações semânticas.

As anotações semânticas estão diretamente ligados ao alto nível de compreensão e

Figura 3 – A imagem CT do fígado anotada com termos semânticos.



Fonte: Adaptada de [Kurtz et al. \(2014\)](#).

descrições do usuário sobre as características da imagem ([RUBIN, 2011](#)). Com base nessas considerações, incorporar características semânticas em sistemas de decisão sobre diagnóstico pode ser uma tentativa promissora para preencher a lacuna semântica entre a descrição visual de uma imagem e seu significado ([MA et al., 2010](#)), Figura 3.

Anotação Semântica

De acordo com [Nagarajan \(2006\)](#), anotação é o processo de associar metadados a recursos como áudio, vídeo, texto estruturado, páginas web, imagens, etc. Enquanto que anotação semântica é o processo de marcar com recursos de metadados semânticos. No contexto da biomedicina, esse processo é conhecido como: anotação semântica de imagens médicas.

Em radiologia, o processo de marcar imagens é um trabalho manual, lento e inconsistente devido ao grande número de imagens que são geradas por dia em hospitais modernos ([DEPEURSINGE et al., 2014](#)). É por essa razão que várias abordagens para enfrentar o desafio da anotação semântica de imagens médicas já foram estudados. Por exemplo, em [Möller, Regel e Sintek \(2009\)](#) uma abordagem para a extração de informações de *DICOM headers* e de relatórios estruturados DICOM (Digital Imaging and Communications in Medicine) foi apresentada, [Seifert et al. \(2009\)](#) introduziram um novo método para a análise automática de imagens (em anatomia e com detetores de tecidos específicos). Em outra frente, [Rubin et al. \(2008\)](#) integraram a anotação manual de imagens no fluxo de trabalho dos radiologistas e [Hu et al. \(2003\)](#) introduziram uma abordagem de anotação de imagens para melhorar o diagnóstico do câncer da mama. Todas essas abordagens fizeram uma contribuição importante

para melhorar o acesso às informações em imagens médicas, especificando a "semântica" das regiões nas imagens.

Anotação de Imagens Médicas com Vocabulários Estruturados

Embora o processo de descrição de imagens médicas utilizando conteúdo semântico possa melhorar o acesso à informação médica, ele pode levar a se ter diferentes descrições para uma mesma imagem. Isso prejudica o bom desempenho de sistemas baseados nessas descrições, como por exemplo os sistemas de apoio à decisão de diagnóstico. Para lidar com esse problema, os últimos trabalhos no domínio semântico usam vocabulários controlados para anotar as imagens ([KORENBLUM et al., 2011](#)).

Um vocabulário controlado fornece um conjunto de termos pré-definidos para descrever características em imagens médicas. Esses vocabulários podem facilitar a anotação de grandes conjuntos de imagens. Trabalhos, como ([NAPEL et al., 2010](#)), investigaram métodos assistidos por computador no apoio ao diagnóstico. Os autores utilizaram um banco de dados de imagens onde as imagens foram anotadas semanticamente. O resultado do estudo indicam que as anotações semânticas, usando um vocabulário controlado, podem levar a diagnósticos mais precisos.

Medical Background Knowledge

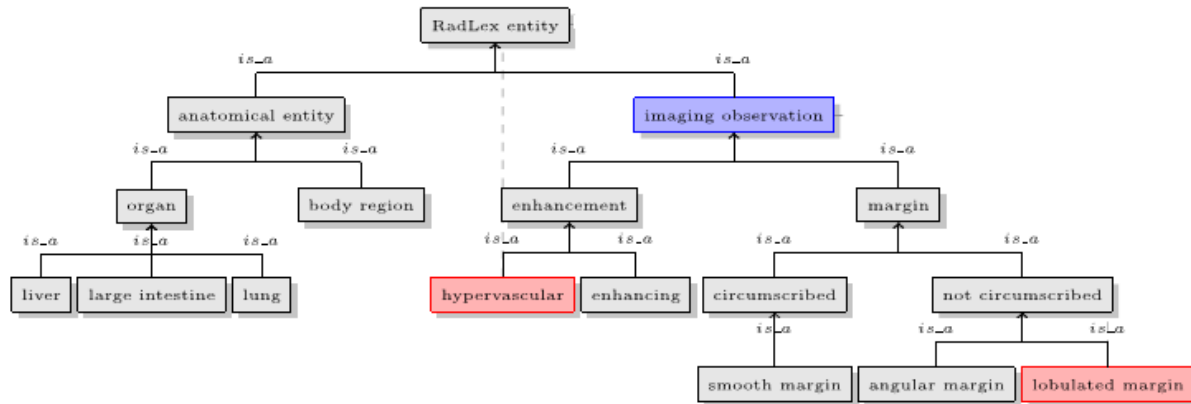
Embora os vocabulários são os blocos de construção básicos para a inferência sobre as técnicas de Web Semântica. Não há uma divisão clara entre o que é referido como "vocabulários controlados" e "ontologias". Considerando que o vocabulário é usado quando o formalismo rigoroso não é necessariamente usado ou apenas em um sentido muito solto, podemos dizer que a diferença fundamental entre uma ontologia e um vocabulário controlado é o nível de abstração e relações entre conceitos.

As ontologias são vocabulários controlados expresso em uma linguagem de representação de ontologias (como OWL) ([BROWNE; JERMEY, 2004](#)). Elas fornecem uma maneira formal para modelar conhecimento (*Background Knowledge*) e geralmente são construídas a partir de um consenso entre especialistas de um domínio. As ontologias representam uma poderosa forma de estruturar termos semânticos pertencentes a uma fonte de conhecimento particular.

No contexto da medicina, várias ontologias estão sendo desenvolvidas para organizar conceitos biomédicos de uma forma abrangente, por exemplo: *Medical Subject Headings (MeSH)*, *International Classification of Diseases (ICD Taxonomy)*, *Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT)*, *Unified Language of Radiology Terms*

(RadLex) (KURTZ et al., 2014). Na Figura 4 a seguir, vemos uma parte da ontologia RadLex.

Figura 4 – Um extrato da ontologia RadLex que é uma terminologia controlada em radiologia.



Fonte: Adaptada de Kurtz et al. (2014).

A fim de atingir nosso objetivo de classificar pacientes com câncer é necessário representar um conhecimento geral do domínio (*background knowledge*). No caso deste trabalho, esse conhecimento é chamado *sistema de staging*. No entanto, ter representações formais das definições de um sistema de *staging* numa ontologia não é suficiente para fornecer ajuda automática na classificação de tumores de câncer. Esse conhecimento deve ser representado também em regras e axiomas (reasoning). Portanto, representar o conhecimento dos sistemas de *staging* de cancer como o TNM em ontologias e regras parece uma contribuição desejável para a realização da classificação automática de tumores em imagens médicas, e por conseguinte uma contribuição aos sistemas de apoio à decisão auxiliada por computador. Os Conceitos como ontologias, regras e *reasoning* serão melhor detalhados no próximo capítulo.

Estágio do Câncer (*Staging System*)

O estágio do câncer categoriza a progressão de um câncer no corpo, em termos da extensão do tumor primário e de se espalhar para alguma região próxima da origem do câncer. A rotina para determinar o estágio de um câncer (*cancer staging*) em pacientes tem uma série de vantagens reconhecidas por organizações relacionadas a câncer em todo o mundo. O *cancer staging* permite que os médicos possam determinar o tratamento mais apropriadamente, avaliar os resultados de forma mais confiável e comparar as estatísticas a nível local, regional e nacional com mais confiança (MCCOWAN; MOORE; FRY, 2006).

Esses benefícios têm motivado a criação de padrões internacionais para *cancer staging*, como, por exemplo, o padrão *TNM* definido pela *AJCC* (*American Joint Committee on Cancer*) e *UICC* (*International Union Against Cancer*).

O Padrão TNM

O *cancer staging* no padrão TNM Classification of Malignant Tumors (TNM) é um processo de duas etapas. A primeira etapa consiste em dar notas para três conceitos: descrição do tumor (T), difusão em nós linfáticos (N) e possível metástase (M). Esses conceitos estão resumidos na Tabela 1.

A segunda etapa consiste na determinação do estágio de acordo com as pontuações da fase anterior. As pontuações sobre T, N e M definem um estágio único do tumor de 0 a IV. Note-se que várias combinações de pontuações T, N e M podem levar ao mesmo estágio (DAMERON et al., 2006). Esses conceitos estão resumidos na Tabela 2.

Considerações Finais

Recentemente pode ser visto que houve uma escalada da complexidade na recoleção dos dados médicos pelo fato dos avanços em tecnologias relacionadas com aquisição de imagem. As novas tecnologias como dispositivos médicos e sistemas de medição, geram grandes volumes de imagens como também outros dados por paciente, e torna difícil para os médicos analisar e fornecer diagnóstico ou prognósticos na hora.

Este capítulo forneceu uma base teórica para compreensão de conceitos como vocabulários controlados, anotações de imagem, anotação semântica e representação do conhecimento. Foi mostrado também como esses conceitos podem ser utilizados na área de informática biomédica para melhorar processos como a interpretação de imagens médicas em câncer.

Tabela 1 – Os critérios de T, N e M para câncer de pulmão.

T: Primary Tumor	Tis	The cancer is found only in the layer of cells lining the air passages
	T1	The cancer is no larger than 3 cm, has not spread to the membranes that surround the lungs (visceral pleura), and does not affect the main branches of the bronchi.
	T2	The cancer has 1 or more of: (1) It is larger than 3 cm. (2) It involves a main bronchus but is not closer than 2 cm to the point where the windpipe branches into the left and right main bronchi. (3) It has spread to the visceral pleura. (4) The cancer may partially clog the airways, but this has not caused the entire lung to collapse or develop pneumonia.
	T3	The cancer has 1 or more of: (1) It has spread to the chest wall, the diaphragm, the membranes surrounding the space between the 2 lungs (mediastinal pleura), or membranes of the sac surrounding the heart (parietal pericardium). (2) It involves a main bronchus and is closer than 2 cm to the point where the windpipe branches into the left and right main bronchi, but does not involve this area. (3) It has grown into the airways enough to cause 1 lung to entirely collapse or to cause pneumonia of the entire lung
	T4	The cancer has 1 or more of: (1) It has spread to the space behind the chest bone and in front of the heart, the heart, the windpipe, the esophagus, the backbone, or the point where the windpipe branches into the left and right main bronchi (carina). (2) Two or more separate tumor nodules are present in the same lobe. (3) There is a fluid containing cancer cells in the space surrounding the lung
N: Regional Lymph Nodes	N0	The cancer has not spread to lymph nodes.
	N1	The cancer has spread to lymph nodes within the lung, hilar lymph nodes (located around the area where the bronchus enters the lung). The cancer has metastasized only to lymph nodes on the same side as the cancerous lung.
	N2	The cancer has spread to lymph nodes around the point where the windpipe branches into the left and right bronchi or to lymph nodes in the mediastinum (space behind the chest bone and in front of the heart). The lymph nodes on the same side of the cancerous lung are affected.
	N3	The cancer has spread to lymph nodes near the collarbone on either side, to hilar or mediastinal lymph nodes on the side opposite the cancerous lung.
M: Distant Metastasis	M0	The cancer has not spread to distant sites.
	M1	The cancer has spread to distant sites such as other lobes of the lungs, lymph nodes farther than those mentioned in N stages, and other organs or tissues such as the liver, bones, or brain

Fonte: Adaptada de [Dameron et al. \(2006\)](#).

Tabela 2 – Definições para os diferentes estágios de tumores.

Stage 0	(Tis, N0, M0)
Stage IA	(T1, N0, M0)
Stage IB	(T2, N0, M0)
Stage IIA	(T1, N1, M0)
Stage IIB	(T2, N1, M0) or (T3, N0, M0)
Stage IIIA	(T1, N2, M0) or (T2, N2, M0) or (T3, N1, M0) or (T3, N2, M0)
Stage IIIB	(T1, N3, M0) or (T2, N3, M0) or (T3, N3, M0) or (T4, N0, M0) or (T4, N1, M0) or (T4, N2, M0) or (T4, N3, M0)
Stage IV	(Any T, Any N, M1)

Fonte: Adaptada de [Dameron et al. \(2006\)](#).

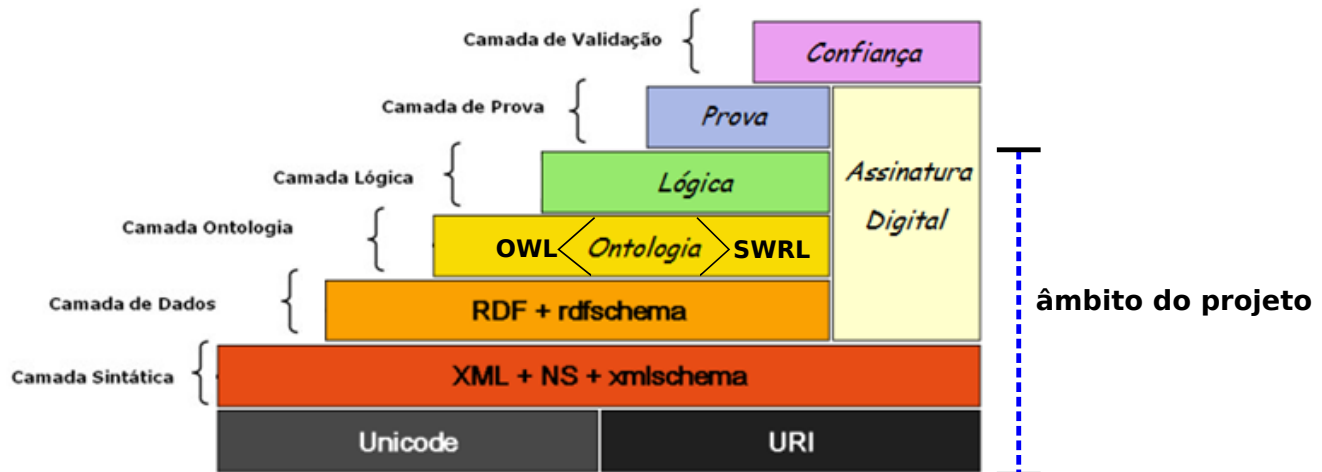
Web Semântica

O objetivo fundamental da Web Semântica é fazer com que as máquinas entendam o significado das informações disponíveis na Web; daí a palavra "semântica". No caso de seres humanos, entender uma sinal ou uma palavra não é uma tarefa extraordinária, nosso cérebro associa os conceitos que acumulou ao longo dos anos, portanto, nossa interpretação semântica é fornecida pelas estruturas neurais. Mas para as máquinas atuais, "entendimento" não deve ser relacionado à "compreensão humana", mas sim, à "inferência e dedução".

Tendo em conta que, os dados de entrada para realizar o processo de *staging* de câncer neste trabalho são basicamente terminologias que descrevem imagens usando o formato XML, é vital abordar a semântica destes termos com as tecnologias de Web Semantica. Tecnologias como as Ontologias, OWL-SWRL e reasoners são essenciais para realizar tarefas complexas de inferencia como o *staging*.

Conceitos relacionados a Web Semântica e ontologias foram mencionados no capítulo anterior, neste capítulo esses conceitos vão ser apresentados e discutidos. O capítulo está organizado da seguinte maneira, na Seção 3.1 a Web Semântica é discutida. Na Seção 3.2 serão expostos conceitos referentes à Ontologias. A Web Ontology Language (OWL) é exposta na seção 3.3, assim como a Semantic Web Rule Language (SWRL) na seção 3.4. Em seguida, na seção 3.6 apresentaremos as principais bio-ontologias e terminologias envolvidas nas pesquisas deste trabalho. Finalmente na seção 3.7 serão exposto a infraestruturas de dados AIM, e sua interação com ferramentas de anotação como ePAD.

Figura 5 – Arquitetura em camadas da Web Semântica,



Fonte: Adaptada de [Berners-Lee, Hendler e Lassila \(2001\)](#).

Web Semântica

[Berners-Lee, Hendler e Lassila \(2001\)](#) propuseram em 2001 a Web semântica como uma extensão da Web atual que fornece informações com significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação. A partir dessa visão conceitual sobre a Web semântica, Berners-Lee propôs a arquitetura da Web Semântica conhecida como “bolo de noiva”, que está ilustrada na Figura 5. A idéia central dessa arquitetura é de que cada camada vai gradativamente trazendo uma nova contribuição como maior expressividade, possibilidade de se realizar inferências e autenticação. A arquitetura da Web Semântica está categorizada nas seguintes camadas: camada URI/Unicode, camada sintática, camada de dados, camada de ontologia, camada lógica, camada de prova, camada de validação e camada de assinatura digital.

A base da arquitetura de camadas da Web Semântica é formada pelos padrões Unicode e URIs. Eles facilitam o intercâmbio de dados: o Unicode permite aos computadores representar e manipular caracteres em quase todas as línguas existentes e URIs (*Uniform Resources Identifiers*) permitem identificar unicamente recursos disponíveis na Web através de uma string.

A camada sintática é formada pela linguagem XML (*eXtensible Markup Language*) e permite a criação de marcações para descrição de informações. As syntaxes das camadas superiores são baseadas em XML. A camada de dados representa conceitos e regras lógicas. A linguagem RDF (*Resource Description Framework*) provê um modelo de descrição lógica de dados permitindo descrever informações sobre um determinado recurso e a RDF Schema

permite a criação de um vocabulário para a camada RDF;

A camada da ontologia estende a camada de dados, provendo um maior nível de expressividade para a definição da semântica das informações. A camada lógica permite definir regras lógicas para inferir novos conhecimentos. A camada de prova e confiança, situadas na parte alta da pirâmide, provêm um mecanismo para avaliar o nível de confiabilidade das fontes de recursos e informações. A camada de assinatura digital, inserida na estrutura da pirâmide, permite incorporar mecanismos de segurança que garantam a confiabilidade da informação.

