

---

Classificação automática de tumores cancerosos  
usando anotações em imagens e ontologias

***Edson Francisco Luque Mamani***

---



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Edson Francisco Luque Mamani**

## Classificação automática de tumores cancerosos usando anotações em imagens e ontologias

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *EXEMPLAR DE DEFESA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Dilvan de Abreu Moreira

**USP – São Carlos**  
**Novembro de 2016**



Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados fornecidos pelo(a) autor(a)

L965c      Luque M., Edson Francisco  
            Classificação automática de tumores cancerosos  
            usando anotações em imagens e ontologias / Edson  
            Francisco Luque M.; orientador Dilvan de Abreu  
            Moreira. -- São Carlos, 2016.  
            107 p.

            Dissertação (Mestrado - Programa de Pós-Graduação  
            em Ciências de Computação e Matemática  
            Computacional) -- Instituto de Ciências Matemáticas  
            e de Computação, Universidade de São Paulo, 2016.

            1. OWL. 2. SWRL. 3. ESTÁGIO DE CÂNCER. 4.  
            BIOINFORMÁTICA. 5. SISTEMAS BASEADOS EM  
            CONHECIMENTO. I. Moreira, Dilvan de Abreu , orient.  
            II. Título.



**Edson Francisco Luque Mamani**

**Automatic classification of cancer tumors using image  
annotations and ontologies**

Master dissertation submitted to the Instituto de  
Ciências Matemáticas e de Computação – ICMC-  
USP, in partial fulfillment of the requirements for the  
degree of the Master Program in Computer Science  
and Computational Mathematics. *EXAMINATION  
BOARD PRESENTATION COPY*

Concentration Area: Computer Science and  
Computational Mathematics

Advisor: Prof. Dr. Dilvan de Abreu Moreira

**USP – São Carlos  
November 2016**





*A meus pais, José e Dina por sua capacidade de acreditar em mim*



# AGRADECIMENTOS

---

Queria deixar um agradecimento especial ao meu orientador Prof. Dr. Dilvan de Abreu Moreira, pelo apoio, especialmente por sua paciência, sabedoria, confiança e por toda a ajuda dedicada ao meu trabalho, serei eternamente grato por tudo. Também deixar um agradecimento especial ao Prof. Dr. Daniel L. Rubin pelo apoio no desenvolvimento deste trabalho e oportunidade de participar nos projetos desenvolvidos na Stanford University, ainda agradeço a todos demais docentes do ICMC que contribuíram para minha formação. Agradeço aos meus colegas, do ICMC e do Intermídia, especialmente a Silvio, Flor, Kleberson e João, que me ajudaram várias vezes no desenvolvimento deste mestrado. Agradecer às pessoas que me deram a sua ajuda nesta fase, e, finalmente, ao equipe do CAASO.



*“El éxito es fácil de obtener. Lo difícil es merecerlo.  
–Albert Camus”*



# Resumo

Informação sobre o estágio de câncer num paciente é fundamental quando médicos avaliam o progresso de seu 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, propensa a erros e gerar incertezas. Com o intuito de amenizar esses problemas, este trabalho tem como objetivo auxiliar radiologistas fornecendo uma segunda opinião na avaliação do estágio de câncer de um paciente. Para isso, tecnologias da Web Semântica, como ontologias e *reasoning*, foram usadas para classificar automaticamente estágios de câncer. Essa classificação usou anotações semânticas feitas por radiologistas, usando a ferramenta ePAD, e armazenadas no formato AIM. Um protótipo de classificador, baseado no padrão TNM, foi criado. Ele transforma anotações AIM em indivíduos da ontologia AIM4-O e, usando axiomas e regras (escritos na linguagem OWL-SWRL) representando o padrão TNM, ele automaticamente calcula o estágio de câncer de fígado de pacientes. A ontologia AIM4-O foi desenvolvida, como parte desse trabalho, para representar anotações AIM 4 em OWL. Esse classificador TNM foi avaliado, usando-se dados reais de relatórios radiológico de pacientes do NCI's Genomic Data Commons (GDC, em termos de precisão e revocação, com resultados respectivos de 85,7% e 81,0% (quando comparado aos valores reais de estágio de câncer dos relatórios). Todo o processo foi validado com radiologistas do Radiology Dept. of the Stanford University.

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





# Abstract

Information about cancer staging in a patient is crucial when clinicians assess treatment progress. Determining cancer stage is a process that takes into account the description, location, characteristics and possible metastasis of cancerous tumors in a patient. It should follow classification standards, such as the TNM Classification of Malignant Tumors. However, in clinical practice, the implementation of this process can be tedious, error-prone and create uncertainty. In order to alleviate these problems, we intend to assist radiologists by providing a second opinion in the evaluation of cancer stage in patients. For doing this, Semantic Web technologies, such as ontologies and reasoning, were used to automatically classify cancer stages. This classification used semantic annotations made by radiologists, using the ePAD tool, and stored using standard AIM format. A TNM based classifier prototype was created. It transforms AIM annotations in AIM4-O ontology individuals and, using axioms and rules (written in OWL-SWRL languages) representing the TNM standard, it automatically calculates patient's liver cancer stage. The AIM4-O ontology was developed, as part of this work, to represent AIM 4 annotations in OWL. This TNM classifier was evaluated, using real patient's radiology reports, from the NCI's Genomic Data Commons (GDC), in term of precision and recall, with 85.7% and 81.0% respective results (when compared to the actual cancer stages from the reports). The whole process was validated with radiologists from the Radiology Dept. of the Stanford University.

**Palavras Chave:** *OWL, SWRL, cancer staging, ePAD, TNM.*



# Lista de ilustrações

---

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. . . . .	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. . . . .	40
Figura 3 – A imagem CT do fígado anotada com termos semânticos. . . . .	42
Figura 4 – Um extrato da ontologia RadLex, que é uma terminologia controlada para radiologia. . . . .	43
Figura 5 – A) Uma imagem axial CT, contrastada, apresenta múltiplos tumores HCC, identificados e anotados utilizando a ferramenta ePAD; B) O diagrama do <i>stage</i> IIIA mostra múltiplos HCCs com pelo menos um deles $> 5$ cm. . . . .	45
Figura 6 – A) Uma imagem axial CT, contrastada, apresenta um tumor HCC, identificado e anotado utilizando a ferramenta ePAD; B) O diagrama do <i>stage</i> II mostra um tumor HCC solitário com invasão vascular. . . .	46
Figura 7 – Arquitetura em camadas da Web Semântica, . . . . .	50
Figura 8 – Mecanismos de <i>Reasoners</i> , baseados em lógica de descrição. . . . .	54
Figura 9 – Anotações AIM de uma imagem que descrevem localizações anatômicas, medidas e tipos de lesões . . . . .	57
Figura 10 – Ontologia de Bulu e Rubin (2015), baseada no modelo AIM 3.0. . . . .	68
Figura 11 – Este diagrama representa as entidades como ovais e retângulos. As propriedades são mostradas nas setas (em vermelho). . . . .	70
Figura 12 – Neste diagrama, descrevemos 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 <b>basic</b> e agora elas têm um sufixo "Entity", o que significa que essas classes são uma instância da classe abstrata principal " <b>Entity</b> ". . . . .	71

Figura 13 – Ilustração gráfica da Listagem 5.1. Classes de tipo <b>Extended</b> e seus indivíduos são mostrados. Ovais e retângulos representam entidades e as propriedades são mostradas em setas (cor vermelha). Entidades ovais, cercado por linhas pontilhadas, são classes abstratas de tipo <b>Entity</b> .	74
Figura 14 – OWL ObjectProperties adicionadas à ontologia de Bulu e que fazem parte da nova ontologia <b>AIM4-O</b> . Pode-se ver que a OWL Object-Property <b>hasAnnotations</b> é uma propriedade funcional inversa da propriedade <b>hasPerson</b> .	75
Figura 15 – Instâncias de anotações (indivíduos) AIM4-O, que são mostrados no painel esquerdo, e os detalhes e propriedades de uma instância são mostrados na direita.	75
Figura 16 – O diagrama da Ontologia Geral, seus componentes TNM (regras e axiomas) e importações necessárias a fim de se ter os mecanismos para classificar imagens de câncer de fígado.	76
Figura 17 – Relações entre as classes e outros conceitos da ontologia ONLIRA utilizados neste trabalho.	77
Figura 18 – Módulo da ontologia Radlex e classes que representam grupos de órgãos, tais como <b>adjacentOrganGroup</b> e <b>noadjacentOrganGroup</b> .	78
Figura 19 – Hierarquia de subclasses da ontologia estabelecidas usando-se Radlex.	82
Figura 20 – Image CT do fígado sendo anotada usando o ePAD e a template TNM (mostrada no apinel à direita da imagem).	91
Figura 21 – Matriz de escala de cores para a Matriz de Confusão.	93
Figura 22 – Resumo dos Histogramas para cada estágio após da classificação TNM dos metadados dos 51 pacientes	94