A expressividade neste trabalho é o fator muito importante, ela permite a modelagem de *reasoning* para responder a questões complexas automaticamente, por essa razão o trabalho desenvolvido engloba até a camada lógica da arquitetura da Web Semântica (ver na Figura 5). Nas seções seguintes iremos detalhar as tecnologias da Web Semântica que serão utilizadas para este trabalho, como: Ontologias, linguagens de construção: OWL, XML, linguagens de regras: SWRL, Reasoners, Terminologias.

Ontologias

Segundo [Smith et al. \(2007\)](#), a Ontologia, um ramo da filosofia, é a ciência “do que é”, dos tipos e estruturas dos objetos, propriedades, eventos processos e relações em todas as áreas da realidade. Na área das ciências de computação e informação, o termo ontologia vem se tornando muito popular, principalmente nas áreas de engenharia do conhecimento, processamento de linguagem natural, sistemas de informação cooperativos, integração inteligente da informação e gerência do conhecimento ([SMITH; MANN, 2003](#)).

[Gruber \(1993\)](#) define o termo "ontologia" como uma especificação formal e explícita de uma conceitualização compartilhada de um domínio de interesse e de informações de como os indivíduos são agrupados e se encaixam em um determinado domínio.

O principal papel das ontologias na Web Semântica é explicitar o vocabulário utilizado e servir como padrão para compartilhamento de informação entre agentes, softwares e aplicações ([CARDOSO et al., 2014](#)).

A fim de representar ontologias para que elas possam ser consumidas por computadores, a representação formal é usada. Segundo [Patel-schneider \(2005\)](#) quanto a representação formal, existem linguagens para representação de ontologias. Normalmente, usa-se lógica de predicados, lógica descritiva. Atualmente, as linguagens mais populares para descrever ontologias são RDF /RDF-S e OWL.

Web Ontology Language – OWL

Embora a proposta inicial da Web Semântica fosse utilizar o RDF, essa linguagem demonstrou limitações em poder de expressividade ([HORROCKS; PATEL-SCHNEIDER; HARMELEN, 2003](#)). O W3C lançou em 2004 a Web Ontology Language (OWL), um resultado da formação do *Web Ontology Working Group*. Ontologias em OWL foram desenvolvidas em várias áreas, como e-Science, medicina, biologia, geografia, astronomia, defesa e nas indústrias automobilística e aeronáutica ([GRAU et al., 2008](#)). Apesar do sucesso da linguagem OWL foram encontradas algumas deficiências. Em resposta aos comentários e requisições dos usuários da OWL, atualmente, a linguagem OWL encontra-se em sua segunda versão (OWL 2) que é subdividida em três sub-linguagens, OWL EL, OWL QL e OWL RL, cada uma com um poder de expressividade diferente, porém todas permitem a criação de ontologias ([BECHHOFFER et al., 2004](#)).

A OWL EL é baseada na família EL++ de lógica descritiva, sendo que sua utilização é particularmente útil em aplicações que contém um grande número de propriedades e classes para definir uma ontologia. Além disso, a OWL EL utiliza um padrão comum em ontologias para conceitos e planejamento, ou seja, a combinação de conjunções e qualidades existenciais ([BECHHOFFER et al., 2004](#)).

A OWL QL é estruturada a partir da família DL-Lite de lógica de descrição (Description Logic). Ela foi criada para permitir o raciocínio (reasoning) eficiente em grandes quantidades de dados estruturados de acordo com esquemas relativamente simples. Ela fornece vários recursos para capturar modelos conceituais, tais como diagramas de classe UML, diagramas de Entidade de Relacionamento, e esquemas de banco de dados ([BECHHOFFER et al., 2004](#)).

Por fim, a OWL RL foi criada para dar suporte a aplicações que exigem raciocínio escalável em troca de alguma restrição de poder expressivo. Através de um subconjunto sintático, é possível implementar o raciocínio (*reasoning*) usando tecnologias baseadas em regras que geralmente são mais escaláveis e fáceis de implementar ([BECHHOFFER et al., 2004](#)) como o caso do SWRL.

Infelizmente a expressividade de OWL nem sempre é suficiente para modelar os tipos de problemas na área biomédica e na Web Semântica ([ORLANDO, 2012](#)). A necessidade de ter suporte para sentenças de implicação (*horn rules* que OWL não suporta) usadas em regras para sistemas de decisão na área médica, levou para criar SWRL.

Semantic Web Rule Language - SWRL

Dado que a interoperabilidade é uma das metas principais da Web Semântica e que as regras são uma parte fundamental dessas metas, um passo fundamental para fazer essa interoperação na Web é a Semantic Web Rule Language (SWRL)¹, que foi projetada para ser o idioma de regras da Web Semântica (CONNOR et al., 2005).

A SWRL é baseada numa combinação das sub-linguagens OWL-DL e OWL-Lite da OWL e as sub-linguagens Unary/Binary Datalog da *Rule Markup Language*².

A SWRL permite aos usuários escrever regras na forma de expressões Horn e usar termos de conceitos OWL para raciocinar (to reason) sobre indivíduos OWL. As regras podem ser usados para inferir novos conhecimentos a partir de bases de conhecimento em OWL.

A especificação da SWRL³ não impõe restrições sobre como o raciocínio deve ser realizado com regras SWRL. Assim, pesquisadores são livres para usar uma variedade de mecanismos de regras para raciocinar com regras SWRL armazenadas em uma base de conhecimento OWL. Dessa forma, a SWRL fornece um ponto de partida conveniente para a integração de sistemas de regras para trabalhar com a Web Semântica (CONNOR et al., 2005).

As regras são escritas em termos de classes, propriedades, indivíduos, e valores de dados OWL, por exemplo:

$$pessoa(?x) \wedge temLesao(?x, ?y) \wedge localizadoEm(?y, ?z) \wedge figado(?z) \rightarrow TemCancer(?x)$$

O efeito da regra acima, seria classificar todos os indivíduos da classe *pessoa* que possuíssem uma lesão hepática como também pertencentes à classe de pessoas com câncer. SWRL também permite o uso de *built-ins* que torna ele como uma linguagem muito rica em recursos de representação.

Para a execução de regras tais como a regra acima, as linguagens da Web Semântica usam *reasoners* automatizados. Os Reasoners são ferramentas muito importantes para realizar inferência de dados.

Reasoners

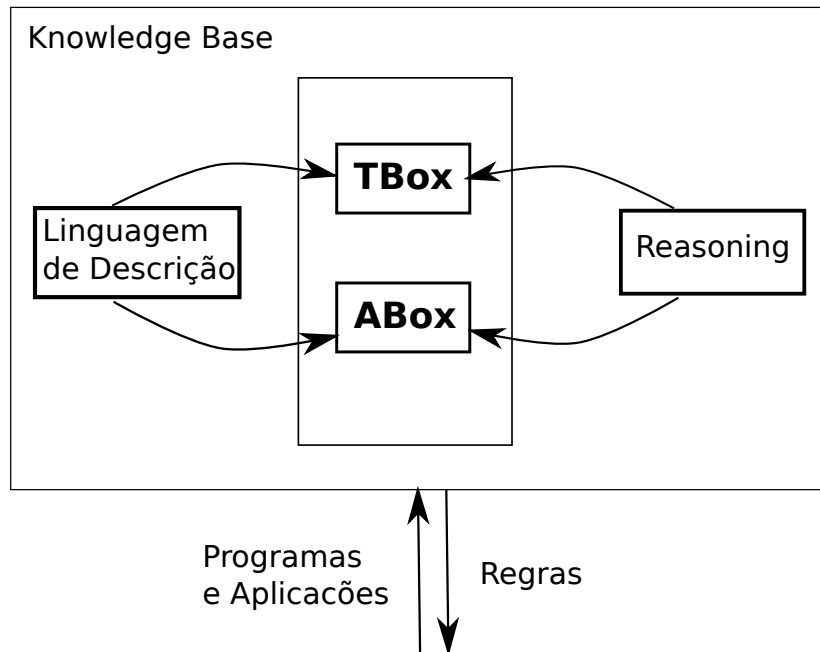
Segundo (AMANQUI, 2014) a finalidade dos *reasoners* na Web Semântica é realizar inferência sobre os dados, com o objetivo de obter novas informações. Para realizar o processo

¹ <http://www.daml.org/rules/proposal/>

² <http://www.ruleml.org/>

³ <http://www.w3.org/Submission/SWRL/>

Figura 6 – Mecanismos de *Reasoners*, baseados em lógica de descrição



Fonte: Adaptada de [Baader \(2003\)](#).

de inferência, os *reasoners* utilizam um motor de inferência associado com um conjunto de regras descritas em linguagens como OWL,RDF(*S*).

Os *reasoners* contêm dois mecanismos internos para processar bases de conhecimento (*knowledge base*), esses mecanismo são TBOX e ABOX a Figura 6 apresenta esses mecanismos associados a uma base de conhecimento(*knowledge base*). Ambos são descritos conforme [AMANQUI \(2014\)](#):

TBox: armazena um conjunto de afirmações universalmente quantificadas para descrever a estrutura de um domínio. Além disso, contém frases que descrevem as relações entre conceitos. por exemplo, uma afirmação deste tipo é a que refere que um conceito representa uma especialização de um outro conceito:

$$\text{Homem} \sqsubseteq \text{Pessoa}$$

Nesse exemplo, a classe (conceito) Homem é declarado como uma Pessoa com especialização.

ABox: Compreende afirmações em objetos individuais (*instant assertions*). A afirmação típica na Abox é a que refere que um indivíduo é uma instância de um certo conceito,

como no exemplo a seguir onde é afirmado que o joaquin é um indivíduo da classe Homem:

$$\text{Homem} \equiv \text{joaquin}$$

Vale destacar que parte da base de conhecimento utilizado pelo *reasoner* neste projeto, é representado por ontologias em OWL destinadas às áreas de medicina e biologia, essas ontologias são chamados Bio Ontologies.

Bio Ontologies

Grandes volumes de dados digitais sobre conhecimento médico e biológico sobrecarregam os pesquisadores dessas áreas. Eles têm a difícil tarefa de gerenciar terabytes de informação, com conteúdos semânticos, em grandes repositórios de dados que usam tecnologias de gerenciamento cada vez mais sofisticadas. Pesquisadores da área biomédica podem trabalhar com uma enorme quantidade de dados, que não pode ser interpretada sem o auxílio de computadores ([SERIQUE, 2012](#)).

Os pesquisadores também devem agregar e integrar toda estas informação. Para tanto, eles precisam de ferramentas que permitam a descoberta de conhecimento nesse paradigma rico em dados. Portanto, eles começaram a utilizar ontologias para descrever a estrutura de seus domínios complexos e relacionar seus dados para compartilhar e interoperar informações Biomédicas ([RUBIN et al., 2014](#)).

Muitas áreas, a exemplo da Informática Biomédica, estão utilizando ontologias para promover o entendimento, interoperabilidade e controle de entropia dos dados. Isso é feito através do uso e desenvolvimento de novas ferramentas semânticas, capazes de recuperar e organizar dados distribuídos apoiadas, em sua grande parte, por Aplicações Web ([SERIQUE, 2012](#)). Nos campos de medicina como: radiologia e oncologia, as duas ontologias mas conhecidas, usadas são o FMA e o *RadLex*.

Foundational Model of Anatomy – FMA

Uma bio-ontologia importante no escopo deste estudo é a Foundational Model of Anatomy (FMA). Segundo [Serique \(2012\)](#) a anatomia é o estudo das estruturas físicas dos organismos biológicos. Por tanto, descrever e capturar de forma computável as terminologias e ontologias que envolvem entidades anatômicas é uma tarefa não trivial. A falta de uma representação generalizada da anatomia levou os desenvolvedores de terminologias e ontologias, a representar a anatomia a partir dos seus próprios pontos de vista. Tanto na área de medicina

clínica quanto de biologia isso gera um obstáculo para correlacionar os termos anatômicos entre si, devido a heterogeneidade das representações anatômicas(SERIQUE, 2012).

A estrutura ontológica da FMA fornece um modelo de inferência, permitindo que ferramentas computacionais possam inferir informações sobre dados anatômicos. Especificamente, a FMA representa um corpo coerente de conhecimentos declarativos sobre a anatomia humana (ROSSE; MEJINO JOSÉL.V., 2008). No entanto, seu quadro ontológico pode ser aplicado e estendido para outras espécies. A FMA tem sido desenvolvida e mantida pelo *Structural Informatics Group* da *University of Washington*.

RadLex

O RadLex é a mais importante bio-ontologia do escopo deste trabalho. Ela surgiu da necessidade de disseminar o conhecimento radiológico entre radiologistas. De forma geral, os radiologistas utilizam constantemente imagens médicas, relatórios de imagens e registros médicos que se encontram online. Para tanto, eles necessitam de uma linguagem unificada para organizar e recuperar esses dados.

Por muitas décadas, o *American College of Radiology* (ACR) desenvolveu o *Index for Radiological Diagnoses*, conhecido como o Índice ACR. Ele tem características atraentes para indexação de imagens. Para tanto, utiliza um sistema de dois códigos numéricos, separados por pontos. O primeiro é para a localização da anatomia e o segundo para a entidade patológica: por exemplo, o código ACR 642.3212 significa 642 = “adenocarcinoma primário” e 3212 = “língua” (LANGLOTZ, 2006).

Como o Índice ACR possui apenas milhares de índices, fica fácil para as pessoas se lembrarem. Todavia, comparado às novas terminologias e ontologias, como o SNOMED-CT e FMA, fica evidente que ele oferece poucos termos específicos.

Com o surgimento do mundo digital, a necessidade de criar sistemas de indexação informatizados levou a RSNA a desenvolver a ontologia RadLex (RNSA, 2012), como resposta as lacunas encontradas em outras terminologias radiológicas, criando assim uma única e completa terminologia. O RadLex é uma forma unificada de reunir os termos usados em radiologia; ela é projetada para atender às necessidades dos radiologistas (LANGLOTZ, 2006).

Ferramentas como ePAD tentam atender às necessidades dos radiologistas, essa ferramenta utiliza o Radlex como a terminologia para descrever imagens em forma de anotações e a sua vez a infraestrutura de tipo AIM.

O RadLex está disponível para download no *National Center for Biomedical Computing's Bioportal site* ⁴.

⁴ <http://bioportal.bioontology.org/ontologies/40885>

Annotation Imaging Markup(AIM)

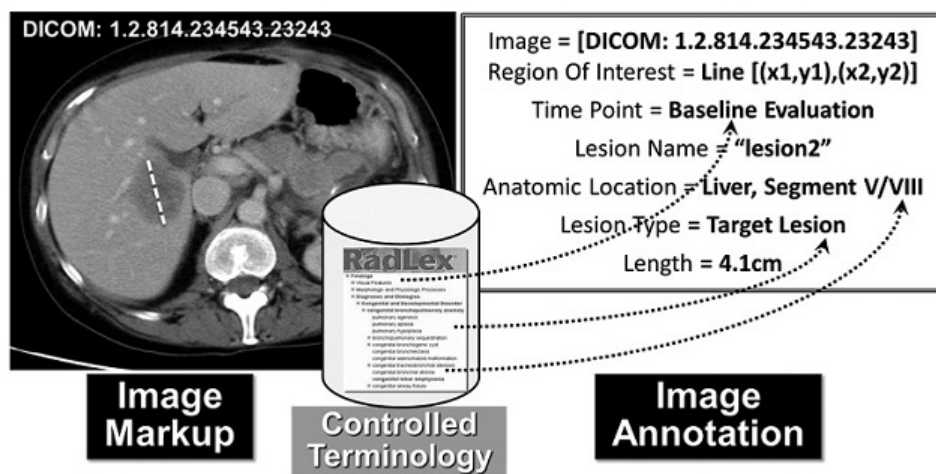
As modernas imagens médicas contêm uma vasta quantidade de informação no formato padrão DICOM. No entanto, essas informações são relacionadas apenas a metadados sobre as imagens. As informações mais importantes relacionadas as imagens ficam codificadas nos seus pixels. Elas contêm um rico conteúdo que, porém, não está explícito ou acessível às máquinas. Essas informações são interpretadas por observadores humanos, como os radiologistas, e não são capturadas ou relacionadas às imagens de uma forma estruturada ([RUBIN et al., 2008](#)).

O projeto Annotation Imaging Markup (AIM), que atualmente encontra-se em sua quarta versão, trata de um modelo de informação que oferece uma infra-estrutura de apoio para a criação de uma coleção de anotações médicas de imagens (que podem ser lidas por máquinas). Ele é resultado do cumprimento de novas exigências da comunidade de imagens médicas e vem conseguindo atingir os requisitos estabelecidos pela *caBIG In Vivo Imaging Workspace* ([RUBIN et al., 2008](#)). O projeto é centrado no modelo AIM de anotação e marcação de imagens DICOM. As anotações e marcações de imagens são consideradas como um objeto de informação da imagem, porém não estão ligadas fisicamente as mesmas. Esse modelo de informação, no entanto, é compatível com outros formatos de imagens além do DICOM. O modelo AIM não cria nenhuma nova ontologia para descrever anotações, ele usa termos do RadLex para descrever, lesões e junta a eles informações numéricas, como volume ou diâmetro.

O modelo AIM de informação, ou AIM Schema, é descrito em [Rubin et al. \(2008\)](#), [Rubin et al. \(2014\)](#) usando um diagrama de classe UML. O modelo é usado para expressar a captura de informações contidas nas imagens médicas através de anotações e marcações de informações relevantes. Uma anotação pode ser explicativa ou uma descrição de informação, gerada por seres humanos ou por máquinas, que se relaciona diretamente com o conteúdo de uma imagem ou várias imagens de referência. As anotações tornam-se uma coleção de descrições de conteúdo que podem ser usadas para fins de mineração de dados semânticos. Uma marcação de imagem é composta de símbolos gráficos ou descrições textuais associados a uma imagem.

Marcações podem ser usadas para descrever uma informação textual e as regiões de interesse visual. As informações nas anotações e marcações são usadas para preencher o AIM Schema, através de funções da biblioteca de software AIM, com a finalidade de geração de objetos DICOM SR, documentos XML AIM e documentos no formato HL7 CDA.