# Lista de tabelas

---

Tabela 1 – Os critérios de T, N e M para câncer de fígado. . . . .	45
Tabela 2 – Definições para os diferentes estágios de tumores. . . . .	46
Tabela 3 – Comparativo entre os trabalhos relacionados. . . . .	61
Tabela 4 – Prefijos e Namespaces usados na ontologia <b>AIM4-O</b> . . . . .	68
Tabela 5 – The American Joint Committee on Cancer/International Union against cancer TNM classification system and formal representation. . . . .	83
Tabela 6 – The American Joint Committee on Cancer/International Union against cancer TNM classification system and formal representation. . . . .	84
Tabela 7 – The American Joint Committee on Cancer/International Union against cancer TNM classification system and formal representation. . . . .	85
Tabela 8 – Queries em Description Logic (DL query) e palavras chaves para as quatro queries usadas na avaliação quantitativa da ontologia AIM4-O. . . . .	89
Tabela 9 – Resultados mostrando os valores de R=recall e P=precision para as 4 queries. . . . .	89
Tabela 10 – Matriz de Confusão dos estágios de câncer calculados pelo classificador (predicted) versus os valores que os médicos colocaram nos relatórios (actual). . . . .	92



# Lista De Códigos-Fonte

---

5.1	Representação em sintaxe Manchester de um individuo 8ixmtfhbiqujoz.. da ontologia AIM4-O (indivíduo anônimo, sem URI). Ele representa uma anotação semântica em uma imagem . . . . .	72
-----	--	----





# Lista De Abreviaturas e Siglas

---

AAS	Ann Arbor Staging
ACR	American College of Radiology
AIM	Annotation and Image Markup Project
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
GDC	NCI Genomic Data Commons
NCI	National Cancer Institute
NCI	National Cancer Institute
NGS	Nottingham Grading System
OWL	Web Ontology Language
PET	Positron Emission Tomography
RadLex	Unified Language of Radiology Terms
RDF	Resource Description Framework
RSNA	Radiological Society of North America
SWRL	Semantic Web Rule Language
TCIA	The Cancer Imaging Archive
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 . . . . .	19
Lista de tabelas . . . . .	21
Sumário . . . . .	27
<b>1</b>	<b>INTRODUÇÃO . . . . . 31</b>
1.1	Contextualização . . . . . 32
1.2	Motivação . . . . . 33
1.3	Objetivo . . . . . 33
1.4	Resultados Principais . . . . . 34
1.5	Organização . . . . . 35
<b>2</b>	<b>INFORMÁTICA BIOMÉDICA . . . . . 37</b>
2.1	Informática Biomédica para Sistemas de Suporte à Decisão . . . . . 38
2.2	Tecnologias para Imagens Médicas . . . . . 39
2.2.1	<i>Tomografia Computadorizada . . . . . 39</i>
2.2.2	<i>Tomografia por Emissão de Pósitrons . . . . . 39</i>
2.3	Anotação de Imagens . . . . . 41
2.3.1	<i>Anotação Semântica . . . . . 41</i>
2.3.2	<i>Anotação de Imagens Médicas com Vocabulários Estruturados . . . . . 42</i>
2.4	Medical Background Knowledge . . . . . 43
2.4.1	<i>Estágio do Câncer (Staging System) . . . . . 44</i>
2.4.2	<i>O Padrão TNM . . . . . 44</i>
2.5	Considerações Finais . . . . . 47
<b>3</b>	<b>WEB SEMÂNTICA . . . . . 49</b>
3.1	Web Semântica . . . . . 49
3.2	Ontologias . . . . . 51
3.3	Web Ontology Language – OWL . . . . . 51
3.4	Semantic Web Rule Language - SWRL . . . . . 52
3.5	Reasoners . . . . . 53
3.6	Bio-Ontologias . . . . . 54
3.6.1	<i>Foundational Model of Anatomy – FMA . . . . . 55</i>

3.6.2	<i>RadLex</i> . . . . .	55
3.7	Annotation Imaging Markup (AIM) . . . . .	56
3.8	AIM e ePad . . . . .	57
3.9	Considerações Finais . . . . .	58
4	TRABALHOS RELACIONADOS . . . . .	59
4.1	Planeamento e Execução do Mapeamento Sistemático . . . . .	59
4.2	Apresentação e Análise dos Resultados Obtidos no Mapeamento . . . . .	60
4.3	Considerações Finais . . . . .	66
5	METODOLOGIA . . . . .	67
5.1	Representação ontológica do modelo de informação AIM . . . . .	67
5.1.1	<i>Namespaces</i> . . . . .	68
5.1.2	<i>Basic Terms</i> . . . . .	69
5.1.3	<i>Expanded Terms</i> . . . . .	70
5.2	Gerar as condições para implementar o reasoning baseado em regras TNM. . . . .	73
5.2.1	<i>Ontologia geral</i> . . . . .	76
5.2.2	<i>ONLIRA (Ontologia de fígado para Radiologia)</i> . . . . .	77
5.2.3	<i>Terminologia Radlex</i> . . . . .	77
5.2.4	<i>Criando Classes para Representar Conceitos do Sistema TNM</i> . . . . .	78
5.3	Representação Formal do Staging de Câncer . . . . .	78
5.3.1	<i>Condição 1: O staging deve considerar a existência de tumor soli- tário ou múltiplo no mesmo site</i> . . . . .	79
5.3.2	<i>Condição 2: O staging deve considerar se os tumores são maiores ou menores que certos tamanhos "X" em cm</i> . . . . .	80
5.3.3	<i>Condição 3 : Staging deve considerar lesões em órgãos adjacentes</i> . . . . .	81
5.4	Considerações Finais . . . . .	82
6	EXPERIMENTOS . . . . .	87
6.1	Dados utilizados . . . . .	87
6.2	Avaliação quantitativa da ontologia AIM4-O . . . . .	88
6.3	Determinação do estágio de câncer (cancer staging) baseado em TNM . . . . .	90
6.4	Considerações Finais . . . . .	94
7	CONCLUSÃO . . . . .	97
7.1	Contribuições . . . . .	97
7.2	Dificuldades e Limitações . . . . .	98
7.3	Trabalhos Futuros . . . . .	99

---

<b>REFERÊNCIAS . . . . .</b>	<b>101</b>
------------------------------	------------



---

# 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 interpretáveis 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 tecnologias, como a web semântica, geram um interesse crescente para aplicação em biomedicina, já que elas podem tornar informações biomédicas explícitas e computáveis. A comunidade biomédica está em busca de ferramentas para ajudar 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, baseadas em imagens, têm sido desenvolvidas para área médica. 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 algumas 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. Mais especificamente quando 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 cancerosas, por meio de imagens, e grava medições detalhadas sobre essas lesões utilizando anotações em texto. 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 na 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 da 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 no 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 a 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. 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 obter uma maior taxa de precisão na interpretação.

Por outro lado, embora sistemas como o ePAD permitam a criação de anotações sobre imagens 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 cancerosas utilizando anotações sobre imagens médicas codificadas em 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 em meios padronizados (como AIM) de adição de informação e conhecimento sobre uma imagem. Sendo assim, a principal motivação deste trabalho é a possibilidade de desenvolver 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. Outra motivação é incorporar esses métodos aos requisitos atendidos por sistemas de anotação de imagens, como o ePAD.