Figura 7 – AIM. Anotações de imagem que descreve, localizações anatômicas, medidas e tipos de lesões



Fonte: Adaptada de [HAGE \(2014\)](#).

AIM e ePad

A ferramenta ePAD usa um vocabulário controlado radiológico como RadLex para definir observações, e impõe a descrição completa dos aspectos necessários da lesão visualizada ([GIMENEZ et al., 2011](#)). Para estudos do câncer, as observações incluem identificação da imagem, momento de obtenção de imagem, nome da lesão, localização anatômica, tipo de lesão, comprimento do segmento de marcação visual da lesão na imagem. Na figura 7 podemos ver os dados de uma anotação(observação) AIM que a ferramenta ePAD armazena normalmente.

Para garantir que as informações sob marcação mínimas são recolhidos, ePAD usa um mecanismo chamado "*AIM templates*". *AIM templates* especifica um formulário de coleta de dados eletrônicos, análogo a um formulário eletrônico de relatório do caso, contendo elementos de dados cujos valores são especificados pelo usuário durante a visualização das imagens. Os elementos de dados em *AIM templates* especifica os tipos de valores válidos, a cardinalidade, e que valores são necessários. Cada *template* é usado para garantir que o mínimo exigido de dados são coletados para um determinado tipo de anotação. *Templates* como RECIST⁵ (*Response Evaluation Criteria In Solid Tumors*) por exemplo, definem quando pacientes de câncer melhoram, ficam na mesma ou pioram. Atualmente não existe uma representação ontológica da infraestrutura AIM que permita inferir e classificar marcações armazenadas por ferramentas como ePAD. Além disso, atualmente não existe o *template*

⁵ <http://www.irrecist.com/>

general para o sistema de staging *TNM*.

Considerações Finais

A fim de melhorar a forma de trabalho na área de saúde, as tecnologias da Web Semântica podem auxiliar as pesquisas médicas e encontrar soluções através do auxílio de computadores, como, por exemplo, o uso dos formatos OWL, SWRL e AIM para processar (*reasoning*) anotações de imagens médicas. O AIM vem sendo desenvolvido e elaborado, tanto para ajudar na estruturação dos relatórios radiológicos, principalmente imagens, como também para ser entendido por máquinas. Esses avanços vem contribuindo no entendimento, interoperabilidade e compartilhamento de conhecimento biomédico na área de radiologia. No caso de radiologia, os padrões como o Radlex e o AIM são utilizados por ferramentas como o ePAD para gerar relatórios diretamente processáveis por máquinas. O próximo capítulo tem como objetivo listar os principais trabalhos que abordam os conceitos de web semântica na classificação de câncer na área de bioinformática.

Trabalhos Relacionados

Neste capítulo apresentaremos o mapeamento sistemático realizado com o objetivo de investigar, catalogar e classificar os trabalhos recentes e relacionados com a classificação de câncer na área de bioinformática.

O mapeamento sistemático é um tipo de revisão sistemática onde se realiza uma revisão mais ampla dos estudos primários, em busca de identificar quais evidências estão disponíveis e as lacunas existentes no conjunto dos estudos primários (KITCHENHAM; CHARTERS, 2007).

Além dos resultados obtidos ao final da revisão, este capítulo também inclui o detalhamento das atividades intermediárias realizadas, sendo elas: o planejamento da revisão, a estratégia adotada para selecionar e utilizar as máquinas de busca, a seleção de trabalhos, entre outros itens.

Planeamento e Execução do Mapeamento Sistemático

O foco deste mapeamento sistemático é identificar, catalogar, e classificar os trabalhos recentes na literatura na área, com o intuito de contribuir de forma substancial no entendimento dos mesmos.

As palavras-chave utilizadas como *strings* de busca são as seguintes: "staging cancer", "semantic image annotations" e "Image metadata reasoning".

O método utilizado para o levantamento de fontes primárias compreendeu a realização de buscas em bibliotecas digitais, como ACM Digital Library¹, IEEE Xplore Digital Li-

¹ <http://dl.acm.org/>

brary², SciELO (Scientific Electronic Library Online)³, DBLP (Computer Science Bibliography)⁴, BDBComp (Biblioteca Digital Brasileira de Computação)⁵, Google Acadêmico (Scholars), Springer⁶, PubMed(US National Library of Medicine National Institutes of Health)⁷.

Os critérios definidos para inclusão dos estudos são apresentados a seguir:

1. Os estudos devem ter sido publicados nas Conferências, Journal e Workshops .
2. Os estudos devem estar escritos em inglês ou português;
3. Os estudos devem estar disponíveis na web;
4. Os estudos devem apresentar alguma das strings de busca em seu título, resumo/abstract ou palavras-chave;
5. Os estudos devem apresentar a proposta de um ou mais sistemas do *cancer staging*.

Cada estudo, analisado de acordo com o método estabelecido para a pesquisa de fontes primárias, é avaliado de acordo com os critérios para inclusão. O processo de busca foi executado utilizando as palavras chave definidas. A consulta obteve 120 estudos publicados entre 2003 e 2016. Esses estudos foram selecionados para uma análise mais detalhada.

Dessa forma, depois de ler o resumo e conclusões 23 estudos foram pré-selecionados, os quais foram lidos e verificados através dos critérios de inclusão estabelecidos.

Dos 23 estudos pré-selecionados, 12 estavam de acordo com o critério de qualidade previsto no protocolo de revisão e tiveram seus dados extraídos e analisados.

Na seção seguinte, os artigos selecionados serão descritos e comparados com o trabalho proposto.

Apresentação e Análise dos Resultados Obtidos no Mapeamento

Nesta seção foi realizada a catalogação bibliográfica dos 11 estudos selecionados no mapeamento sistemático, a qual é apresentada na Tabela 3. Os artigos foram classificados de acordo com: quais tipos de câncer são considerados (Cancer Type), os dados de entrada utilizados (Data Input), método de classificação (*Staging or Grading*) e se os trabalhos têm resultados explícitos (Results).

² <http://ieeexplore.ieee.org/>

³ <http://www.scielo.org/>

⁴ <http://dblp.uni-trier.de/>

⁵ <http://www.lbd.dcc.ufmg.br/bdbcomp/>

⁶ <http://www.springer.com/gp/>

⁷ <http://www.ncbi.nlm.nih.gov/pubmed>

Tabela 3 – Comparativo entre os trabalhos relacionados.

Title (year)	Cancer Type	Data Input	Knowledge source – Grading / Staging method	Results
Zillner e Sonntag (2012)	Lymphoma	Automatic tagging of medical image using OWL ontologies	Ontology-based representation of AAS ⁸ – Staging (OWL-DL reasoning)	No
Zillner (2009)	Lymphoma	Semantic annotation of patient records by tool support (mapped to OWL)	Ontology-based representation of AAS – Staging (OWL-DL reasoning)	No
Tutac et al. (2008)	Breast	Image recognition (pathology images) - automatic tagging	Ontology-based representation of NGS ⁹ – Grading (OWL-DL, SWRL reasoning)	Yes
Tutac, Cretu e Racoceanu (2010)	Breast	Automatic tagging (pathology images)	Ontology-based Spatial representation of NGS – Grading (OWL-DL, SWRL reasoning)	No
Meriem, Yamina e Pathology (2012)	Breast	WFML annotations mapped to OWL (pathology images)	Ontology-based representation of NGS – Grading (OWL-DL, SWRL reasoning)	Yes
Levy, O'Connor e Rubin (2009)	Liver	AIM image annotations mapped to OWL	No grading or staging system – OWL, SWRL, SQWRL reasoning	No
Marquet et al. (2007)	Glioma	Pathologic reports (OWL classes filed manually)	Ontology-based representation of WHO Grading System – Grading (OWL-DL reasoning)	No
Dameron et al. (2006)	Lung	Simulated patient conditions data represented using classes	Ontology-based representation of TNM ¹⁰ – Staging (OWL-DL reasoning)	Yes
Oberkampff et al. (2012)	Lymphoma	Semantic annotations (mapped to OWL)	No grading system	Yes
Kurtz et al. (2014)		Manual semantic annotations from ePAD	No representation of Staging System – Classification (Ontological similarity between terms)	Yes
Racoceanu e Capron (2015)	Breast	WFML annotations mapped to OWL (pathology images)	Ontology-based representation of NGS – Grading (OWL, N3Logic reasoning)	No
Massicano et al. (2015)	Breast	Simulated patient conditions data	Ontology -based represtantion of TNM linked to ICN-O	Yes

A seguir, esses trabalhos são descritos e comparados com o trabalho proposto:

- [Dameron et al. \(2006\)](#) mostram que a classificação de tumores pode ser executada automaticamente usando OWL-DL. Ele concentra-se na classificação dos tumores de pulmão. Os autores usam a ontologia Foundational Model of Anatomy (FMA) para descrever entidades anatômicas a fim de representar os critérios da *TNM*. Além disso, os autores destacam algumas das limitações da Lógica de Descrição (*DL*), para a tarefa de classificação dos tumores de pulmão. Finalmente, as condições dos pacientes são simuladas por instâncias de classes para os experimentos. Esse trabalho não apresenta resultados explícitos usando conjuntos de dados reais.
- Em [Marquet et al. \(2007\)](#) desenvolveram uma ontologia, baseada na ontologia NCI Thesaurus, para classificação dos tumores de glioma. Neste trabalho o sistema de referência de classificação utilizado foi o da *World Health Organization* (WHO). Em seguida, o *reasoning* baseado em classes foi utilizado, o trabalho consegue classificar corretamente dez relatórios clínicos simulados dos onze utilizados no teste. Finalmente os autores não apresentam resultados explícitos usando dados reais.
- Em [Levy, O'Connor e Rubin \(2009\)](#), uma metodologia e ferramentas foram desenvolvidas para transformar o modelo de informações *Annotation and Image Markup* (AIM) na sua versão 3.0 em instâncias OWL. Em seguida, um mecanismo de *reasoning* baseado em OWL e sua linguagem de regras associada SWRL, foi desenvolvido para trabalhar com essas instâncias a fim de classificar e calcular a carga tumoral RECIST ([NISHINO et al., 2010](#)). Embora [Levy, O'Connor e Rubin \(2009\)](#) permite a inferência automatizada de informação semântica sobre lesões de câncer em imagens, as tarefas mais complexas como de avaliar automaticamente o *staging* de câncer (a resposta do tumor ao tratamento) não são suportadas.
- O trabalho proposto em [Tutac et al. \(2008\)](#) introduz uma abordagem que usa OWL-DL a fim de gerar uma correspondência entre os conceitos médicos relacionados com a classificação de câncer de mama (baseados no sistema *Nottingham Grading System – NGS*), para conceitos de visão computacional: Como, por exemplo, *Cell Cluster* \rightarrow *union of cells* e *Lumina* \rightarrow *white compact segments*). Os autores criaram um método baseado em regras para gerar automaticamente o *NGS* e modelar todos esses conceitos em OWL usando o Protégé (uma ferramenta para a edição de ontologias), com a finalidade de definir regras que são utilizadas para a indexação semântica das imagens de pacientes com câncer de mama. Embora a proposta de [Tutac et al. \(2008\)](#) seja desenvolvido no campo da análise de imagens histológicas para a indexação e não exatamente para uma

classificação de pacientes, a abordagem apresentada é importante no desenvolvimento desta proposta, já que descreve claramente as regras utilizadas para a classificação do câncer com base num sistema de classificação, como o *NGS* similar do TNM.

- Em [Zillner \(2009\)](#) uma ontologia OWL-DL foi projetada para realizar o *staging* automático de pacientes com linfoma, usando anotações semânticas. Essas anotações são armazenadas numa ontologia pela ferramenta de anotação que faz parte de projeto *ME-DICO-ontology*. A fim de classificar os pacientes com linfoma, os autores capturaram as regras do sistema de classificação *Ann Arbor* em numa ontologia. Os dados utilizados pertencem ao projeto *privativo* por tanto não estão disponíveis para verificação dos resultados e não são mostrados de forma explícita.
- O principal objetivo em [Tutac, Cretu e Racoceanu \(2010\)](#) é orientar a exploração de imagens dentro de um microscópio virtual e auxiliar no processo de *grading* de imagens histológicas do câncer de mama. Para atender a esses requisitos, os autores estenderam a ontologia desenvolvida em seu trabalho anterior [Tutac et al. \(2008\)](#), ela representa o *grading* do câncer da mama. Nesse trabalho, eles incorporam conhecimento espacial, usando OWL-DL, baseando-se numa teoria formal para desenvolver relações espaciais de *grading* do câncer da mama. Em seguida, eles descrevem como fazer o *reasoning* a partir dessas relações. Mas o trabalho de [Tutac, Cretu e Racoceanu \(2010\)](#) não faz nenhuma avaliação do modelo e os dados não estão disponíveis para verificação dos resultados.
- O trabalho proposto em [Zillner e Sonntag \(2012\)](#) introduz uma aplicação que utiliza anotações de imagens médicas e tecnologias de *reasoning* para classificar automaticamente pacientes com linfoma, com base na posição relativa do linfoma. A fim de verificar ou falsificar as posições de linfoma, conhecimento baseado em oncologia foi utilizado *spatial-anatomical reasoning* ([MÖLLER; SONNTAG; ERÑST, 2013](#)). Em seguida, os autores estabelecem alinhamentos entre conceitos do *RADLEX* e *FMA* a fim de criar a *Patient-ontology*. Essa ontologia representa as informações do paciente. Finalmente, o sistema de *staging Ann Arbor* de linfoma e a *Patient-ontology* foram alinhados para executar um processo de *reasoning*. Os dados utilizados não estão disponíveis para verificação dos resultados.
- O trabalho em [Meriem, Yamina e Pathology \(2012\)](#) propõe uma metodologia para melhorar a usabilidade dos algoritmos de análise de imagem de baixo nível, por exemplo, no modelo clínico que realiza a pontuação de câncer de mama. Essa pontuação é baseada no *Nottingham Grading System (NGS)* para realizar o *grading* do câncer da

mama. Eles desenvolveram uma ontologia chamada *Breast Cancer Ontology (BCO)* com base em OWL-DL e SWRL ([TUTAC et al., 2008](#)). Mais tarde, a fim de descrever as imagens histopatológicas usando anotações semânticas, os autores usaram o formato Wide Field Markup Language (WFML), que é um formato em XML específica para o campo de histopatologia. Esses arquivos no WFML foram traduzidos em OWL. Por fim, a linguagem de consulta SPARQL é usado para acessar as informações disponíveis no sistema.

- Em [Oberkampff et al. \(2012\)](#), os autores propõem uma ontologia que representa formalmente doenças, sintomas e seus respectivos relacionamentos. Eles realizam inferência (*reasoning*) sobre doenças, baseando-se em anotações que refletem o conteúdo descritivo dos dados do paciente. Os autores descrevem o processo para representar o conhecimento médico em uma ontologia chamada *Disease Symptom Ontology*. Finalmente, eles testaram o modelo usando um protótipo. O trabalho de [Oberkampff et al. \(2012\)](#) não faz *staging* de câncer, os autores detalham o processo apenas para modelar o conhecimento em ontologias para interpretar anotações, que são geradas quando são descritos os relatórios de pacientes. Além disso, os dados utilizados pertencem ao projeto *MEDICO* e não estão disponíveis para verificação dos resultados.
- Em [Kurtz et al. \(2014\)](#) e [Seifert et al. \(2009\)](#) propõem um *framework* para recuperação de imagens baseado em anotações semânticas a partir de imagens do fígado. Esse *framework* incorpora as correlações semânticas, entre os termos usados, para descrever essas imagens. Os autores utilizam a semelhança entre termos semânticos (anotações semânticas), que descrevem as lesões do fígado, utilizando a estrutura da ontologia *Radlex* para avaliar a proximidade semântica entre os termos de interesse. Essa abordagem automática fornece suporte em tempo real para radiologistas, mostrando-lhes imagens semelhantes com diagnósticos associados. No trabalho de [Kurtz et al. \(2014\)](#) e [Seifert et al. \(2009\)](#) ontologias não são usadas para representar um sistema de *staging* de câncer. A similaridade entre as imagens é utilizada para classificar as lesões, mas vale a pena notar que as imagens utilizadas são descritas por anotações semânticas baseadas em observações manuais feitas por radiologistas usando o *ePAD* [Rubin et al. \(2014\)](#).
- Em [Racoceanu e Capron \(2015\)](#) é descrito um protótipo que tem a função de controlar um protocolo de análise de imagens histológicas. Esse protótipo foi desenvolvido no *framework* "Cognitive Microscopy" (MICO) a fim de melhorar a análise de imagens histológicas *Whole Slide Image* (WSI) e tornar-se uma avaliação confiável para a classificação de câncer de mama. O protótipo consiste em dois componentes principais,

um núcleo semântico e um conjunto de algoritmos de processamento de imagem. O núcleo semântico tem dois componentes principais que são um motor de *reasoning* e uma ontologia usando N3logic e OWL. N3logic é uma extensão do modelo de dados RDF de tal modo que a mesma linguagem pode ser utilizada para representar lógica e dados. No trabalho de [Racoceanu e Capron \(2015\)](#), os autores usam algoritmos para extração automática de anotações para descrever as imagens histológicas. Além disso, os autores não descrevem explicitamente a ontologia e as regras utilizadas para *grading* de câncer de mama e também não são apresentados resultados explícitos.

- Em [Massicano et al. \(2015\)](#) desenvolveu-se uma ontologia para representar conceitos e axiomas para inferência do *staging* clínico de tumores de um modo generalizado, baseado no *International Classification of Diseases for Oncology* (ICD-O) e as classes que representam conceitos T, N, M. Fundamentalmente os autores criaram um script para analisar arquivos de texto e criar arquivos no formato RDF/XML com a definição de classes e axiomas para inferência do ***staging***, esses arquivos contêm informações como os pares dos valores T, N ou M e uma topografia (descrição e representação do órgão afetado pelo câncer). Eles criaram manualmente esses arquivos de texto com base nas definições TNM. Embora [Massicano et al. \(2015\)](#) tenta generalizar o sistema de classificação como o TNM representado numa ontologia, ele faz inferência só com conceitos de estágio clínico e não da definição de cada critério do TNM, como por exemplo o critério T que é diferente para cada topografia.

Ao comparar o trabalho proposto com os trabalhos acima, as seguintes diferenças foram evidenciadas:

Trabalhos como [Dameron et al. \(2006\)](#), [Marquet et al. \(2007\)](#) executam o *reasoning* baseado em classes, que teoricamente é correto, mas, na prática, tem limitações como a criação de classes desnecessárias, aumentando a complexidade de uma ontologia. Acompanhando a limitação acima mencionada aparece outra, chamada Open World Assumption (OWA) que é um fundamento da OWL que os esses trabalhos solucionam evitando o reasoning baseado em indivíduos. No entanto, é possível realizar o *reasoning* baseado em indivíduos, suportados por estruturas de dados que serão descritos em mais detalhe na seção seguinte.

Os trabalhos como [Tutac et al. \(2008\)](#) e [Zillner \(2009\)](#), eles criam ontologias em OWL-DL para TNM, no entanto a ideia de ter uma ontologia para cada tipo de topografia (órgão do corpo) é indesejável. Nós acreditamos na abordagem de ter uma ontologia que representa diretamente as características encontradas em uma imagem (como a ontologia do modelo AIM) e, além disso, que o *staging* deve ser guiado apenas por regras e axiomas.

Embora o trabalho de [Meriem, Yamina e Pathology \(2012\)](#) e [Tutac, Cretu e Racoceanu \(2010\)](#) utiliza anotações semânticas em imagens e executa uma classificação baseada no NGS utilizando OWL e SWRL, tem um fator limitante que é criar uma nova ontologia, dependendo da patologia que deseja analisar. Isso entretanto, não ocorre em nossa abordagem que fornece uma ontologia baseada em um meio padrão de adição de informação e conhecimento para uma imagem em um ambiente clínico.

Quanto ao trabalho mais parecido com a nossa proposta é do [Zillner e Sonntag \(2012\)](#), no entanto ignorando o fato que os dados utilizados não estão disponíveis para toda a verificação necessária, o sistema do *staging* de linfoma que foi implementado é relativamente mais simples do que um sistema de classificação(*staging*) mais geral como TNM. Por exemplo, pode-se dizer que, [Zillner e Sonntag \(2012\)](#) não considera a dimensão de uma lesão como um factor importante. No entanto, esse fato é muito importante em sistemas de *staging* de cancer mais complexos como TNM. Além disso, podemos ver que o processo de alinhamento de todas as ontologias geradas em [Zillner e Sonntag \(2012\)](#) não são descritos de forma explícita. Finalmente, acreditamos que nossa abordagem pode ser generalizada para outros tipos de casos de uso de reasoning baseado em imagem, o uso do modelo AIM estabelece um fluxo de trabalho automatizado, onde as anotações da imagens produzido por um sistema de imagem(qualquer que gera arquivos AIM), são transformadas á OWL, em seguida essas anotações são utilizadas para gerar reasoning necessárias para tarefa de diferentes dominios.

Considerações Finais

Este capítulo abordou os principais trabalhos relacionados ao desenvolvimento de *staging systems* com base em anotações semânticas de imagens, nas áreas de informática radiológica e oncologia. Foram , descritas as similaridades e divergências entre o trabalho proposto e os 12 trabalhos relacionados que foram escolhidos.

Na literatura, encontraram-se sistemas semelhantes onde as anotações semânticas são armazenados em diferentes formatos que não permitem a integração das mesmas para processos de *reasoning* e, muitas vezes, esses formatos são proprietários.

Alguns desses estudos também permitem a criação de anotações de imagem no formato AIM (na linguagem XML), mas este não é adequados para o raciocínio (*reasoning*). O formato AIM fornece apenas um formato de transferência e armazenamento.

Este trabalho está focado em ajudar os especialistas em câncer na classificação automática de pacientes (*staging*) usando anotações semânticas em imagens. Essa classificação será feita usando-se *reasoning* semântico sobre as anotações, codificadas em AIM.

Metodologia

O foco deste trabalho é a construção de um sistema para permitir o uso de *reasoning* em anotações contidas em imagens médicas. Este capítulo irá descrever a metodologia utilizada para alcançar esse fim. A nossa abordagem é composta por três tarefas: Na Seção 5.1 [5.1](#) A transformação do modelo de informação AIM em uma representação ontológica que define um modelo de informação semanticamente equivalente. Em seguida, na Seção 5.2 [5.2](#), se descreve o mecanismo para transformar as anotações AIM-XML para instâncias da ontologia AIM4-O, a seguir na seção 5.3 [5.3](#), é criado as condições para implementar o reasoning baseado em regras TNM que utilizam instâncias OWL, finalmente a seção 5.4 [5.4](#) descreve a representação formal necessária para *staging* utilizando apenas expressividade OWL + SWRL.

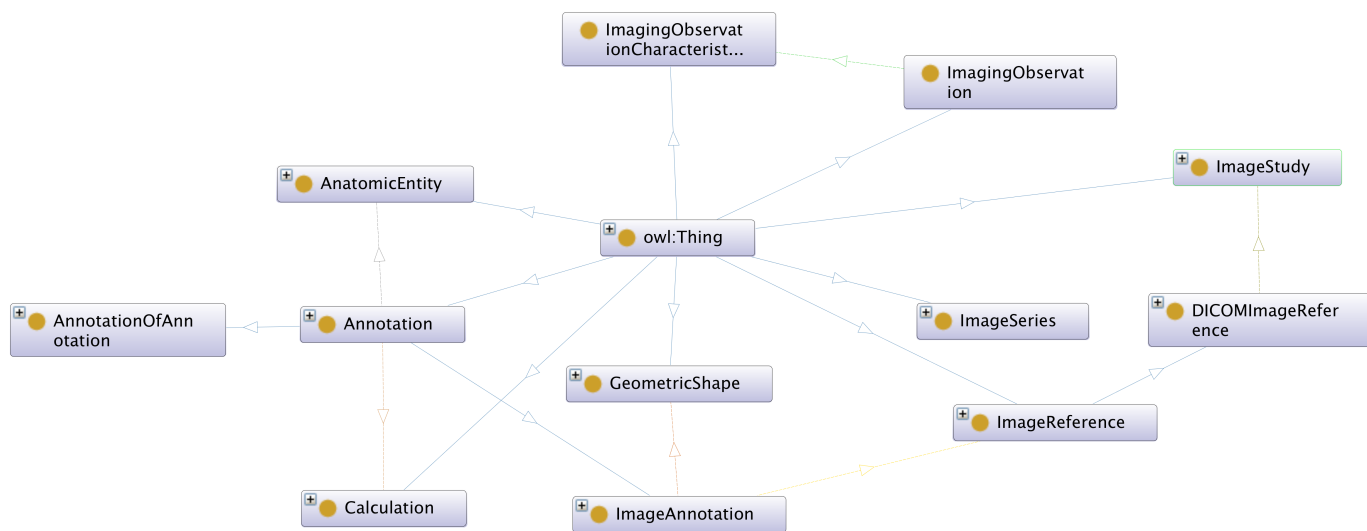
Representação ontológica do modelo de informação AIM

A fim de executar inferência e classificação a partir de anotações e os metadados de marcação de imagem baseados no modelo AIM, nos precisamos de uma linguagem equipado com uma semântica formal. Usando essa semântica, as inferências sobre ontologias e seus indivíduos podem ser feitas. Neste contexto a *Web Ontology Language* OWL descreve uma linguagem para construção de uma representação ontológica semanticamente equivalente de um modelo de informação, tal como por exemplo o modelo de informação *Annotation and Image Markup* AIM . Com base no exposto, neste trabalho transformamos o modelo de dados AIM em uma representação ontológica equivalente, usando OWL2.

Essa transformação foi realizada através da criação de classes e propriedades no OWL , resultando assim compreensível pelo usuário e adequado para inferência. Nós de-

envolvemos uma ontologia com base no trabalho do Hakan Bulu(BULU; RUBIN, 2015), ele desenvolveu uma ontologia. No entanto, essa ontologia representa a versão 3.0 do AIM. Portanto, a fim de representar o modelo AIM na sua versão 4.0, que é a versão utilizada para armazenar anotações de imagem gerados por ferramentas como ePAD(RUBIN et al., 2008), Nós expandimos a ontologia do Hakan Bulu mostrada na figura 8 a seguir:

Figura 8 – Ontologia do Hakan Bulu (BULU; RUBIN, 2015) baseada no modelo AIM 3.0



Fonte: Elaborada pelo autor.

Esta extensão captura lesões e medidas derivadas durante os ensaios clínicos com base em imagens e posteriormente converte as anotações para instâncias da ontologia **AIM4-O**. Essas instâncias são necessárias para o análise das anotações com tecnologias de Web Semântica. A seguir, é descrito em detalhe a ontologia **AIM4-O**.

Ontologia AIM4-O

A ontologia **AIM4-O** (Ontologia do AIM versão 4.0), foi desenvolvido como uma extensão da ontologia do Hakan Bulu 8 e define a codificação OWL2 do modelo de dados AIM 4.0 com o objectivo de apoiar as ferramentas de anotação de imagens especificando a semântica das observações feitas por especialistas e, por conseguinte melhorar a precisão interpretativa dos dados é integrar assistência baseado em computador no processo de interpretação de imagens. As subseções a seguir descrevem o conjunto de classes, propriedades e restrições que constituem a ontologia **AIM4-O**.

Namespaces

Os seguintes prefixos de namespace são usados ao longo deste documento.

Tabela 4 – Prefijos e Namespaces usados na ontologia **AIM4-O**

Prefix	Namespace IRI	Definition
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#	The RDF namespace
xsd	http://www.w3.org/2000/10/XMLSchema#	XML Schema Namespace
owl	http://www.w3.org/2002/07/owl#	The OWL namespace
aim	http://www.w3.org/ns/aim4#	The AIM4-O namespace

Fonte: Elaborada pelo autor.

Dependendo de suas necessidades e de acordo com a quantidade de detalhes que deseja incluir na sua anotação da imagen, o usuário do AIM4-O as vezes só pode precisar usar partes de toda a ontologia. Por esta razão os termos da ontologia AIM4-O (classes e propriedades) são agrupados neste documento em dois categorias a fim de fornecer uma descrição incremental: **Basic Terms** e **Expanded Terms**.

Basic Terms

Fornecem a base para capturar anotações e marcação de informações relevantes para imagens. Através destas classes pode ser descrita explicitamente o tipo de informação que pode coletar o modelo, a informação pode ser de várias disciplinas. O usuario pode estender o modelo para atender suas necessidades, seguindo os critérios descritos no seguinte link ¹.

As classes tipo **Basic Term** já foram definidas na ontologia fornecida por Hakan Bulu (BULU; RUBIN, 2015), em alguns casos, essas classes só mudaram seus nomes adicionando o sufixo "Entity" em comparação com as versões 4.0 e *foundation model* FM do AIM. A seguir encontra-se uma descrição mais detalhada destas classes:

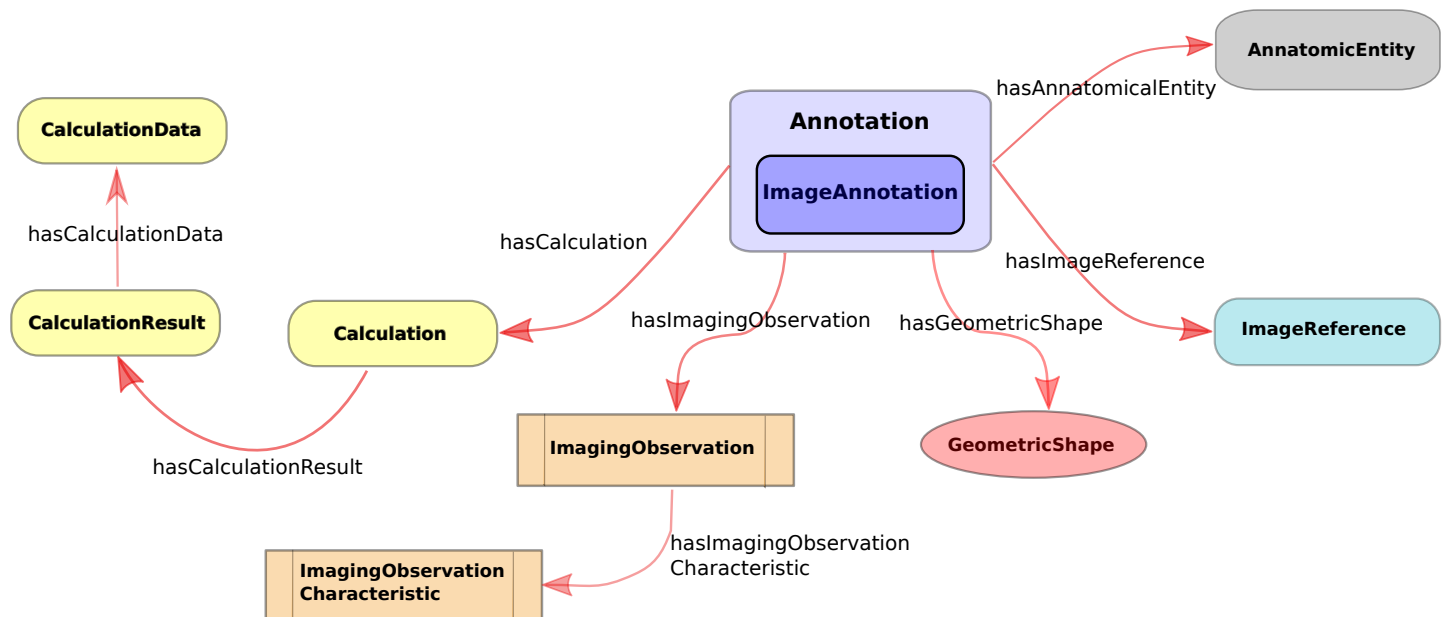
1. **AIM:Calculation:** Representa os resultados do cálculo de uma anotação AIM. Os resultados dos cálculos podem ser o não ser diretamente associado com símbolos gráficos ou marcações.
2. **AIM:GeometricShape:** captura informações e representação gráfica para coordenadas espaciais de tipo bi e tri dimensionais, respectivamente.
3. **AIM:ImagingObservation:** é a descrição de coisas que são vistos em uma imagem.

¹ <https://wiki.nci.nih.gov/display/AIM/Extending+the+AIM+Model#ExtendingtheAIMModel-FromAIM3.0totheAIMFoundationModel>

4. **AIM:ImagingObservationCharacteristics:** a classe contém descritores da classe ImagingObservation.
5. **AIM:Annotation:** captura o nome, uma descrição geral da anotação AIM, o tipo de anotação via terminologia controlada, a data de criação, e uma referência para a versão ao modelo AIM usado.
6. **AIM:ImageReference:** representa uma imagem ou uma coleção de imagens que estão sendo anotada. Os dois tipos possíveis de referências são DICOM e Uniform Resource Identifier (URI).
7. **AIM:ImageAnnotation:** captura informações sobre o significado da informação do pixel em imagens. É usado para anotar uma imagem ou imagens. Ele descreve uma única coisa encontrada.
8. **AIM:CalculationResult:** classe abstrata contém o tipo de resultado (por exemplo , binário, escalar, vetor, histograma, array ou matriz) ,unidade de medida e tipo de dados codificados.
9. **AIM:CalculationData:** classe é usada para armazenar um valor de resultado.

As 9 classes se relacionam entre si utilizando as propriedades mostradas na figura 9 a seguir:

Figura 9 – O diagrama representa as entidades como ovais e retângulos. As propriedades são mostrados em cor vermelho.



Fonte: Elaborada pelo autor.

No seguinte exemplo [5.1](#) descreve os recursos envolvidos na criação da uma anotação sobre uma imagen. O exemplo utiliza apenas termos de tipo **Basic** e serve como ponto para começar a elaboração que será descrita em secções subsequentes.

O exemplo [5.1](#) descreve uma anotação semântica representado como um individuo “:1.2.288.3...” da classe “**AIM:ImageAnnotation**” que captura da imagem; A informação gráfica (coordenadas espaciais), o tamanho, a localização e a descrição da anotação feita na imagem, através dos indivíduos: “:Polyline”, “:LUNG”, “:1.5465798.46468.4566874” cm, “:Conspicuity,Internal Structure”, respectivamente. A ilustração na Figura [10](#), mostra como esses indivíduos do exemplo [5.1](#) se relacionam uns com os outros.

Listagem 5.1 – Representação em sintaxe Manchester do individuo "1.2.288.3.." que representa uma anotação semântica em uma imagem

```

1  @prefix xsd:    <http://www.w3.org/2001/XMLSchema#> .
2  @prefix aim:    <http://www.w3.org/ns/aim4#> .
3  @prefix :       <http://example.org#> .
4  Individual: 1.2.288.3.2205383238.1512.1207945935.1
5      Types:                      ImageAnnotation
6      Facts:
7          aim:hasGeometricShape      :Polyline,
8          aim:hasCalculation          :1.5465798.46468.4566874,
9          aim:hasImageReference       :1.3.6.1.4.1.9328.50.1.10607,
10         aim:hasAnatomicEntity       :Lung,
11         aim:hasImagingObservation   :Conspicuity,
12         aim:uniqueIdentifier         "1.2.288.3.2205383238.1512.1207945935.1"^^xsd:string,
13         aim:dateTime                "2008-04-11T15:32:15"^^xsd:dateTime,
14 .
15 Individual: Polyline
16     Types:                      GeometricShape
17 .
18 Individual: 1.3.6.1.4.1.9328.50.1.10607
19     Types:                      ImageReference
20     Facts:
21         type                      "DICOMImageReference"^^xsd:string
22 .
23 Individual: 1.5465798.46468.4566874
24     Types:                      Calculation
25     Facts:
26         aim:hasCalculationResult     :1.5465798.46468,
27         aim:description              "Linear Measurement"^^xsd:string,
28         aim:uid                      "1.5465798.46468.4566874"^^xsd:string,
29 .
30 Individual: 1.5465798.46468
31     Types:                      ExtendedCalculationResult
32     Facts:
33         aim:hasCalculationData       :1.5465798,
34         aim:type                     "ExtendedCalculationResult"^^xsd:string
35 .
36 Individual: 1.5465798
37     Types:                      CalculationData
38     Facts:
39         aim:values                   3.3422492f^^xsd:float,
40 .
41 Individual: Lung
42     Type:                      AnatomicEntity
43     Facts:
44         aim:codeMeaning              "LUNG"^^xsd:string,
45         aim:codeValue                "REX0001"^^xsd:string,
46         aim:codingSchemeDesignator   "RADLEX"^^xsd:string,
47 .
48 Individual: Conspicuity
49     Type:                      ImagingObservation
50     Facts:
51         aim:ImagingObservationCharacteristic :LIDC Soft Tissue
52         aim:codeMeaning              "Conspicuity"^^xsd:string,
53         aim:codeValue                "REX4001"^^xsd:string,
54         aim:codingSchemeDesignator   "RADLEX"^^xsd:string,
55 .
56 Individual: LIDC Soft Tissue
57     Type:                      ImagingObservationCharacteristic
58     Facts:
59         aim:codeMeaning              "LIDC Soft Tissue"^^xsd:string,
60         aim:codeValue                "REX4021"^^xsd:string,
61         aim:codingSchemeDesignator   "RADLEX"^^xsd:string,
62 .