## Objetivo

Este trabalho tem como objetivo gerar automaticamente, utilizando tecnologias da web semântica, o *Cancer Staging* (estágios do câncer) de lesões cancerosas a partir de anotações (metadados no padrão AIM) sobre essas lesões, presentes em imagens médicas. Essas anotações, feitas por radiologistas, descrevem observações visuais e a estrutura anatômica das lesões, utilizando a ontologia Radlex.

O *Cancer Staging* gerado fornece aos médicos uma segunda opinião sobre o estágio de câncer em que um paciente se encontra, ajudando a obter uma maior taxa de precisão na interpretação das lesões. O trabalho aqui descrito, se concentrou no *staging* de câncer de fígado, devido à disponibilidade de dados.

Para atingir o objetivo proposto, foi necessário se atingir 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 TNM de classificação (*staging*) de câncer. Esse conhecimento teve de ser codificado usando a Web Ontology Language (OWL) e regras codificadas na Semantic Web Rule Language (SWRL).
3. Encontrar e analisar dados com informações do *TNM staging*, como relatórios de diagnóstico clínico com as suas respectivas imagens por Computed Tomography (CT). Foram usados os repositórios *The NCI's Genomic Data Commons* (GDC) e *The Cancer Imaging Archive*.
4. Criar um protótipo do sistema de classificação automática de estágio de câncer para o ePAD.
5. Usar o protótipo para classificar dados, encontrados no item 3, e comparar os resultados da classificação com a classificação feita pelos médicos, com a ajuda de radiologistas da Stanford University School of Medicine.

## Resultados Principais

Os resultados principais deste trabalho de mestrado são:

- Um classificador automático para determinação do estágio de câncer (*cancer staging*) de um paciente. Ele é baseado no standard TNM e usa anotações em imagens médicas (no formato AIM). Essas anotações são convertidas para o formato OWL e usadas, em combinação com um conjunto de axiomas e regras, em um reasoner OWL DL e SWRL para calcular estágios de câncer. Esses estágios funcionam como uma segunda opinião médica para melhorar a taxa de precisão e padronização na determinação de estágio de câncer em pacientes. Esse classificador pode ser usado para detectar casos clínicos com estagio errôneo e, por conseguinte, potenciais erros de tratamento médico.
- Teste da precisão e revocação do classificador TNM desenvolvido. O classificador foi testado com relatório médicos reais abertos de 51 pacientes. Os estágios de câncer, calculados pelo classificador, foram comparados com os valores calculados pelos

médicos (nosso *golden standard*). A precisão e revocação obtidos pelo classificador foram 85.7% e 81,0%, respectivamente. Esses são valores muito bons, próximo a valores encontrados quando se compara opiniões de médicos de diferentes instituições.

- Uma nova ontologia, a AIM4-O, que fornece uma representação semântica do modelo de dados AIM 4.0. AIM4-O é uma ontologia leve que foi desenvolvida com a intenção de poder ser adotada em aplicações em outras disciplinas relacionadas ao processamento de metadados de imagens.
- Avaliação, centrada em tarefa, da ontologia AIM4-O. Relatórios sobre *image findings* foram descritos como indivíduos da ontologia e recuperados, usando-se a linguagem de consulta, baseada em lógica de descrição, DL query (busca semântica). O mesmo conjunto de relatórios foram usados para recuperação usando-se palavras-chave. Comparando-se os dois métodos, a busca semântica obteve resultados sensivelmente melhores que a busca por palavras-chave, com precisão 89.0% x 52.0% e revocação 98.0% x 48.0%.

Com os resultados acima alcançados, demostramos que as técnicas propostas podem obter uma boa qualidade na determinação de estágio de câncer em pacientes e, por tanto, podem aumentar a qualidade e uniformidade na interpretação de imagens por especialistas.

Este projeto de mestrado resultou na publicação de alguns artigos em eventos científicos que são listadas a seguir:

- E. F. Luque, D. L. Rubin and D. A. Moreira, "Automatic Classification of Cancer Tumors Using Image Annotations and Ontologies," 2015 IEEE 28th International Symposium on Computer-Based Medical Systems, Sao Carlos, 2015, pp. 368-369. doi: 10.1109/CBMS.2015.83.
- D. A. Moreira, C. Hage, E. F. Luque, D. Willrett and D. L. Rubin, "3D Markup of Radiological Images in ePAD, a Web-Based Image Annotation Tool," 2015 IEEE 28th International Symposium on Computer-Based Medical Systems, Sao Carlos, 2015, pp. 97-102. doi: 10.1109/CBMS.2015.46

## Organização

Este trabalho está estruturado da seguinte forma:

Capítulo 2: Apresenta tópicos sobre Informática Biomédica e aborda Tecnologias da Informação (TI), utilizadas nessa área, que são pertinentes para infra-estrutura necessária para este trabalho.

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

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 criação da ontologia AIM4-O e a codificação de conhecimento do critério TNM, usando as linguagens OWL e SWRL, para permitir o *staging* dos pacientes com base nas anotações em seus exames.

Capítulo 6: Mostra e discute os resultados obtidos.

Capítulo 7: Conclui esta dissertação de mestrado, comenta as limitações encontradas e propõe trabalhos futuros.

---

# 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 ao estudo do contexto da biomedicina, surgindo 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 de imagens, 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 integradas em aplicações médicas avançadas, tais como aquelas para o apoio à decisão clínica. Para facilitar o processamento dessas anotações por máquinas, é necessário adicionar informações semânticas que sejam baseadas em fontes de conhecimento. Além disso, é preciso que essas informações sejam codificadas de formas compreensíveis para máquinas, utilizando-se ontologias para facilitar 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 pa-

cientes com câncer, a partir de anotações em imagens, usando *reasoning*, este capítulo abordará os principais conceitos referentes a sistemas de suporte à decisão assistidos por computador. Também serão abordados Informática Radiológica, Imagens Médicas e 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

A partir dos anos 70, apareceram os primeiros sistemas de suporte à decisão para a área médica. 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 para informática biomédica [Belle, Kon e Najarian \(2013\)](#). Mais tarde, houve toda uma evolução significativa com grande reconhecimento de seu sucesso na melhoria de desempenho desses sistemas, o que é descrito por [Pearson et al. \(2009\)](#). Essa evolução permitiu o nascimento de mais subgêneros específicos de métodos de informática biomédica e suas implementações na forma de sistemas de diagnóstico assistidos 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 que viram sistemas de apoio à decisão auxiliada por computador 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 métodos como redes neurais e sistemas híbridos através da combinação de informações heterogêneas. O uso de ferramentas de diagnóstico auxiliado por computador tem o potencial de reduzir a variabilidade entre as opiniões de especialistas, bem como, melhorar a precisão diagnóstica para a interpretação e classificação de câncer ([JIANG et al., 2001](#)).

No caso da radiologia, o processamento de imagens baseado em computador é uma área de pesquisa ativa. O processamento de imagens, combinado com a visualização e aprendizagem de máquina para a tomada de decisões, tem proporcionado uma mais-valia para aplicações clínicas. Mais recentemente, com o surgimento de múltiplas tecnologias de imagens médicas, como CT, raios-X, ressonância magnética (MRI) e ressonância magnética funcional (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 tecnologias de 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 ([POLIDAIS, 2006](#)).

Antes de analisar como é que a semântica pode ser adicionada ao 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 a Tomografia Computadorizada e a Tomografia por Emissão de Pósitrons.

### 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 detecção da 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, Figura 1.