```


Expanded Terms

As classes deste tipo são uma extensão que representam explicitamente resultados de lesão e medidas derivadas durante os ensaios clínicos com base em imagens, as classes nesta categoria estão listados abaixo:

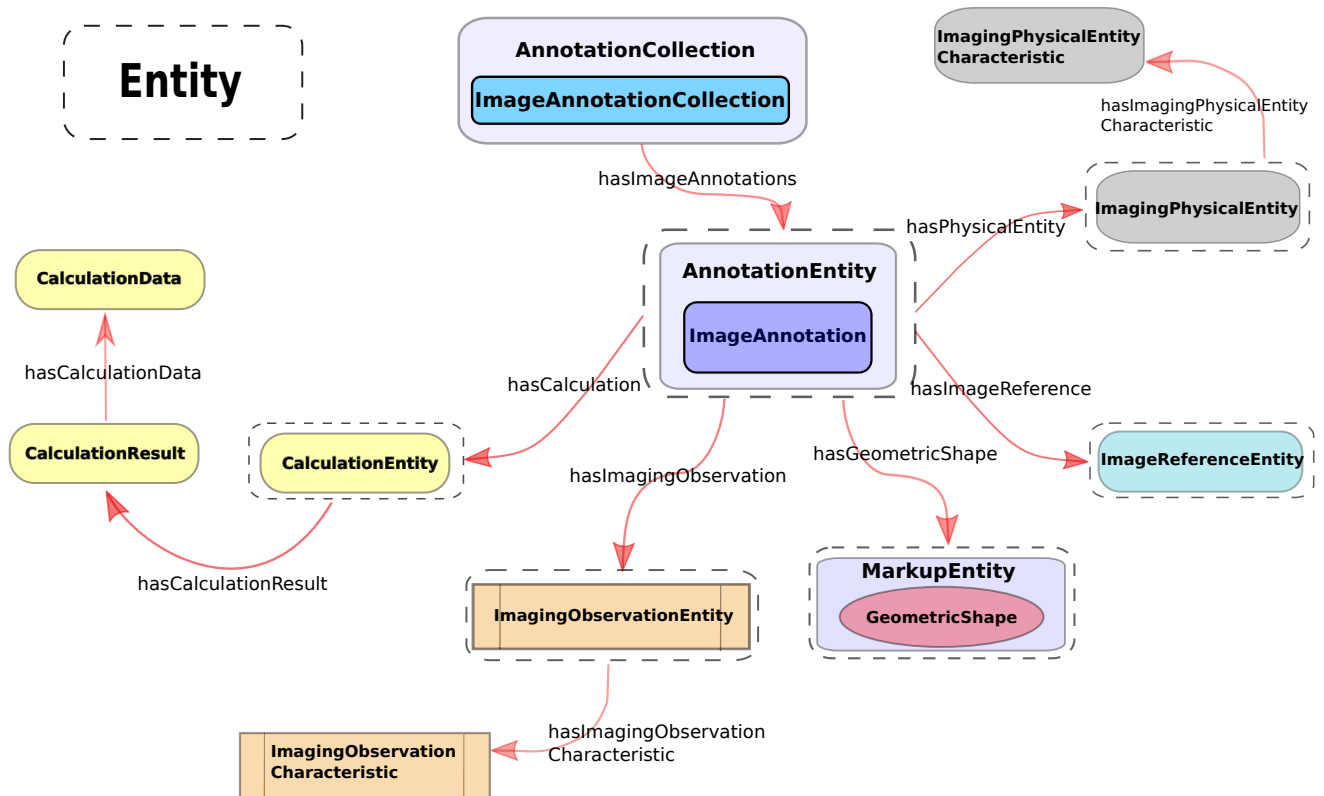
Os termos apresentados nesta seção fornecem as classes adicionais que são necessárias para representar o modelo "AIM 4.0", esse modelo representa explicitamente resultados de lesão e medidas derivadas durante os ensaios clínicos. As classes adicionais encontram-se ilustrados na figura 11:

1. **AIM:Entity:** é uma classe abstrata que representa a existência de uma coisa, conceito, observação, cálculo, medição e desenho gráfico no AIM.
2. **AIM:AnnotationCollection:** que é um conceito abstrato de um recipiente que reúne um tipo de anotações como ImageAnnotation ou AnnotationOfAnnotation.
3. **AIM:imageAnnotationCollection:** armazena instâncias de ImageAnnotation. Se associa com a classe Person que contém informação demográfica do paciente.
4. **AIM:ImagingPhysicalEntity²:** a classe armazena uma localização anatômica como um termo codificado (por exemplo, RID2662 , fêmur, RadLex) a partir de um vocabulário controlado reconhecido como RadLex®, SNOMED CT®, Sistema Único de linguagem médica (UMLS).
5. **AIM:ImagingPhysicalCharacteristic:** descreve ImagingPhysicalEntity tais como termo codificado.
6. **AIM:MarkupEntity:** captura informações textuais e representação gráfica como DICOM-SR e SCORD3D para coordenadas espaciais tri e bi-dimensionais.

As novas classes se relacionam entre si utilizando as propriedades mostradas na figura 11 a seguir:

² A única classe no modelo que foi alterada em totalidade do seu nome para **AnatomicEntity na versão 3.0 do AIM**

Figura 11 – Neste diagrama descreve-se as classes abstratas como ovais roxos envolvendo sua instância (imageAnnotationCollection, ImageAnnotation , geometricShape). Pode-se ver que o nome de algumas classes mudaram em comparação com o modelo **basic** e agora eles tem um sufixo "Entity" o que significa que essas classes são uma instância da classe abstrata principal “**Entity**”.



Fonte: Elaborada pelo autor.

O exemplo seguinte ilustra os termos “**Expanded**” e mostra a codificação OWL (em sintaxe Manchester) de um individuo da ontologia **AIM4-O** que representa uma anotação sobre uma imagen. Como no exemplo da seção anterior o exemplo 5.2 descreve os novos recursos de tipo **Expanded Term** usados na criação da uma anotação.

Como no exemplo da seção anterior, uma anotação e sua representação na forma de individuo **AIM4-O** foi descrita, mas neste exemplo podemos ver que destaca o surgimento de um novo conceito abstrato chamado “**ImageAnnotationCollection**”, ele é um recipiente que armazena instâncias AIM (ImageAnnotations) a partir do mesmo estudo de imagem. Entre algumas outras diferenças em contraste com o exemplo da seção anterior, podemos citar: o surgimento de dataProperties como **typeCode**, que é um atributo que pode capturar informações sobre terminologia controlada em um esquema de codificação simplificado (codeSystemName=**RadLex**, codeSystem=**liver**, code=**RID58**,codeSystemVersion=).

Listagem 5.2 – Representação em sintaxe Manchester de um individuo "Sixmtfhibiqujoz.." da ontologia **AIM4-O**, que representa uma anotação semântica em uma imagem

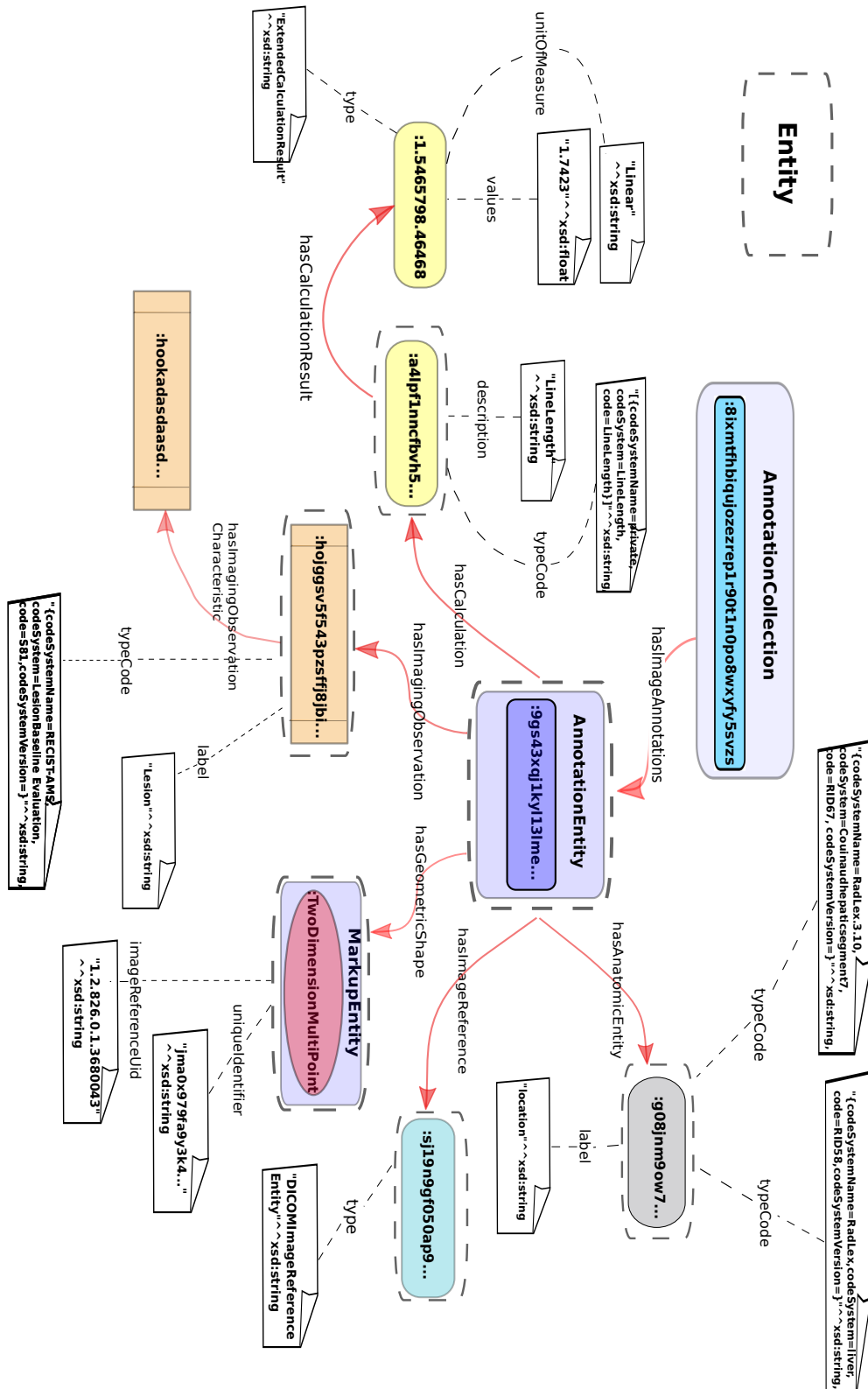
```

1  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
2  @prefix aim: <http://www.w3.org/ns/aim4#> .
3  Individual: 8ixmtfhibiqujozezrep1r90t1n0po8wxyfy5svzs
4      Types: ImageAnnotationCollection
5      Facts:
6          aim:hasImageAnnotations :9gs43xqj1kyl13lmega0zhoenzvgeakkprft8fw8 ,
7      .
8  Individual: 9gs43xqj1kyl13lmega0zhoenzvgeakkprft8fw8
9      Types: ImageAnnotation
10     Facts:
11         aim:hasImageReference :sj19n9gf050ap9uzkc2h98ye4l3rxd71jac28g3g ,
12         aim:hasMarkupEntity :jma0x979fa9y3k4cwypqjvtechwitqb3glvdzjyw ,
13         aim:hasCalculationEntity :a4lpf1nncfbvh5dljf6gfw6r6i3om1ece6270nix ,
14         aim:hasImagingObservation :hojggs5f543pzsfj8jbijgp11thm9qdjow4idd ,
15         aim:hasPhysicalEntity :g08jnm9ow79tijgj1br3q8wd89nq69epjdnxxo30 ,
16         aim:dateTime "2014-10-05T22:15:43"^^xsd:dateTime ,
17         aim:uniqueIdentifier "9gs43xqj1kyl13lmega0zhoenzvgeakkprft8fw8"^^xsd:string ,
18         aim:comment "CT/THORAX 1.0 B45F/273"^^xsd:string ,
19         aim:name "Liver2"^^xsd:string
20     .
21 Individual: sj19n9gf050ap9uzkc2h98ye4l3rxd71jac28g3g
22     Types: DicomImageReferenceEntity
23     Facts:
24         aim:type "DicomImageReferenceEntity"^^xsd:string ,
25         aim:uniqueIdentifier "sj19n9gf050ap9uzkc2h98ye4l3rxd71jac28g3g"^^xsd:string
26     .
27 Individual: jma0x979fa9y3k4cwypqjvtechwitqb3glvImageAnnotationCollectiondzjyw
28     Types: TwoDimensionMultiPoint
29     Facts:
30         aim:type "TwoDimensionMultiPoint"^^xsd:string ,
31         aim:uniqueIdentifier "jma0x979fa9y3k4cwypqjvtechwitqb3glvdzjyw"^^xsd:string
32         aim:imageReferenceUid "1.2.826.0.1.3680043"^^xsd:string ,
33         aim:referencedFrameNumber 273
34     .
35 Individual: a4lpf1nncfbvh5dljf6gfw6r6i3om1ece6270nix
36     Types: CalculationEntity
37     Facts:
38         aim:hasCalculationResult :a4lpf1nncnixExtendedCalculationResult ,
39         aim:description "LineLength"^^xsd:string ,
40         aim:typeCode "[{codeSystemName=private, codeSystem=LineLength,
41         aim:uniqueIdentifier "a4lpf1nncfbvh5dljf6gfw6r6i3om1ece6270nix"^^xsd:string
42     .
43 Individual: a4lpf1nncnixExtendedCalculationResult
44     Types: ExtendedCalculationResult
45     Facts:
46         aim:dateTime "[{codeSystemName=, codeSystem=}]"^^xsd:string ,
47         aim:type "ExtendedCalculationResult"^^xsd:string ,
48         aim:unitOfMeasure "linear"^^xsd:string ,
49         aim:values 1.7432112f
50     .
51 Individual: hojggs5f543pzsfj8jbijgp11thm9qdjow4idd
52     Types: ImagingObservationEntity
53     Facts:
54         aim:annotatorConfidence 0.0f ,
55         aim:label "Lesion"^^xsd:string ,
56         aim:typeCode "{codeSystemName=RECIST-AMS, codeSystem=LesionBaseline,
57         aim:uniqueIdentifier "hojggs5f543pzsfj8jbijgp11thm9qdjow4idd"^^xsd:string
58     .
59 Individual: g08jnm9ow79tijgj1br3q8wd89nq69epjdnxxo30
60     Types: ImagingPhysicalEntity
61     Facts:
62         aim:annotatorConfidence 0.0f ,
63         aim:label "Location"^^xsd:string ,
64         aim:typeCode "{codeSystemName=RadLex, codeSystem=liver,
65         aim:uniqueIdentifier "g08jnm9ow79tijgj1br3q8wd89nq69epjdnxxo30"^^xsd:string
66         aim:uniqueIdentifier "g08jnm9ow79tijgj1br3q8wd89nq69epjdnxxo30"^^xsd:string ,
67         aim:uniqueIdentifier "g08jnm9ow79tijgj1br3q8wd89nq69epjdnxxo30"^^xsd:string ,
68         aim:uniqueIdentifier "g08jnm9ow79tijgj1br3q8wd89nq69epjdnxxo30"^^xsd:string ,
69         aim:uniqueIdentifier "g08jnm9ow79tijgj1br3q8wd89nq69epjdnxxo30"^^xsd:string
70     .

```

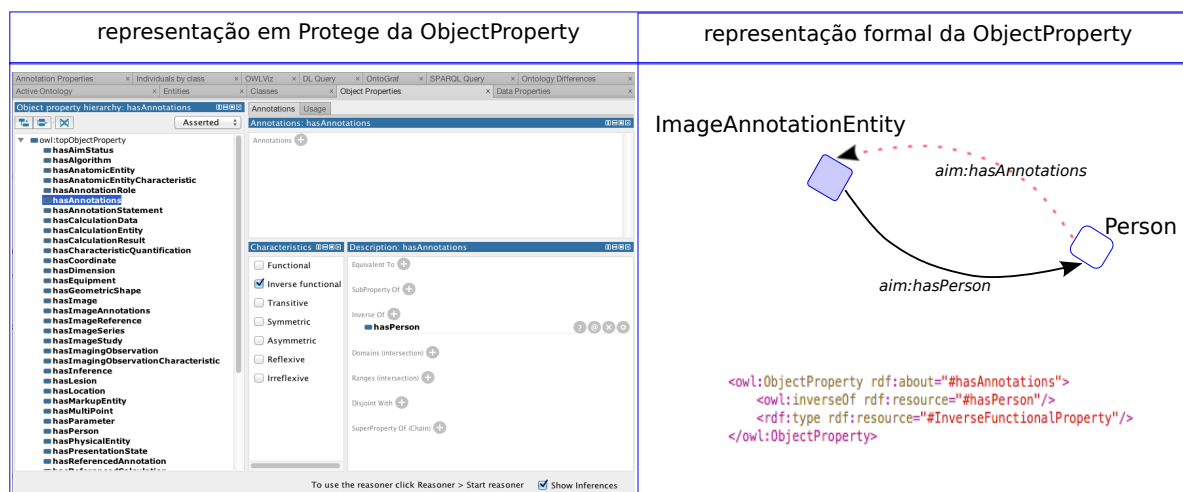
Uma ilustração gráfica do exemplo [5.2](#), mostra como as classes de tipo **Extended** referem se uma a outra na figura [12](#).