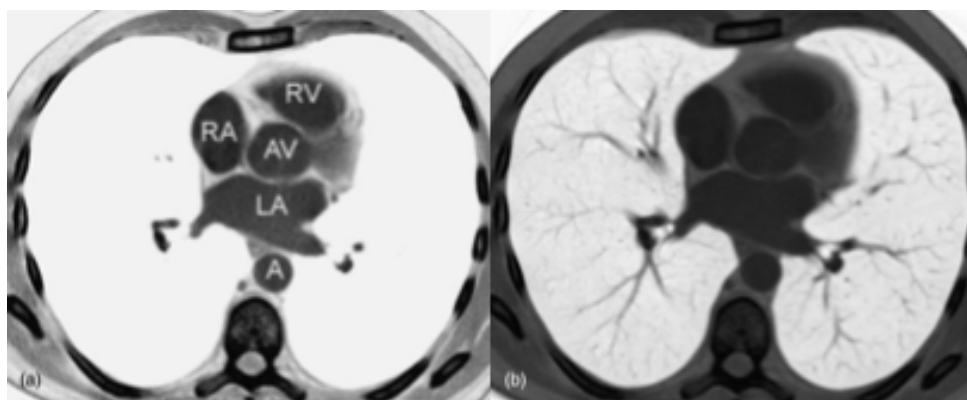
Tomografias estão entre as tecnologias de imagem mais comuns utilizados no diagnóstico, bem como no planejamento e monitoramento do tratamento de câncer; especialmente na detecção de câncer de fígado, pâncreas, pulmões e ossos([POLIDAIS, 2006](#)).

### Tomografia por Emissão de Pósitrons

Os exames de Tomografia por Emissão de Pósitrons, ou 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.

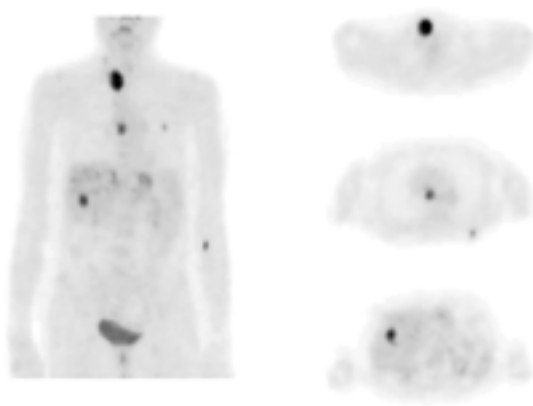
Nesses estudos oncológicos PET, o paciente recebe uma injeção do açúcar radioactivo que pode ajudar na localização de um tumor, porque as células cancerosas utilizam o açúcar (glicose) mais avidamente do que outros tecidos no corpo([ZIEGLER, 2005](#)).

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 \(2012\)](#).

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 \(2005\)](#).

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 CT e PET([RUBIN et al., 2014](#)).

Todas as modalidades de análise, acima mencionadas, fornecem uma visão detalhada sobre a anatomia humana, suas funções e 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 das 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 dados. Com base nisso, as aplicações médicas avançadas devem possuir descrições semânticas sobre dados clínicos, como imagens médicas.

## Anotação de Imagens

A necessidade de anotar imagens é reconhecida em uma ampla variedade de aplicações diferentes, abrangendo tanto o uso profissional quanto 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 assistida 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 desse tipo de sistema baseia-se na escolha dos termos que são usados para descrever o conteúdo das imagens. Essa escolha é altamente dependente da aplicação, necessidades e 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.

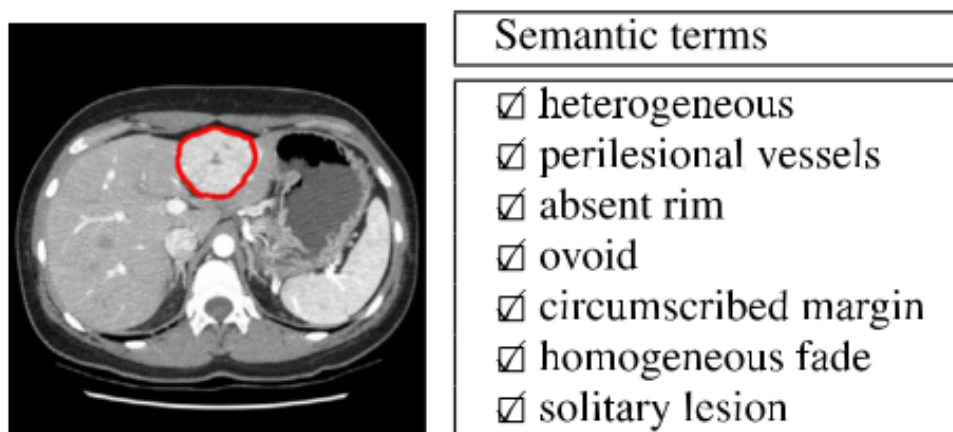
As anotações semânticas estão diretamente ligadas ao alto nível de compreensão e descrição do usuário sobre as características da imagem ([RUBIN, 2011](#)). Com base nessas considerações, incorporar características semânticas em sistemas de decisão sobre diagnóstico pode ser uma tentativa promissora 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

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



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

modernos (DEPEURSINGE et al., 2014). É por essa razão que várias abordagens para enfrentar o desafio da anotação semântica de imagens médicas já foram estudadas. 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 ao 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 contribuições importantes para melhorar o acesso às informações em imagens médicas, especificando a "semântica" dos achados médicos 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, 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.

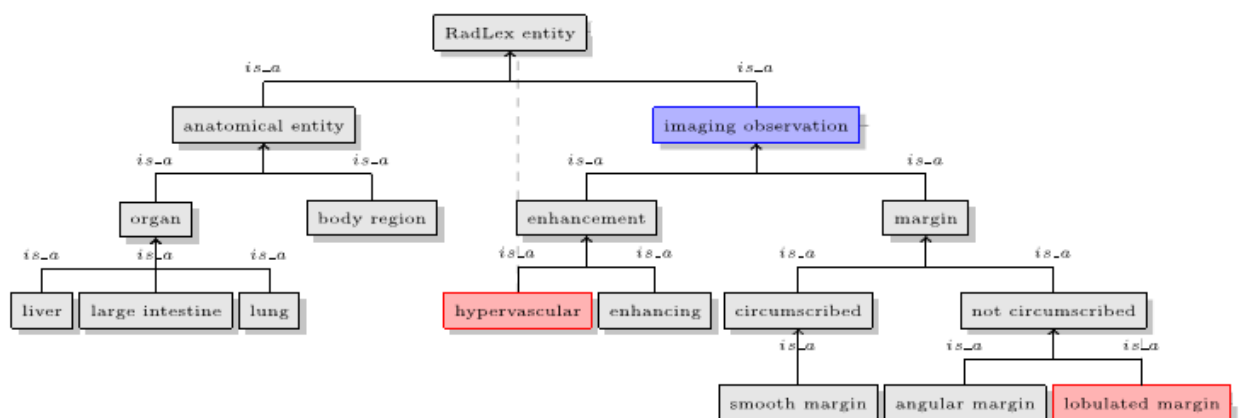
## Medical Background Knowledge

Embora os vocabulários sejam os blocos de construção básicos para a inferência usando técnicas da Web Semântica, não há uma divisão clara entre o que é referido como "vocabulários controlados" e "ontologias". Considerando-se 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 expressos 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 usadas para organizar conceitos biomédicos de uma forma abrangente, por exemplo: *Medical Subject Headings (MeSH)*, *International Classification of Diseases (ICD)*, *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 para 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 câncer, como o TNM, em ontologias e regras é 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 dele ter 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). Para câncer de fígado, esses conceitos são resumidos na Tabela 1.

A segunda etapa consiste na determinação do estagio 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.

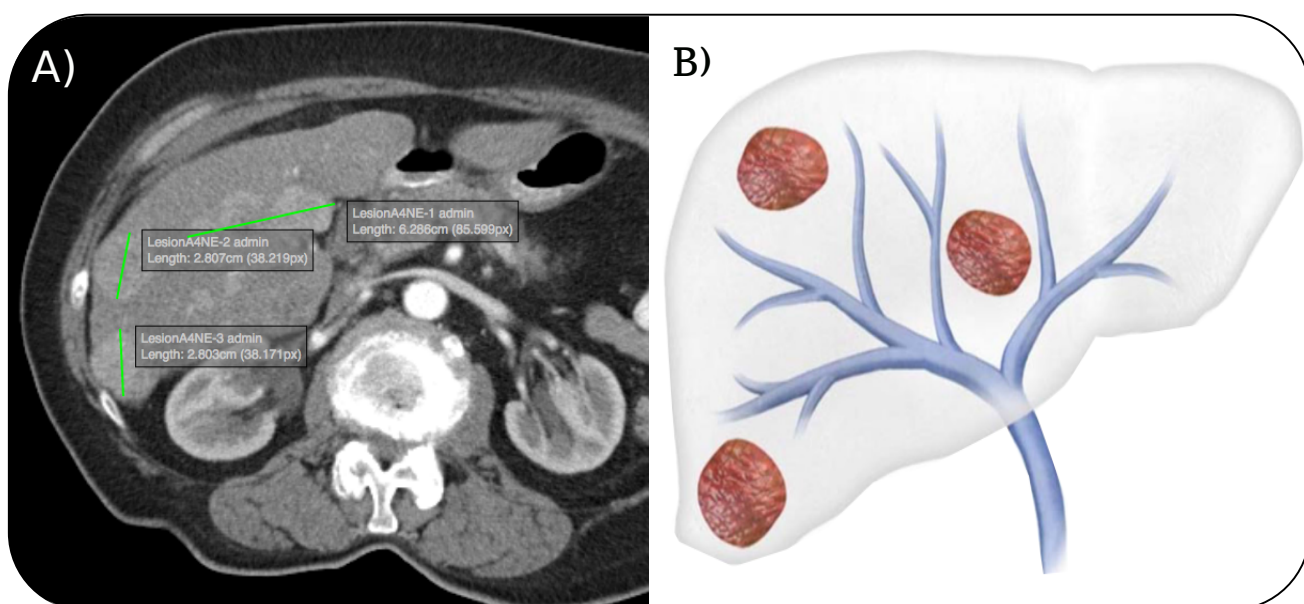
A Figura 5 mostra múltiplos tumores de fígado (linhas verdes) com pelo menos um deles > 5 cm. Não houve envolvimento linfonodal regional ou metástase. Este paciente seria classificado no estágio TNM **IIIA** (T3a, N0, M0).

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

T: Primary Tumor	Tis	Primary tumor cannot be assessed
	T0	No evidence of primary tumor
	T1	Solitary tumor without vascular invasion
	T2	Solitary tumor with vascular invasion or multiple tumors, none > 5 cm
	T3a	Multiple tumors > 5 cm
	T3b	Single tumor or multiple tumors of any size involving a major branch of the portal or hepatic vein
	T4	Tumor(s) with direct invasion of adjacent organs other than gallbladder or with visceral peritoneum
N: Regional Lymph Nodes	NX	Regional lymph nodes cannot be assessed
	N0	The cancer has not spread to lymph nodes
	N1	The cancer has metastasized only to lymph nodes on the same side as the cancerous liver.
M: Distant Metastasis	M0	The cancer has not spread to distant sites.
	M1	The cancer has spread to distant sites such as lymph nodes farther than those mentioned in N stages, and other organs or tissues such as the lung and colon

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

Figura 5 – A) Uma imagem axial CT, contrastada, apresenta múltiplos tumores HCC, identificados e anotados utilizando a ferramenta ePAD; B) O diagrama do *stage* IIIA mostra múltiplos HCCs com pelo menos um deles > 5 cm.



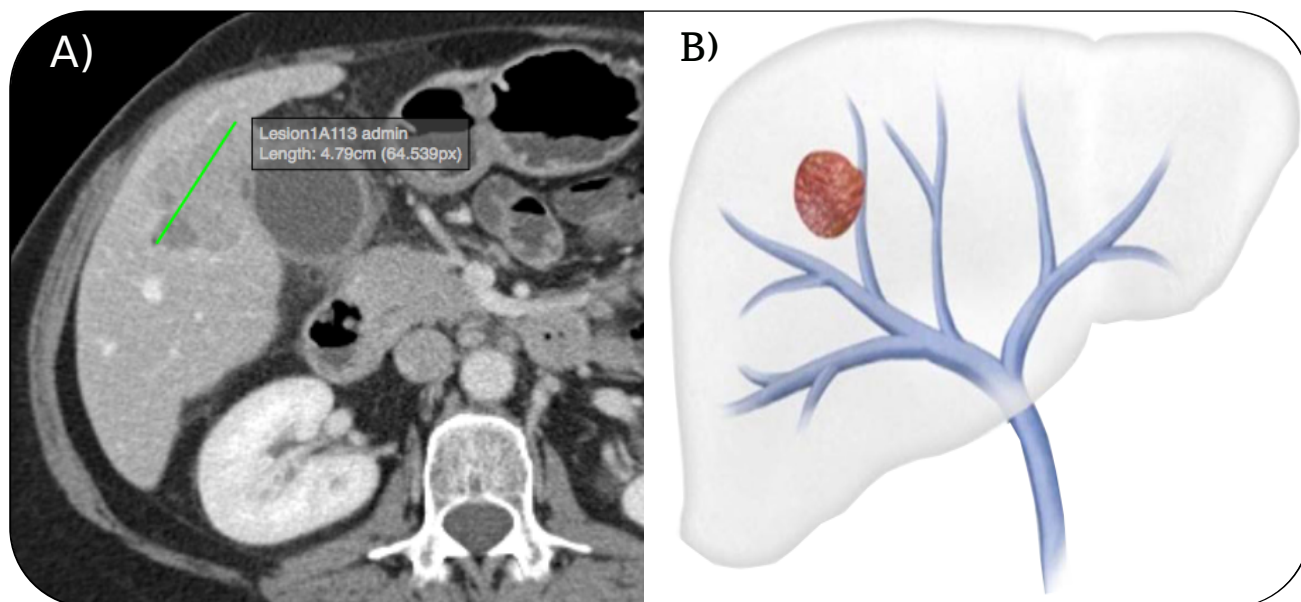
Fonte: Adaptada de [Faria et al. \(2014\)](#).

Tabela 2 – Definições para os diferentes estágios de tumores.

Stage 0	(Tis, N0, M0)
Stage I	(T1, N0, M0)
Stage I	(T2, N0, M0)
Stage IIIA	(T3a, N0, M0)
Stage IIIB	(T3b, N0, M0)
StageIIIC	(T4, N0, M0)
Stage IVA	(Any T, any N, M1)

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

Figura 6 – A) Uma imagem axial CT, contrastada, apresenta um tumor HCC, identificado e anotado utilizando a ferramenta ePAD; B) O diagrama do *stage* II mostra um tumor HCC solitário com invasão vascular.



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

A figura 6 mostra um único tumor de fígado (linha verde) com micro invasão vascular, sem envolvimento de linfonodos regionais ou metástases. Esse paciente seria classificado no estágio TNM **II** (T2, N0, M0).

Segundo [Faria et al. \(2014\)](#), o sistema de *staging* TNM tornou-se um dos sistemas

de *staging* patológico mais amplamente utilizados para carcinomas Hepato celular e foi validado para prever resultados pós-cirúrgicos em pacientes. As taxas de sobrevivência por 5 anos de pacientes com câncer de fígado, com base no sistema TNM, foram no stage I, 55%; stage II, 37%; e stage III, 16%. Portanto, podemos ver que existe uma relação entre o estágio clínico e as taxas de sobrevivência.

## Considerações Finais

Recentemente pode ser visto que houve uma escalada da complexidade na geração dos dados médicos pelo fato dos avanços em tecnologias relacionadas com aquisição de imagens. Novas tecnologias, como dispositivos médicos e sistemas de medição, geram grandes volumes de imagens como também outros dados por paciente, e tornam difícil para os médicos analisar e fornecer diagnósticos ou prognósticos em tempo hábil.

Este capítulo forneceu uma base teórica para a 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 de câncer.