Figura 12 – ilustração gráfica do exemplo 5.2 onde as classes de tipo **Extended** e seus indivíduos são mostrados. Ovais e retângulos representam entidades e as propriedades são mostrados em cor vermelha. A figura mostra entidades ovas cercado por linhas pontilhadas, o que significa que estas entidades são classes abstratas de tipo **Entity**.



Neste trabalho, além de adicionar as novas classes, nós também adicionamos algumas propriedades de objetos "**ObjectProperties**" e de dados "**DataProperties**". A fim de conseguir isso, decidimos criar um conjunto de novas relações como: **isImageStudyOf**; **isImageReferenceOf**; **hasAnnotations**; que foram adicionados à ontologia do Hakan Bulu 8 e fazem parte da nova ontologia **AIM4-O**. O objectivo destas relações é para adicionar à AIM4-O a maior expressividade em tarefas de classificação complexas, tal como os critérios do classificação TNM “**_MoreThanOneTumor**”. Cada um de essas relações é uma propriedade inverso funcional das propriedades: **hasImageStudy**; **hasImageReference** e **hasPerson** respectivamente, de modo eles só pode ter um valor. Mais formalmente, se estabelecermos que *hasAnnotations* é *inverseFunctionalProperty* de *hasPerson* podemos dizer que não pode haver duas pessoas diferentes, com a mesma anotação.

Figura 13 – A figura mostra as Object Properties adicionadas à ontologia do Hakan Bulu e fazem parte da nova ontologia **AIM4-O**, podemos ver na figura o Object-Property **hasAnnotations** como propriedade funcional inversa da propriedade **hasPerson**.



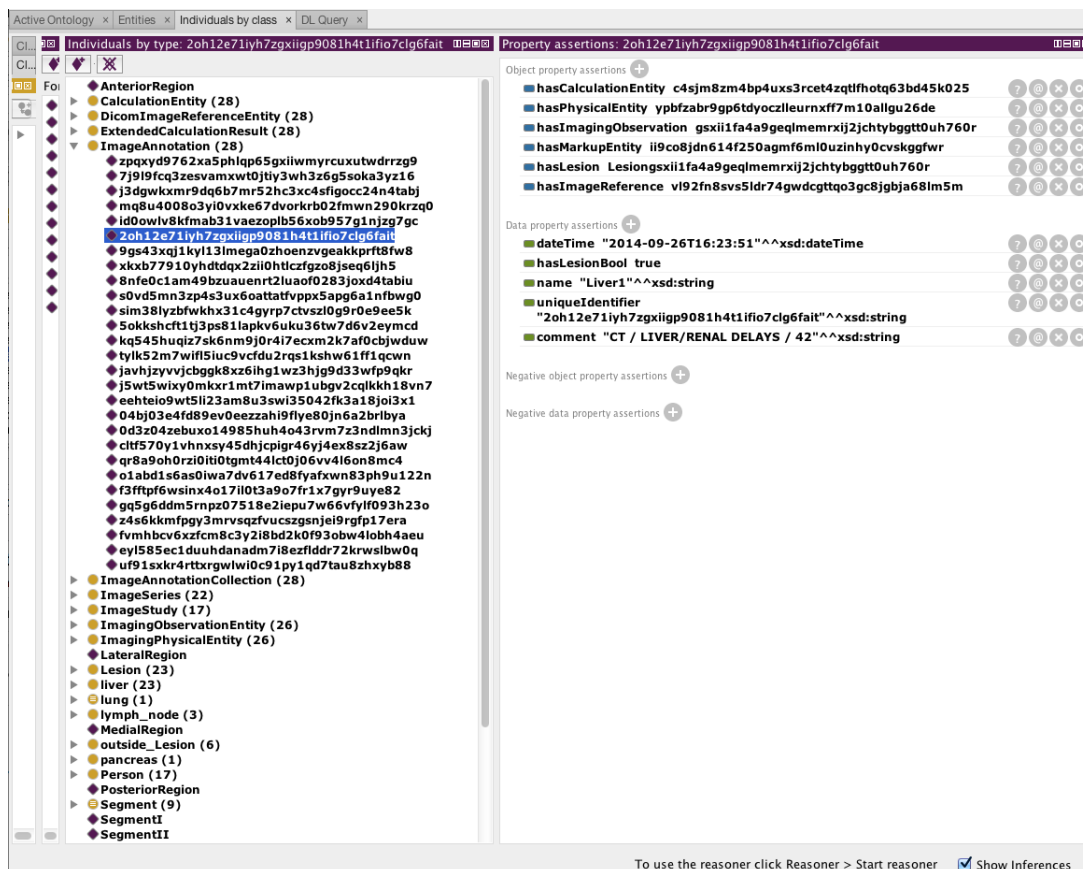
Fonte: Elaborada pelo autor.

AIM4-O é uma ontologia leve que foi desenvolvido com a intenção de ser adotado em uma gama de aplicações e outras disciplinas relacionadas ao processamento de metadados de imagem. **AIM4-O** está de acordo com o perfil **OWL-DL**. Assim, a ontologia espera para ser usado directamente em aplicações, bem como servir como um modelo de referência para a criação de ontologias de domínio específico como por exemplo em radiologia. Para demonstrar o uso das classes AIM4-O e suas propriedades, esse documento usa um cenário de exemplos como acima mencionados.

Desenvolver um mecanismo para transformar as anotações no formato AIM-XML para instâncias da ontologia AIM4-O.

O segundo passo foi definir um mecanismo para transformar as anotações no formato AIM-XML às suas equivalentes codificados em OWL(indivíduos da ontologia AIM4-O). Para isso, desenvolvemos um script para executar esta tarefa. O script usa a linguagem Groovy para mapear automaticamente documentos XML para instâncias das classes java-**AIM**. As classes tipo java-**AIM** representam o conjunto de classes **Java** baseada na representação UML do modelo AIM 4.0. Posteriormente, o **OWL-api** (o java API e implementação para criar, manipular e serializar Ontologias) foi usada para criar instâncias **OWL** destes classes Java, permitindo assim ao script processar fluxos de documentos XML e preencher uma base de conhecimento OWL com o seu conteúdo 14.

Figura 14 – Instâncias de anotações(indivíduos) AIM são mostrados no painel esquerdo, e os detalhes e propriedades de uma instância são mostrados na direita.



Fonte: Elaborada pelo autor.

O objectivo da representação OWL acima, é ter as condições para desenvolver mecanismos de reasoning baseados em regras e trabalhar com anotações do AIM. Para atingir o objetivo da classificação da lesões de câncer, usamos as subclasses, interseção, equivalência e disjunção entre as classes, object properties e algumas regras como mecanismos de raciocínio. Este processo será descrito na próxima etapa.

Gerar as condições para implementar o reasoning baseado em regras TNM que utilizam instâncias OWL.

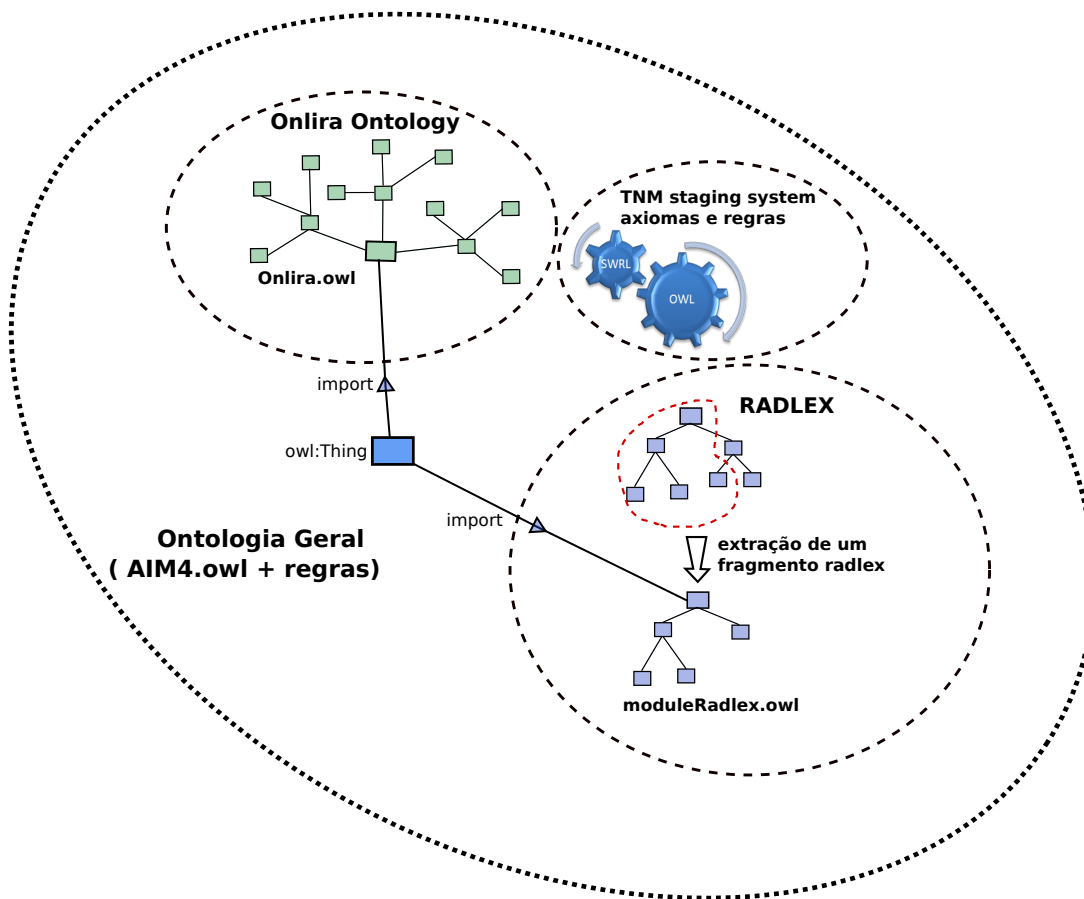
A fim de classificar lesões do um tipo de cancer apartir das imagens e suas anotações precisamos ter o apoio de uma ontologia capaz de especificar a semântica das observações de imagem de um domínio particular, neste caso o corpo humano e mais especificamente um organo que vai ser seleccionado de acordo como o tipo de câncer que vão avaliar(neste trabalho vai ser o câncer de figado). Além disso, considerou-se necessário a inclusão de uma representação em **owl** de uma parte do vocabulário Radlex, a fim de facilitar o trabalho de classificação e tratamento de indivíduos, estes indivíduos têm terminologias de Radlex em sua estrutura. Finalmente, foi adicionado à ontologia gerada o conjunto de regras baseadas no sistema TNM. O resultado é uma **Ontologia geral** construída a partir das exigências acima mencionadas.

Ontologia geral

A **ontologia geral** é dividido em 3 arquivo, como motra a Figura 15. Os Arquivos contêm as seguintes informações:

1. Os conceitos gerais do **TNM**(regras SWRL e axiomas)
2. Onlira.owl: a topografia do fígado (com base na Ontologia Onlira ([KOKCIYAN et al., 2014](#)))
3. ModuleRadlex.owl: Uma parte do vocabulário Radlex ³

Figura 15 – O diagrama da ontologia geral, seus componentes TNM (regras e axiomas) e importações necessárias a fim de ter os mecanismos para classificar imagens de câncer de fígado.



Fonte: Elaborada pelo autor.

As seções a seguir fornecem uma descrição das ontologias utilizadas como parte da **Ontologia Geral**.

Criando classes do conceitos do sistema TNM

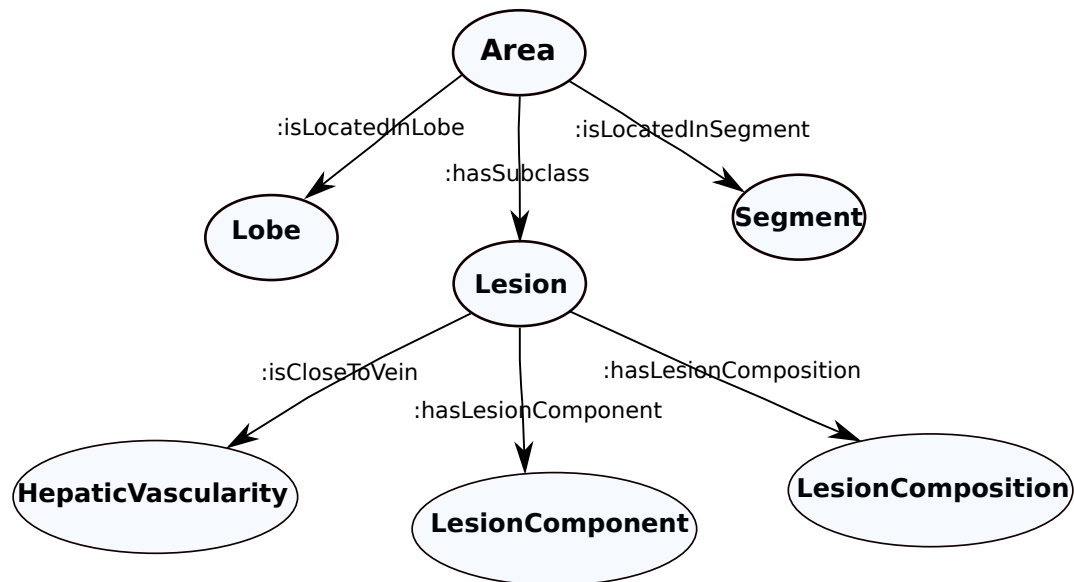
A fim de obter uma representação OWL de cada estágio TNM, precisamos de uma interpretação da sua definição. Embora não seja mencionado explicitamente, os critérios TNM são exclusivos, por isso as classes OWL correspondentes foram feitas disjuntas. Por exemplo, o critério do staging T2 é representado pela união de duas restrições: a primeira, representa um único tumor (de qualquer tamanho) que tem crescido em vasos sanguíneos (classe **T2_a**), ea segunda representa um único tumor não maior do que 5 cm (classe **T2_b**) respectivamente.

ONLIRA (Ontologia de fígado para Radiologia)

A ontologia ONLIRA foi desenvolvida como parte do projeto CaReRa(KOKCIYAN et al., 2014). ONLIRA⁴ visa modelar observações de imagem do domínio do fígado com ênfase em propriedades e as relações entre o fígado, as veias hepáticas e lesões hepáticas. O desenho de ONLIRA baseou-se em sessões de elicitación com radiologistas.

Partes desta ontologia foram utilizadas para representar a topografia do fígado descrito nas anotações AIM-XML gerado pelo ePAD (CARDOSO et al., 2014).

Figura 16 – mostra as relações entre as classes e outros conceitos de ONLIRA utilizados neste trabalho.



Fonte: Elaborada pelo autor.

Terminologia Radlex

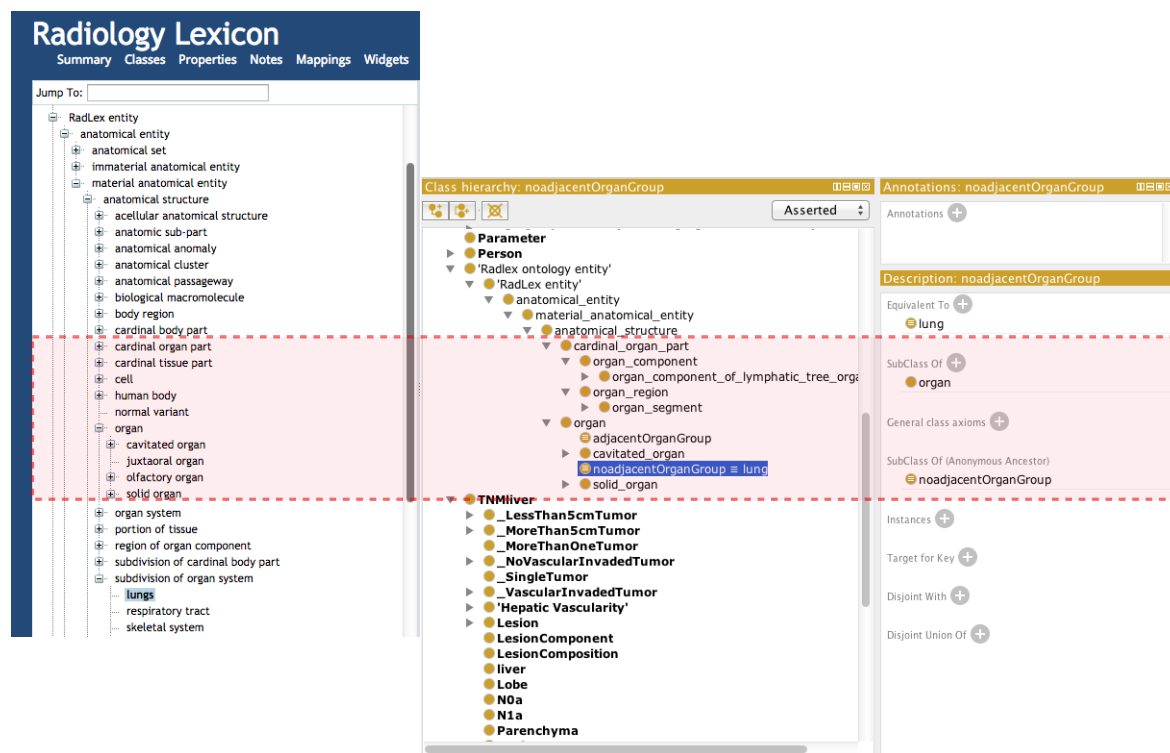
AIM fornece um esquema XML que descreve as estruturas anatômicas e observações visuais em imagens utilizando a terminologia RadLex. Foi usado partes de RadLex para representar essas estruturas anatômicas e representar critérios de classificação TNM em OWL.