---

## 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 um 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".

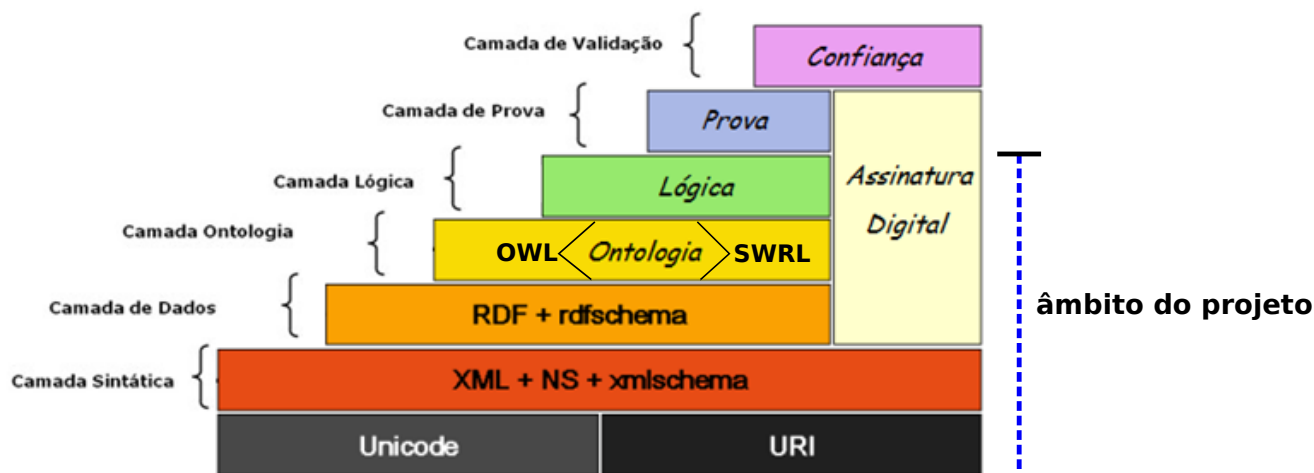
Os dados de entrada para realizar o processo de *staging* de câncer, neste trabalho, são basicamente terminologias que descrevem imagens, usando o formato XML. Tendo em conta isso, é vital abordar como associar e processar a semântica destes termos usando, neste caso, as tecnologias da Web Semântica. Tecnologias como ontologias, linguagens OWL-SWRL e reasoners são essenciais para realizar tarefas complexas de inferência exigidas pelo *staging*.

Conceitos relacionados a Web Semântica e ontologias foram mencionados no capítulo anterior. Neste capítulo, esses conceitos serão apresentados e discutidos, incluindo uma introdução à Web Semântica (Seção 3.1), ontologias (Seção 3.2), a Web Ontology Language - OWL (Seção 3.3) e a Semantic Web Rule Language - SWRL (seção 3.4). Na Seção 3.6, serão apresentadas as principais bio-ontologias e terminologias envolvidas nas pesquisas deste trabalho. Finalmente, na Seção 3.7, serão discutidas a linguagem de dados AIM e sua interação com ferramentas de anotação, como o ePAD.

### 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

Figura 7 – Arquitetura em camadas da Web Semântica,



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

conceitual sobre a Web, Berners-Lee propôs a arquitetura da Web Semântica, conhecida como “bolo de noiva”, que está ilustrada na Figura 7. A idéia central dessa arquitetura é de que cada camada vai gradativamente trazendo uma nova contribuição com 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 dessa arquitetura de camadas 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 sintaxes 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 um fator muito importante, ela permite a modelagem do *reasoning* para responder a questões complexas automaticamente,. Por essa razão, o trabalho desenvolvido engloba até a camada lógica da arquitetura da Web Semântica (Figura 7). Nas seções seguintes iremos detalhar as tecnologias da Web Semântica que serão utilizadas neste trabalho: Ontologias, linguagens de construção: OWL, XML , linguagens de regras: SWRL, Reasoners e 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 \(2005\)](#) existem linguagens para representação de ontologias onde usa-se lógica de predicados ou 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 contêm 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).

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 a criação da linguagem SWRL.

## Semantic Web Rule Language - SWRL

A linguagem de regras de Web Semântica (SWRL) emergiu como um primeiro passo para aumentar a interoperabilidade de sistemas baseados em regras a partir da perspectiva da Web Semântica<sup>1</sup>. Dado que a interoperabilidade é uma das metas principais da Web Semântica e que as regras são uma parte fundamental dessas metas. Essa linguagem 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*<sup>2</sup>.

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.

<sup>1</sup> <http://www.w3.org/Submission/SWRL/>

<sup>2</sup> <http://www.ruleml.org/>

A especificação da SWRL<sup>3</sup> não impõe restrições sobre como o raciocínio, com as regras, deve ser realizado. 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 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 possuem uma lesão hepática como também pertencentes à classe de pessoas com câncer. SWRL também permite o uso de *built-ins* que tornam a 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 de inferência, os *reasoners* utilizam um motor de inferência associado com um conjunto de regras descritas em linguagens como OWL e RDF(S).

Os *reasoners* contêm dois mecanismos internos para processar bases de conhecimento (*knowledge bases*), esses mecanismos são as TBOX e ABOX. A Figura 8 apresenta esses mecanismos associados a uma base de conhecimento. 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 desse tipo pode se referir a um conceito que representa uma especialização de um outro conceito:

---

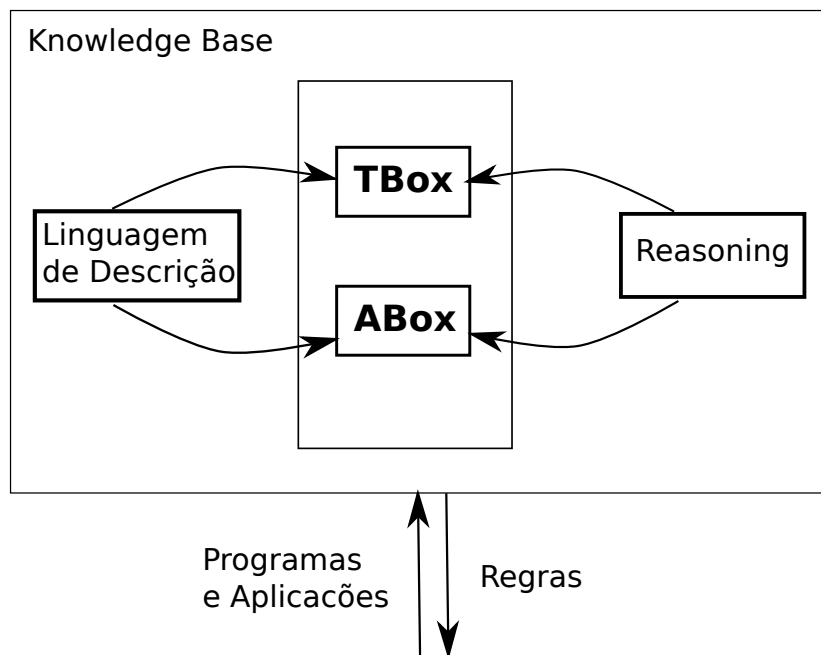

$$Homem \sqsubseteq Pessoa$$


---

Nesse exemplo, a classe (conceito) Homem é declarada como uma especialização de Pessoa.

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

Figura 8 – Mecanismos de *Reasoners*, baseados em lógica de descrição.



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

**ABox:** Compreende afirmações em objetos individuais (*instant assertions*). A afirmação típica na Abox é a que diz que um indivíduo é uma instância de um certo conceito, como no exemplo a seguir onde é afirmado que joaquin é um indivíduo da classe Homem:

---

$Homem \equiv joaquin$

---

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

## Bio-Ontologias

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 essa 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, algumas dessas ferramentas de engenharia ontológica são *FMA* (ROSSE; MEJINO JOSÉ L.V., 2008) e *RadLex* (LANGLOTZ, 2006).

## Foundational Model of Anatomy – FMA

Segundo Serique (2012) A FMA nasceu a falta de uma representação generalizada da anatomia (o estudo das estruturas físicas dos organismos biológicos) e tentar correlacionar os termos anatômicos entre si, devido a heterogeneidade das representações anatômicas. A FMA vem sendo desenvolvida para oferecer uma ontologia de referência, destinada ao uso por qualquer ferramenta computacional que necessite de informações anatômicas.

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 JOSÉ L.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*.

A FMA é de código aberto e está disponível para download no formato OWL<sup>4</sup> e em outros formatos<sup>5</sup>.

## RadLex

O RadLex é a mais importante bio-ontologia no 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

<sup>4</sup> (<http://www.bioontology.org/wiki/index.php/FMAInOwl>)

<sup>5</sup> (<http://sig.biostr.washington.edu/projects/fma/release/index.html>)

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 (LANGLOTZ, 2006), 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 e atender às necessidade dos radiologistas.