Os critérios TNM referem-se a locais de entidades e estruturas anatômicas. Em seguida, é importante que a ontologia contém não apenas as estruturas anatômicas mencionados nos critérios de staging, mas também as partes ligadas direta e indiretamente a eles. Por exemplo, a fim de implementar os critérios N e M do TNM, nós adicionamos 2 Super classes: **adjacentOrganGroup** que descreve o conjunto de órgãos adjacentes a um órgão

⁴ <https://biportal.bioontology.org/ontologies/ONLIRA>

principal (neste caso o fígado) e **noadjacentOrganGroup** que descreve órgãos com base nos locais mais comuns de disseminação do tumor, conforme ao trabalho do (FARIA et al., 2014). Como neste trabalho foi tomado como caso de estudo a classificação do câncer de fígado, incluímos pulmões e ossos como órgãos no adjacentes, isso pode ser visto na Figura 17.

Figura 17 – módulo da ontologia Radlex e as classes que representam grupos de órgãos, tais como **adjacentOrganGroup** e **noadjacentOrganGroup**.



Fonte: Elaborada pelo autor.

Reasoning sobre lesões de câncer

Nas seções anteriores, várias operações foram necessárias para melhorar a **Ontologia Geral**. Em primeiro lugar, nós adicionamos novas classes e propriedades para preencher a lacuna entre as características do tumor descritos no modelo AIM-xml e as classes presentes na ontologia AIM4-O. Em segundo lugar, fornecemos definições para: todos os estadios TNM; a topografia do órgãos como o fígado; parte da terminologia RadLex, a fim de representar os seus conhecimentos em OWL. Finalmente esta seção descreve a representação formal necessária para *staging* do câncer utilizando apenas expressividade OWL se possível (apenas OWL + SWRL).

Representação Formal do *staging* do câncer

O staging do câncer é dividido em duas etapas principais: A primeira etapa consiste em dar uma contagem a partir da descrição do tumor (T); a propagação em gânglios linfáticos (N) e possível metástase (M). O segundo passo consiste em determinar o estágio de acordo com as pontuações anteriores.

A fim de descobrir os limites da capacidade de OWL e as capacidades fornecidas pela adição de regras e axiomas, buscou-se definir formalmente e implementar as condições que exige o sistema TNM, usando apenas expressividade OWL + SWRL.

Decidimos que as seguintes condições reflectiu um processo de staging desejável:

Condição 1: O staging do câncer deve considerar a existência de tumor solitário ou múltiplo no mesmo site

Condição 2: Os tumores deve ser maior ou menor que 5 cm

Condição 3 : Staging deve considerar lesões em órgãos adjacentes

Representação formal das condições usando OWL

Condição 1: "O staging do câncer deve considerar a existência de tumor solitário ou múltiplo no mesmo site"

A ontologia **AIM4-O**, previamente desenvolvida, não tem mecanismos tais como classes, subclasses ou propriedades que nos permitem inferir se um paciente tem tumores simples ou múltiplos explicitamente. No caso de tumores múltiplos, construímos a seguinte regra "**MoreThanOneTumor** " (em notação SWRL) :

$$\begin{aligned} &ImageStudy(?X) \wedge isImageStudyOf(?x, ?y) \wedge isImageStudyOf(?x, ?z) \wedge \\ &DifferentFrom(?y, ?z) \rightarrow MoreThanOneTumor(?x) \end{aligned}$$

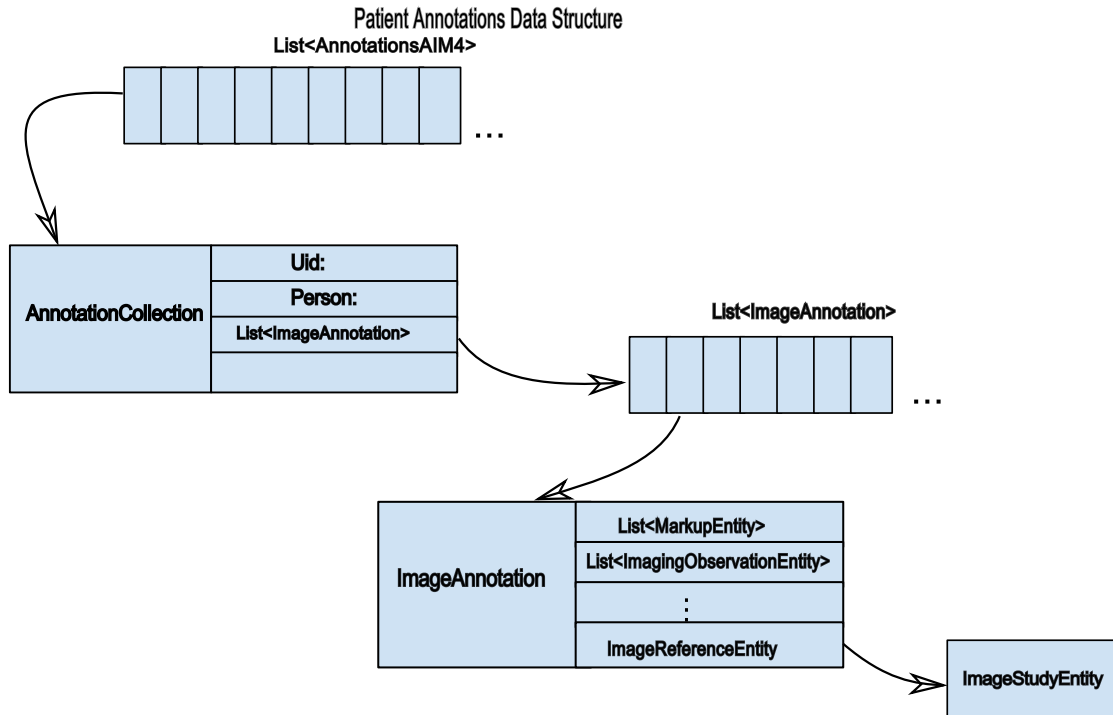
Esta regra classifica um individuo como membro da classe **MorethanOneTumor**, se um estudo de imagem "X" é referenciado por mais de uma anotação da imagem. Para classificar algo tão **MorethanOneTumor**, precisávamos criar um novo conceito chamado: **isImageStudyOf**. Este conceito é a inversa da Objectproperty **hasImageStudy**. A propriedade **hasImageStudy** relaciona uma entidade de tipo **ImageAnnotation** a uma entidade **ImageStudy**.

A fim de classificar os pacientes com exatamente um tumor identificado. Nós não conseguimos encontrar axiomas ou regras que satisfazem esta exigência, porque **OWL** tem

uma pressuposição de mundo aberto ou **Open World**. “O mundo aberto significa que só porque nós não dizer algo, isso não significa que não é verdade”. Por exemplo, pode-se dizer que nós podemos descrever uma lesão de cancer em uma imagen usando a entidade de tipo *ImagingObservation* do modelo AIM, mas, a menos que eu diga explicitamente que não há outras lesões nessa imagen, presume-se que pode haver outra lesão que eu não mencionei explicitamente.

Para resolver o problema de **mundo aberto**(Open World), consideramos algumas alternativas como por exemplo: A necessidade de reformular a nossa ontologia para modelar "**hasImagingObservation**" como uma classe, e não como uma relação ObjectProperty como actualmente na ontologia **AIM4-O**. Mas, isso não parece intuitivo para nós. Em vez disso, decidimos indicar o número de lesões de forma explícita através da criação de um novo conceito chamado: **singleLesion**, como uma ObjectProperty da entidade **ImageStudy**. Este conceito indica se o estudo descreve apenas uma lesão do câncer na imagen. Assumiu-se que o estudo descreve exatamente uma lesão câncer ("**singleLesion { true}**"), se ele é referenciado por uma única anotação na imagem. Mas, Não foi possível formular o pressuposto acima usando apenas **OWL**. Em vez disso, utilizou-se uma estrutura de dados como referencia, a estrutura de dados e mostrada na Figura 18. Esta estrutura de dados foi gerado como parte do processo de *parsing* das anotacoes de imagem no formato XML para criar indivíduos da ontologia AIM4-O.

Figura 18 – é apresentada a definição da estrutura de dados gerado na etapa de *parsing*



Fonte: Elaborada pelo autor.

Finalmente, para classificar anotações que descrevem uma única lesão, construímos a regra "**Tumor Single**" (em notação SWRL):

$$ImageStudy(?X) \wedge singleLesion(?X, ?val) \wedge equal(?val, true) \rightarrow singleTumor(?X)$$

Condição 2: "Os tumores deve ser maior ou menor que 5 cm"

Esta condição é facilmente implementado, obtendo o valor da propriedade "*values*" pertencente à entidade **CalculationResult**. Esta entidade está relacionado com outra entidade de tipo **ImageAnnotation** através da propriedade **hasCalculationEntity**. A fim de satisfazer esta condição de medição, formulamos as seguintes regras (na notação SWRL):

LessThan5cmTumor

$$ImageAnnotation(?x) \wedge hasCalculationEntity(?x, ?y) \wedge hasCalculationResult(?y, ?z) \wedge values(?z, ?val) \wedge lessThan(?val, 5) \rightarrow lessThan5cmTumor(?x)$$

MoreThan5cmTumor

$$ImageAnnotation(?x) \wedge hasCalculationEntity(?x, ?y) \wedge hasCalculationResult(?y, ?z) \wedge values(?z, ?val) \wedge greaterThan(?val, 5) \rightarrow MoreThan5cmTumor(?x)$$

Condição 3 : "Staging deve considerar lesões em órgãos adjacentes"

Para satisfazer o critério mais complicado de classificação tivemos que considerar o fato de que uma lesão pode espalhar-se no corpo. Para isso, foi necessário criar um novo conceito baseado na classe **Lesion** da ontologia **Onlira**¹⁶. Enquanto a classe **Lesion** lida com muitas características importantes de uma lesão tais como: a composição, densidade, tamanho, forma, etc, infelizmente isso não é suficiente para satisfazer os criterios da classificação TNM. Por esta razão, foram adicionados 3 propriedades a ele e, como resultado definimos un novo conceito, a subclasse **OutsideLesion**, As 3 novas propriedades são:

hasLocation ObjectProperty: Esta propriedade indica a localização das lesões com base na taxonomia **Radlex**. A propiedade **hasLocation** relaciona as instâncias da classe **Lesion** da Ontologia Onlira aos membros da classe **AnatomicalEntity** do modulo Radlex usado neste trabalho.

isRegionalLymphNodeAffected ObjectProperty: Especifica se uma lesão está localizada em algum nódulo linfático. Esse conceito foi útil para os critérios de classificação como N0 e N1.

isAdjacentOrgan object property: Indica se a lesão presente no local "X" está perto de órgão "Y". De acordo com o critério de classificação TNM do cancer de fígado, considerou-se como órgãos adjacentes ao fígado: o pâncreas, estômago, duodeno, fígado, cólon, etc. (ver figura 2). Nós agrupamos esses conceitos(usando a taxonomia radlex), criando duas novas classes "**AdjacentOrganGroup**" e "**NoAdjacentOrganGroup**":

$$NoAdjacentOrganGroup \subseteq RadlexEntity.Organ$$

$$AdjacentOrganGroup \subseteq RadlexEntity.Organ$$

AdjacentOrganGroup e *NoAdjacentOrganGroup*, indicam se um órgão é considerado adjacente ou não em referência ao órgão pai. O órgão pai representa o tipo do *staging* de cancer, para o nosso estudo de caso o órgão pai é o fígado. Finalmente, construímos a seguinte regra (em notação SWRL) para classificar se uma lesao fora do figado (representado pelo conceito *OutsideLesion*) está localizada em um órgão adjacente:

$$Lesion(?x) \wedge adjacentOrganGroup(?y) \wedge \\ hasLocation(?x, ?y) \rightarrow isAdjacentOrgan(?x, true)$$

Neste ponto, uma vez formulados os criterios acima adequadamente usando **OWL** e **SWRL**, nós usamos eles para construir os seguintes axiomas e regras, a fim de ser capaz de classificar automaticamente lesões cancerosas utilizando o sistema TNM (ver Tabelas 5, 6, 7). Notamos que a maneira que nós modelamos as coisas importava. Por exemplo, foi mais fácil para definir **N1a**, **N0** e reutilizar suas definições para **M0**, ao invés de começar com a definição de **M0** através da manipulação de closures complexos.

Com a utilização da ontologia **AIM4-O**, os conceitos podem ser facilmente relacionados uns com os outros como foi demonstrado anteriormente. Além disso, os requisitos de integridade e de cardinalidade pode ser especificada e executada com fins de *reasoning*.

Considerações Finais

Neste capitulo, foram descritas as etapas necessárias para alcançar a implementação de um sistema de classificação de cancer baseado em anotações contidas em imagens médicas. Para conseguir isso, foi necessário uma representação formal híbrida baseado em regras SWRL e exiomas dos criterios de classificação TNM. Alem disso, o sistema foi suportado pelas ontologias AIM4-O(extensão do trabalho de Bulu e Rubin (2015)), ONLIRA e Radlex. No processo de implementação foram encontrados problemas como o mundo aberto (Open World) , embora houvesse muitas alternativas para resolver o problema, nosso grupo de pesquisa escolheu por utilizar a mais simple usando uma estruturas de dados.

O próximo capitulo conlui a dissertação consolidando os resultados obtidos nas avaliações e resaltando as principais contribuições e limitações desta pesquisa.

Tabela 5 – American Joint Committee on Cancer/International Union against cancer TNM classification system and formal representation

Primary tumor (T)	
Medical representation for T1 criterion:	Solitary tumor (any size) without vascular invasion
Formal representation for T1 in OWL-DL	$\equiv \text{NotVascularInvadedTumor} \wedge$ $\ni \text{hasImageReferenceImageReferenceEntity} \wedge$ $\text{ImageReferenceEntity} \wedge \ni \text{hasImageStudyImageStudy} \wedge \text{SingleTumor}$
Medical representation for T2a criterion :	Solitary tumor (any size) with vascular invasion
Formal representation for T2a in OWL-DL	$\equiv \text{VascularInvadedTumor} \wedge$ $\ni \text{hasImageReferenceImageReferenceEntity} \wedge$ $\text{ImageReferenceEntity} \wedge \ni \text{hasImageStudyImageStudy} \wedge \text{SingleTumor}$
Medical representation for T2b criterion :	multiple tumors none < 5cm
Formal representation for T2b in OWL-DL	$\equiv \text{LessThan5cmTumor} \wedge$ $\ni \text{hasImageReferenceImageReferenceEntity} \wedge$ $\text{ImageReferenceEntity} \wedge \ni \text{hasImageStudyImageStudy} \wedge \text{MoreThanOneTumor}$
Medical representation for T3a criterion :	Multiple tumors, with at least one tumor > 5cm
Formal representation for T3b in OWL-DL	$\equiv \text{MoreThan5cmTumor} \wedge$ $\ni \text{hasImageReferenceImageReferenceEntity} \wedge$ $\text{ImageReferenceEntity} \wedge \ni \text{hasImageStudyImageStudy} \wedge \text{MoreThanOneTumor}$
Medical representation for T4a criterion :	Tumors with direct invasion of adjacent organs other than the gallbladder
Formal representation for T4a in SWRL	$\text{outsideLesion}(?x) \wedge \text{hasLocation}(?x, ?y) \wedge$ $\text{isAdjacentOrgan}(?x, \text{true}) \rightarrow \text{T4a}(?x)$

Tabela 6 – American Joint Committee on Cancer/International Union against cancer TNM classification system and formal representation

Regional lymph nodes (N)	
Medical representation for N0 criterion :	Solitary tumor (any size) without vascular invasion
Formal representation for N0a in SWRL	$Lesion(?x) \wedge isRegionalLymphNodeAffected(?x, false) \rightarrow N0a(?x)$
Formal representation for N0 in OWL-DL	$\equiv \quad ImageAnnotation.hasLesion \quad \ni \quad N0a$
Medical representation for N1 criterion :	Regional lymph node metastasis
Formal representation for N1a in SWRL	$Lesion(?x) \wedge organcomponentoflymphatictree(?y) \wedge outsideLesion(?x) \wedge hasLocation(?x, ?y) \rightarrow N1a(?x)$
Formal representation for N1a in OWL-DL	$\equiv \quad ImageAnnotation.hasLesion \quad \ni \quad N1a$

Tabela 7 – American Joint Committee on Cancer/International Union against cancer TNM classification system and formal representation

Distant metastasis (M)	
Medical representation for M0 criterion :	No distant metastasis
Formal representation for M0 in OWL-DL	$\equiv \quad ImageAnnotation.hasLesion \wedge \quad \ni \quad N0 \quad \vee \quad N1a$
Medical representation for M1 criterion :	Distant metastasis
Formal representation for M1 in OWL-DL	$\equiv ImageAnnotation.hasLesion \wedge \ni OnIra.Lesion \wedge$ $Lesion.hasLocation \wedge \ni Organ.modifiedOrganGroup$

Experimentos

Neste capítulo, serão apresentados os resultados que foram alcançados

Experimentos

a

b

Referências

AMANQUI, F. K. M. *Uma arquitetura para sistemas de busca semântica para recuperação de informações em repositórios de biodiversidade*. Dissertação (Mestrado) — ICMC-USP, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2014. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-03072014-150009/>>. (Citado nas páginas 53 e 54.)

ARNOLD, H. *Introduction to medical imaging*. São Paulo, 2012. Disponível em: <<http://cw.tandf.co.uk/imagingforstudents/sample-material/Chapter-1-Introduction-to-Medical-Imaging.pdf>>. Acesso em: 02 March. 2016. (Citado na página 40.)

BAADER, F. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003. ISBN 9780521781763. Disponível em: <https://books.google.com.br/books?id=e6_hJtM07qwC>. (Citado na página 54.)

BECHHOFFER, S. et al. *OWL Web Ontology Language Reference*. <http://www.w3.org/TR/owl-ref/>, 2004. (Citado na página 52.)

BELLE, A.; KON, M. A.; NAJARIAN, K. Biomedical informatics for computer-aided decision support systems: a survey. *ScientificWorldJournal*, v. 2013, p. 769639, 2013. [PubMed Central:PMC3575619] [DOI:10.1155/2013/769639] [PubMed:23431259]. (Citado na página 38.)

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. *Scientific American*, v. 284, n. 5, p. 34–43, maio 2001. Disponível em: <<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>>. (Citado na página 50.)

BROWNE, G.; JERMEY, J. *Website Indexing: Enhancing Access to Information Within Websites*. Auslib Press, 2004. ISBN 9781875145560. Disponível em: <<https://books.google.com.br/books?id=BBt0Ir1t2msC>>. (Citado na página 43.)

BULU, H.; RUBIN, D. L. *Java Application Programming Interface (API) for Annotation Imaging Markup (AIM)*. 2015. 1–12 p. <<http://web.stanford.edu/group/qil/wikiupload/4/41/AimApiManuscript.pdf>>. [Online; accessed 02-March-2015.]. (Citado nas páginas 17, 37, 70, 71 e 91.)

CARDOSO, D. S. et al. A gazetteer for biodiversity data as a linked open data solution. In: . [S.l.]: Web2Touch 2014 - Modelling the Collaborative Web Knowledge, 2014. (Citado nas páginas 51 e 85.)

CHANNIN, D. S. et al. The cabigTM annotation and image markup project. *Journal of Digital Imaging*, v. 23, n. 2, p. 217–225, 2010. ISSN 1618-727X. Disponível em: <http://dx.doi.org/10.1007/s10278-009-9193-9>. (Citado na página 37.)

CONNOR, M. et al. Supporting rule system interoperability on the semantic web with swrl. In: *Proceedings of the 4th International Conference on The Semantic Web*. Berlin, Heidelberg: Springer-Verlag, 2005. (ISWC'05), p. 974–986. ISBN 3-540-29754-5, 978-3-540-29754-3. Disponível em: http://dx.doi.org/10.1007/11574620_69. (Citado na página 53.)

DAMERON, O. et al. Grading lung tumors using OWL-DL based reasoning Methods. 2006. Disponível em: http://protege.stanford.edu/conference/2006/submissions/abstracts/8.2_2006protege-dameron-short.pdf. Acesso em: 02 March. 2015. (Citado nas páginas 45, 46, 47, 63, 64 e 67.)

DEPEURSINGE, A. et al. Predicting visual semantic descriptive terms from radiological image data: Preliminary results with liver lesions in ct. *IEEE Transactions on Medical Imaging*, v. 33, n. 8, p. 1669–1676, Aug 2014. ISSN 0278-0062. (Citado nas páginas 33 e 42.)

DOMBAL, F. T. de et al. Computer-aided diagnosis of acute abdominal pain. *Br Med J*, v. 2, n. 5804, p. 9–13, Apr 1972. [PubMed Central:PMCID1789017] [PubMed:4552594]. (Citado na página 38.)

FARIA, S. C. et al. Thm/okuda/barcelona/unos/clip international multidisciplinary classification of hepatocellular carcinoma: concepts, perspectives, and radiologic implications. *Abdominal Imaging*, v. 39, n. 5, p. 1070–1087, 2014. ISSN 1432-0509. Disponível em: <http://dx.doi.org/10.1007/s00261-014-0130-0>. (Citado na página 86.)

GIMENEZ, F. et al. On the feasibility of predicting radiological observations from computational imaging features of liver lesions in ct scans. In: *Healthcare Informatics, Imaging and Systems Biology (HISB), 2011 First IEEE International Conference on*. [S.l.: s.n.], 2011. p. 346–350. (Citado nas páginas 41 e 58.)

GRAU, B. C. et al. Owl2: The next step for owl. *Web Semantics: Science, Services and Agents on the World Wide Web*, Elsevier, v. 6, n. 4, 2008. ISSN 1570-8268. Disponível em: <http://www.websemanticsjournal.org/index.php/ps/article/view/156>. (Citado na página 52.)

GRUBER, T. R. A translation approach to portable ontology specifications. *Knowl. Acquis.*, Academic Press Ltd., London, UK, UK, v. 5, n. 2, p. 199–220, jun. 1993. ISSN 1042-8143. Disponível em: <http://dx.doi.org/10.1006/knac.1993.1008>. (Citado na página 51.)

HAGE, C. C. *Marcação de regiões de interesse em 3d sobre imagens radiológicas utilizando a web*. 2014. Dissertação (Mestrado) — ICMC-USP, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2014. Disponível em: <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-19032015-104453/>. (Citado na página 58.)

HORROCKS, I.; PATEL-SCHNEIDER, P. F.; HARMELEN, F. V. From shiq and rdf to owl: The making of a web ontology language. *Journal of Web Semantics*, v. 1, p. 2003, 2003. (Citado na página 52.)

HU, B. et al. Ontology-based medical image annotation with description logics. In: *Tools with Artificial Intelligence, 2003. Proceedings. 15th IEEE International Conference on*. [S.l.: s.n.], 2003. p. 77–82. ISSN 1082-3409. (Citado nas páginas 38 e 42.)

JIANG, Y. et al. Potential of computer-aided diagnosis to reduce variability in radiologists' interpretations of mammograms depicting microcalcifications. *Radiology*, v. 220, n. 3, p. 787–794, Sep 2001. [DOI:10.1148/radiol.220001257] [PubMed:11526283]. (Citado na página 38.)

KITCHENHAM, B.; CHARTERS, S. *Guidelines for performing Systematic Literature Reviews in Software Engineering*. 2007. (Citado na página 61.)

KOKCIYAN, N. et al. Semantic Description of Liver CT Images: An Ontological Approach. *IEEE Journal of Biomedical and Health Informatics*, v. 18, n. 4, p. 1363–1369, jul 2014. ISSN 2168-2194. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6708417>>. (Citado nas páginas 83 e 85.)

KORENBLUM, D. et al. Managing biomedical image metadata for search and retrieval of similar images. *Journal of Digital Imaging*, v. 24, n. 4, p. 739–748, 2011. ISSN 1618-727X. Disponível em: <<http://dx.doi.org/10.1007/s10278-010-9328-z>>. (Citado na página 43.)

KUNDU, S. et al. The IR Radlex Project: An Interventional Radiology Lexicon-A Collaborative Project of the Radiological Society of North America and the Society of Interventional Radiology. *Journal of Vascular and Interventional Radiology*, SIR, v. 20, n. November 2006, p. 433–435, 2009. ISSN 10510443. Disponível em: <<http://dx.doi.org/10.1016/j.jvir.2008.10.022>>. (Citado na página 32.)

KURTZ, C. et al. On combining image-based and ontological semantic dissimilarities for medical image retrieval applications. *Medical Image Analysis*, v. 18, n. 7, p. 1082 – 1100, 2014. ISSN 1361-8415. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1361841514001030>>. (Citado nas páginas 42, 44, 63 e 66.)

LANGLOTZ, C. P. Radlex: A new method for indexing online educational materials. *RadioGraphics*, v. 26, n. 6, p. 1595–1597, 2006. PMID: 17102038. Disponível em: <<http://dx.doi.org/10.1148/rg.266065168>>. (Citado na página 56.)

LEVY, M. A.; O'CONNOR, M. J.; RUBIN, D. L. Semantic reasoning with image annotations for tumor assessment. *AMIA Annu Symp Proc*, v. 2009, p. 359–363, 2009. [PubMed Central:PMC2815449] [PubMed:20351880]. (Citado nas páginas 31, 33, 63 e 64.)

MA, H. et al. Bridging the semantic gap between image contents and tags. *IEEE Transactions on Multimedia*, v. 12, n. 5, p. 462–473, Aug 2010. ISSN 1520-9210. (Citado na página 42.)

MARQUET, G. et al. Grading glioma tumors using OWL-DL and NCI Thesaurus. *AMIA Annu Symp Proc*, p. 508–512, 2007. [PubMed Central:PM2655830] [PubMed:18693888]. (Citado nas páginas 31, 63, 64 e 67.)

MASSICANO, F. et al. An ontology for tnm clinical stage inference. In: *ONTOBRAS*. [S.l.: s.n.], 2015. (Citado nas páginas 63 e 67.)

MCCOWAN, I.; MOORE, D.; FRY, M. J. Classification of cancer stage from free-text histology reports. In: *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*. [S.l.: s.n.], 2006. p. 5153–5156. ISSN 1557-170X. (Citado na página 44.)

MERIEU, B.; YAMINA, T.; PATHOLOGY, A. Interpretation breast cancer imaging by using ontology. *no. 1*, p. 1–6, 2012. Disponível em: <<http://www.cyberjournals.com/Papers/Mar2012/06.pdf>>. Acesso em: 19 September. 2016. (Citado nas páginas 63, 65 e 68.)

MÖLLER, M.; REGEL, S.; SINTEK, M. Radsem: Semantic annotation and retrieval for medical images. In: _____. *The Semantic Web: Research and Applications: 6th European Semantic Web Conference, ESWC 2009 Heraklion, Crete, Greece, May 31–June 4, 2009 Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 21–35. ISBN 978-3-642-02121-3. Disponível em: <http://dx.doi.org/10.1007/978-3-642-02121-3_6>. (Citado na página 42.)

MÖLLER, M.; SONNTAG, D.; ERNST, P. A spatio-anatomical medical ontology and automatic plausibility checks. In: _____. *Knowledge Discovery, Knowledge Engineering and Knowledge Management: Second International Joint Conference, IC3K 2010, Valencia, Spain, October 25-28, 2010, Revised Selected Papers*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 41–55. ISBN 978-3-642-29764-9. Disponível em: <http://dx.doi.org/10.1007/978-3-642-29764-9_3>. (Citado na página 65.)

MYERS, J. D. The background of internist i and qmr. In: *Proceedings of ACM Conference on History of Medical Informatics*. New York, NY, USA: ACM, 1987. (HMI '87), p. 195–197. ISBN 0-89791-248-9. Disponível em: <<http://doi.acm.org/10.1145/41526.41543>>. (Citado na página 38.)

NAGARAJAN, M. Semantic annotations in web services. In: _____. *Semantic Web Services, Processes and Applications*. Boston, MA: Springer US, 2006. p. 35–61. ISBN 978-0-387-34685-4. Disponível em: <http://dx.doi.org/10.1007/978-0-387-34685-4_2>. (Citado na página 42.)

NAPEL, S. A. et al. Automated retrieval of ct images of liver lesions on the basis of image similarity: Method and preliminary results. *Radiology*, v. 256, n. 1, p. 243–252, 2010. PMID: 20505065. Disponível em: <<http://dx.doi.org/10.1148/radiol.10091694>>. (Citado na página 43.)

NISHINO, M. et al. Revised RECIST guideline version 1.1: What oncologists want to know and what radiologists need to know. *AJR Am J Roentgenol*, v. 195, n. 2, p. 281–289, Aug 2010. [DOI:10.2214/AJR.09.4110] [PubMed:20651182]. (Citado na página 64.)

OBERKAMPF, H. et al. *Interpreting Patient Data using Medical Background Knowledge*. 2012. 2–6 p. Disponível em: <<http://ceur-ws.org/Vol-897/session2-paper06.pdf>>. Acesso em: 19 September. 2016. (Citado nas páginas 63 e 66.)

ORLANDO, J. P. *Usando aplicações ricas para internet na criação de um ambiente para visualização e edição de regras SWRL*. Dissertação (Mestrado) — ICMC-USP, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2012. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-25072012-101810/>>. (Citado na página 52.)

PATEL-SCHNEIDER, P. F. Building the semantic web tower from rdf straw. In: *In Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI 2005)*. [S.l.: s.n.], 2005. (Citado na página 51.)

PEARSON, S.-A. et al. Do computerised clinical decision support systems for prescribing change practice? a systematic review of the literature (1990-2007). *BMC Health Services Research*, v. 9, n. 1, p. 1–14, 2009. ISSN 1472-6963. Disponível em: <<http://dx.doi.org/10.1186/1472-6963-9-154>>. (Citado na página 38.)

POLIDAIIS. Medical Imaging in Cancer Care : Charting The Progress. 2006. Disponível em: <<http://www.medicalimaging.org/publications-and-research/>>. Acesso em: 01 March. 2015. (Citado nas páginas 39 e 40.)

RACOCLEANU, D.; CAPRON, F. Towards semantic-driven high-content image analysis: An operational instantiation for mitosis detection in digital histopathology. *Computerized Medical Imaging and Graphics*, v. 42, p. 2 – 15, 2015. ISSN 0895-6111. Breakthrough Technologies In Digital Pathology. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0895611114001487>>. (Citado nas páginas 63, 66 e 67.)

ROSSE, C.; MEJINO JOSÉ L.V., J. The foundational model of anatomy ontology. In: BURGER, A.; DAVIDSON, D.; BALDOCK, R. (Ed.). *Anatomy Ontologies for Bioinformatics*. Springer London, 2008, (Computational Biology, v. 6). p. 59–117. ISBN 978-1-84628-884-5. Disponível em: <http://dx.doi.org/10.1007/978-1-84628-885-2_4>. (Citado na página 56.)

RUBIN, D. L. Finding the Meaning in Images: Annotation and Image Markup. 2011. Disponível em: <<https://web.stanford.edu/group/rubinlab/pubs/18.4.rubin.pdf>>. Acesso em: 19 September. 2016. (Citado na página 42.)

RUBIN, D. L. et al. iPad: Semantic annotation and markup of radiological images. *AMIA Annu Symp Proc*, p. 626–630, 2008. (Citado nas páginas 32, 42, 57 e 70.)

RUBIN, D. L. et al. Automated Tracking of Quantitative Assessments of Tumor Burden in Clinical Trials. *Translational Oncology*, Neoplasia Press, Inc., v. 7, n. 1, p. 23–35, fev. 2014. ISSN 19365233. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1936523314800041>>. (Citado nas páginas 32, 33, 40, 55, 57 e 66.)

SEIFERT, S. et al. *Hierarchical parsing and semantic navigation of full body CT data*. 2009. 725902-725902-8 p. Disponível em: <<http://dx.doi.org/10.1117/12.812214>>. (Citado nas páginas 42 e 66.)

SERIQUE, K. J. A. *Anotação de imagens radiológicas usando a web semântica para colaboração científica e clínica*. Dissertação (Mestrado) — Universidade de São Paulo, São Carlos, 2012. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-10092012-155249/>>. (Citado nas páginas 32, 37, 55 e 56.)

SMITH, B. et al. The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech*, Nature Publishing Group, v. 25, n. 11, p. 1251–1255, nov 2007. ISSN 1087-0156. Disponível em: <<http://dx.doi.org/10.1038/nbt1346>>. (Citado na página 51.)

SMITH, D. A.; MANN, G. S. Bootstrapping toponym classifiers. In: *Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References - Volume 1*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2003. (HLT-NAACL-GEOREF '03), p. 45–49. Disponível em: <<http://dx.doi.org/10.3115/1119394.1119401>>. (Citado na página 51.)

STIVAROS, S. M. et al. Decision support systems for clinical radiological practice – towards the next generation. *Br J Radiol*, v. 83, n. 995, p. 904–914, Nov 2010. [PubMed Central:[PMC3473729](https://pubmed.ncbi.nlm.nih.gov/PMC3473729/)] [DOI:[10.1259/bjr/33620087](https://doi.org/10.1259/bjr/33620087)] [PubMed:[20965900](https://pubmed.ncbi.nlm.nih.gov/20965900/)]. (Citado na página 38.)

TUTAC, A. E.; CRETU, V. I.; RACOCEANU, D. Spatial representation and reasoning in breast cancer grading ontology. In: *Computational Cybernetics and Technical Informatics (ICCC-CONTI), 2010 International Joint Conference on*. [S.l.: s.n.], 2010. p. 89–94. (Citado nas páginas 63, 65 e 68.)

TUTAC, A. E. et al. Knowledge-guided semantic indexing of breast cancer histopathology images. *BioMedical Engineering and Informatics, International Conference on*, IEEE Computer Society, Los Alamitos, CA, USA, v. 2, p. 107–112, 2008. (Citado nas páginas 63, 64, 65, 66 e 67.)

WENNERBERG, P.; SCHULZ, K.; BUITELAAR, P. Ontology modularization to improve semantic medical image annotation. *Journal of Biomedical Informatics*, v. 44, n. 1, p. 155 – 162, 2011. ISSN 1532-0464. Ontologies for Clinical and Translational Research. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1532046410001796>>. (Citado na página 31.)

ZIEGLER, S. I. Positron emission tomography: Principles, technology, and recent developments. *Nuclear Physics A*, v. 752, p. 679 – 687, 2005. ISSN 0375-9474. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0375947405002265>>. (Citado nas páginas 40 e 41.)

ZILLNER, S. Towards the ontology-based classification of lymphoma patients using semantic image annotations. In: *SWAT4LS'09*. [S.l.: s.n.], 2009. p. –1–1. (Citado nas páginas 63, 65 e 67.)

ZILLNER, S. Reasoning-Based Patient Classification for Enhanced Medical Image Annotation. p. 243–257, 2010. (Citado na página [32](#).)

ZILLNER, S.; BRETSCHNEIDER, C. Grammar-Based Lexicon Extension for Aligning German Radiology Text and Images. n. September, p. 105–112, 2013. (Citado nas páginas [32](#) e [40](#).)

ZILLNER, S.; SONNTAG, D. Image metadata reasoning for improved clinical decision support. *Network Modeling Analysis in Health Informatics and Bioinformatics*, v. 1, n. 1-2, p. 37–46, mar. 2012. ISSN 2192-6662. Disponível em: <http://link.springer.com/10.1007/s13721-012-0003-9>. (Citado nas páginas [63](#), [65](#) e [68](#).)