O RadLex está disponível para download no *National Center for Biomedical Computing's Biportal site*<sup>6</sup>. Ferramentas, como ePAD, utilizam o Radlex como a terminologia para descrever imagens em forma de anotações e a infraestrutura AIM para compartilhar essas anotações.

## Annotation Imaging Markup (AIM)

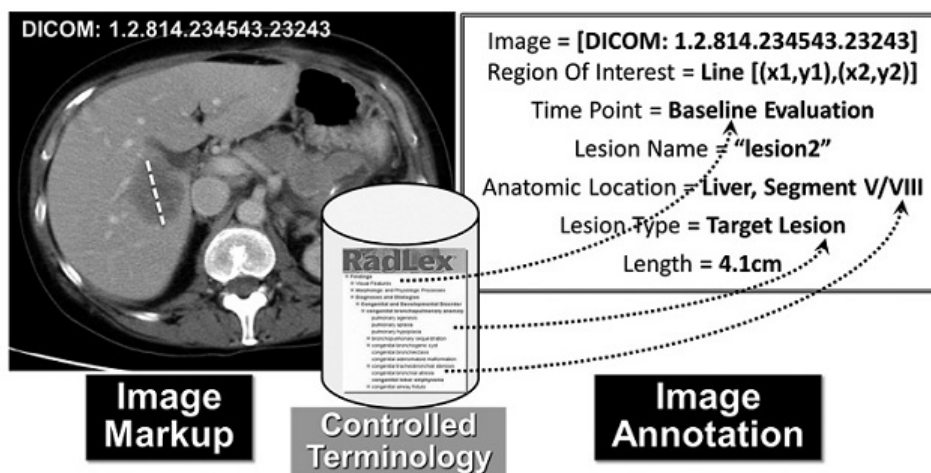
As modernas imagens médicas contêm uma vasta quantidade de informação no formato padrão DICOM, que é um conjunto de normas para uniformizar a comunicação entre equipamentos que geram essas imagens. No entanto, essas informações são relacionadas apenas a metadados sobre as imagens. As informações mais 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.

<sup>6</sup> <http://biportal.bioontology.org/ontologies/40885>



Figura 9 – Anotações AIM de uma imagem que descrevem localizações anatômicas, medidas e tipos de lesões



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

O modelo AIM de informação, ou AIM Schema, é descrito em [Rubin et al. \(2008\)](#), [Rubin et al. \(2014\)](#) usando um diagrama de classes 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.

## 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 9, podemos ver os dados de uma anotação (observação) AIM que a ferramenta ePAD armazena normalmente.

Para garantir que as informações mínimas necessárias para uma anotação sejam recolhidas, o ePAD usa um mecanismo chamado "*AIM templates*". *AIM templates* especificam um formulário de coleta de dados eletrônico, incluindo elementos obrigatórios, tipos de valores válidos e sua cardinalidade. Cada *template* é usada para garantir que o mínimo exigido de informações seja coletado para um determinado tipo de anotação. *Templates*, como, por exemplo, a RECIST<sup>7</sup> (*Response Evaluation Criteria In Solid Tumors*), definem quando pacientes de câncer melhoram, ficam na mesma ou pioram. Antes deste trabalho, não existia uma representação ontológica da infraestrutura AIM que permita inferir e classificar marcações armazenadas por ferramentas, como o ePAD. Além disso, também não existia um *template* 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 em 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. Nessa área, 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.

---

<sup>7</sup> <http://www.irrecist.com/>

---

## 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 Library<sup>1</sup>, IEEE Xplore Digital Library<sup>2</sup>, SciELO (Scientific Electronic Library Online)<sup>3</sup>, DBLP (Computer Sci-

---

<sup>1</sup> <http://dl.acm.org/>

<sup>2</sup> <http://ieeexplore.ieee.org/>

<sup>3</sup> <http://www.scielo.org/>

ence Bibliography)<sup>4</sup>, BDBComp (Biblioteca Digital Brasileira de Computação)<sup>5</sup>, Google Acadêmico (Scholars), Springer<sup>6</sup>, PubMed(US National Library of Medicine National Institutes of Health)<sup>7</sup>.

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 será apresentada 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).

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

---

<sup>4</sup> <http://dblp.uni-trier.de/>

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

<sup>6</sup> <http://www.springer.com/gp/>

<sup>7</sup> <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 <sup>8</sup> – 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 <sup>9</sup> – Grading (OWL-DL, SWRL reasoning)	Yes
Tutac, Cretu e Racoc- eanu (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 <sup>10</sup> – Staging (OWL-DL reasoning)	Yes
Oberkampf 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
Racocceanu 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

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.

- [Marquet et al. \(2007\)](#) desenvolveram uma ontologia, baseada na ontologia NCI Thesaurus, para classificação dos tumores de glioma. Nesse 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. 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 \(2009\)](#) permita a inferência automatizada de informação semântica sobre lesões de câncer em imagens, as tarefas mais complexas, como a 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). A sua finalidade era definir regras que são utilizadas para a indexação semântica das imagens de pacientes com câncer da mama. Embora a proposta de [Tutac et al. \(2008\)](#) seja desenvolvida 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 ao *TNM*.
- Em [Zillner \(2009\)](#), uma ontologia OWL-DL foi projetada para realizar o *staging* automático de pacientes com linfoma, usando anotações semânticas. Essas anota-

ções são armazenadas numa ontologia pela ferramenta de anotação que faz parte de projeto *MEDICO-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; ERNST, 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 específico para o campo de histopatologia. Esses arquivos no WFML foram traduzidos para OWL. Por fim, a linguagem de consulta SPARQL é usado para acessar as informações disponíveis no sistema.
- Em [Oberkamp 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 Oberkamp 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.

- 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 do 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. \(2015\)](#) tentem generalizar o sistema de classificação TNM, representado-o numa ontologia, eles fazem 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 como os trabalho acima, as seguintes diferenças foram evidenciadas:

Trabalhos como [Dameron et al. \(2006\)](#), [Marquet et al. \(2007\)](#) executam o *reasonig* 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 e que os trabalhos solucionam evitando o reasoning baseado em indivíduos. No entanto, é possível realizar o *reasoning* baseado em indivíduos, suportado por estruturas de dados, que serão descritas em mais detalhe na seção seguinte.

Os trabalhos, como [Tutac et al. \(2008\)](#), [Massicano et al. \(2015\)](#) e [Zillner \(2009\)](#), 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, como no caso dos trabalhos de [Tutac et al. \(2008\)](#) e [Zillner \(2009\)](#). Nós acreditamos na abordagem de se 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 os trabalhos de [Meriem, Yamina e Pathology \(2012\)](#) e [Tutac, Cretu e Racocceanu \(2010\)](#) utilizem anotações semânticas em imagens e executem uma classificação baseada no NGS (utilizando OWL e SWRL), eles têm um fator limitante que é criar uma nova ontologia, dependendo da patologia que se 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, que é do [Zillner e Sonntag \(2012\)](#), ele não disponibiliza os dados utilizados, o que impede qualquer verificação dos resultados. Além disso, o sistema de *staging* de linfoma, que foi implementado, é relativamente mais simples do que um sistema de classificação (*staging*) mais geral, como o TNM. Por exemplo, pode-se dizer que [Zillner e Sonntag \(2012\)](#) não considera a dimensão

de uma lesão como um fator importante. No entanto, esse fato é muito importante em sistemas de *staging* de câncer mais complexos, como o TNM. Podemos verificar também que o processo de alinhamento de todas as ontologias geradas em [Zillner e Sonntag \(2012\)](#) não é descritos de forma explícita. Finalmente, acreditamos que nossa abordagem possa 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 produzidas num sistema de imagem qualquer (que gera arquivos AIM) são transformadas em OWL, em seguida essas anotações podem ser utilizadas para o reasoning necessário em tarefas de diferentes domínios.

## 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 em imagem no formato AIM (na linguagem XML), mas AIM 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 classificador TNM que usa *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:

- A transformação do modelo de informação AIM em uma representação ontológica que define um modelo de informação semanticamente equivalente (Seção 5.1).
- A criação das condições para implementar o reasoning baseado em regras TNM utilizando instâncias OWL (Seção 5.2,).
- A representação formal necessária para *staging* do câncer utilizando apenas a expressividade OWL + SWRL (Seção 5.3).

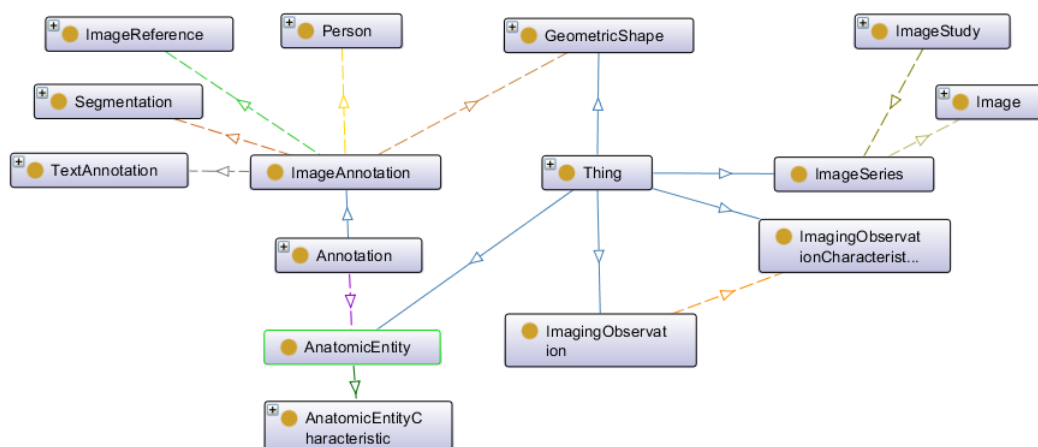
## Representação ontológica do modelo de informação AIM

A fim de executar a inferência e classificação, a partir de anotações e dos metadados de marcação de imagem (baseados no modelo AIM), precisamos de uma linguagem que tenha uma semântica formal. Usando essa semântica, as inferências sobre ontologias e seus indivíduos podem ser feitas. Nesse contexto, a linguagem OWL permite a construção de uma representação ontológica semanticamente equivalente a um modelo de informação, tal como, por exemplo, o modelo de informação usado em AIM. Com base no exposto, optamos neste trabalho por transformar o modelo de dados AIM em uma representação ontológica equivalente em OWL (versão 2).

Essa transformação foi realizada através da criação de classes e propriedades em OWL, resultando em algo compreensível pelo usuário e adequado para inferência. Desenvolvemos uma ontologia com base no trabalho de Hakan Bulu (BULU; RUBIN, 2015). Nesse trabalho, ele desenvolveu uma ontologia para representar 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, modificamos a ontologia de Bulu (Figura 10). Vale destacar que as versões 3.0 e 4.0 do AIM têm diferenças significativas no modo como representam dados.

Figura 10 – Ontologia de Bulu e Rubin (2015), baseada no modelo AIM 3.0.



Usando essa nova versão da ontologia (chamada AIM4-O), capturamos lesões e medidas derivadas durante os ensaios clínicos com base em imagens e, posteriormente, as convertimos em instâncias de anotação na ontologia. Essas instâncias são usadas durante a etapa de reasoning. A seguir, é descrita em detalhes a ontologia **AIM4-O**.

A ontologia **AIM4-O** foi desenvolvida como uma extensão da ontologia de Bulu 10 e define a codificação OWL2 do modelo de dados AIM 4.0. Seu objectivo é apoiar ferramentas de anotação de imagens especificando a semântica das observações feitas por especialistas e, desse modo, melhorar a precisão interpretativa dos dados. 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	<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://www.w3.org/1999/02/22-rdf-syntax-ns#</a>	The RDF namespace
xsd	<a href="http://www.w3.org/2000/10/XMLSchema#">http://www.w3.org/2000/10/XMLSchema#</a>	XML Schema Namespace
owl	<a href="http://www.w3.org/2002/07/owl#">http://www.w3.org/2002/07/owl#</a>	The OWL namespace
aim	<a href="http://www.w3.org/ns/aim4#">http://www.w3.org/ns/aim4#</a>	The AIM4-O namespace

Dependendo das necessidades e de acordo com a quantidade de detalhes que se deseja incluir numa anotação de imagem, o usuário do AIM4-O pode usar apenas partes

da ontologia. Por essa razão, os termos da ontologia **AIM4-O** (classes e propriedades) são agrupados, neste documento, em duas categorias, a fim de fornecer uma descrição incremental: **Basic Terms** e **Expanded Terms**.

## Basic Terms

As classes **Basic** fornecem a base para capturar anotações e marcações de informações relevantes para imagens. Através dessas classes, pode-se descrever explicitamente o tipo de informação que se quer coletar com o modelo. O usuário pode estender o modelo para atender suas necessidades, seguindo os critérios descritos na documentação do AIM<sup>1</sup>.

As classes tipo **Basic Term** foram definidas na ontologia de [Bulu e Rubin \(2015\)](#). A seguir temos uma descrição mais detalhada dessas classes:

1. **AIM:Calculation:** representa os resultados do cálculo de uma anotação AIM. Os resultados dos cálculos podem ser ou não diretamente associados 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 vistas em uma imagem.
4. **AIM:ImagingObservationCharacteristics:** a classe contém descritores da classe AIM: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 está 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 usada para anotar uma imagem ou imagens. Ela 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.

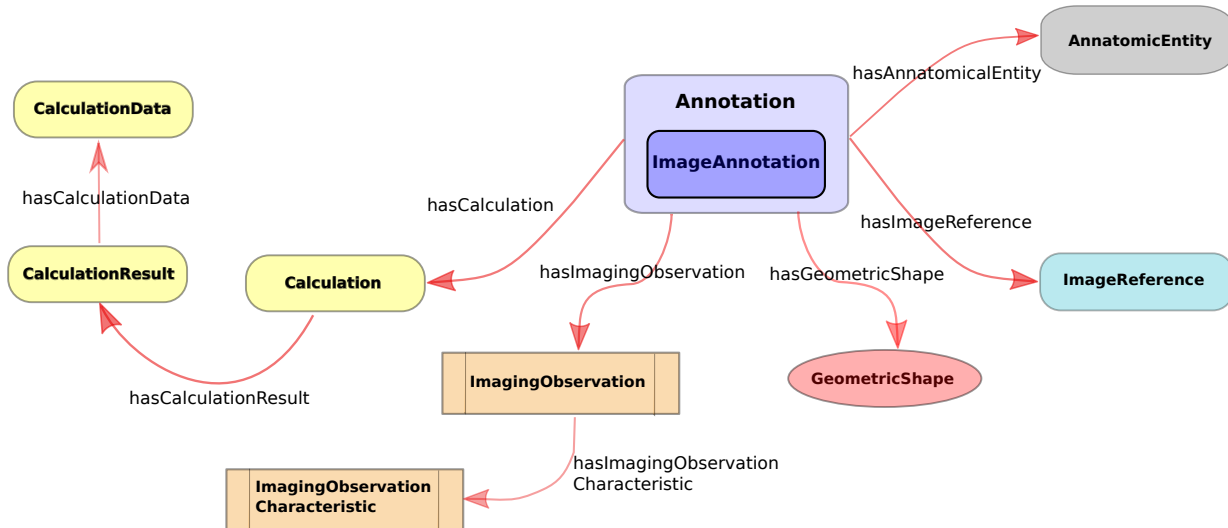
---

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

9. **AIM:CalculationData**: essa classe é usada para armazenar um valor de resultado.

As 9 classes se relacionam entre si utilizando as propriedades mostradas na Figura 11.

Figura 11 – Este diagrama representa as entidades como ovais e retângulos. As propriedades são mostradas nas setas (em vermelho).



## Expanded Terms

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. Em alguns casos, os nomes das classes (presentes nas versões 4.0 e *foundation model* do AIM) foram mudados, adicionando-se o sufixo "Entity". O que significa que essas classes são uma instância da classe abstrata principal **AIM:Entity**. As classes adicionais encontram-se ilustrados na figura 12:

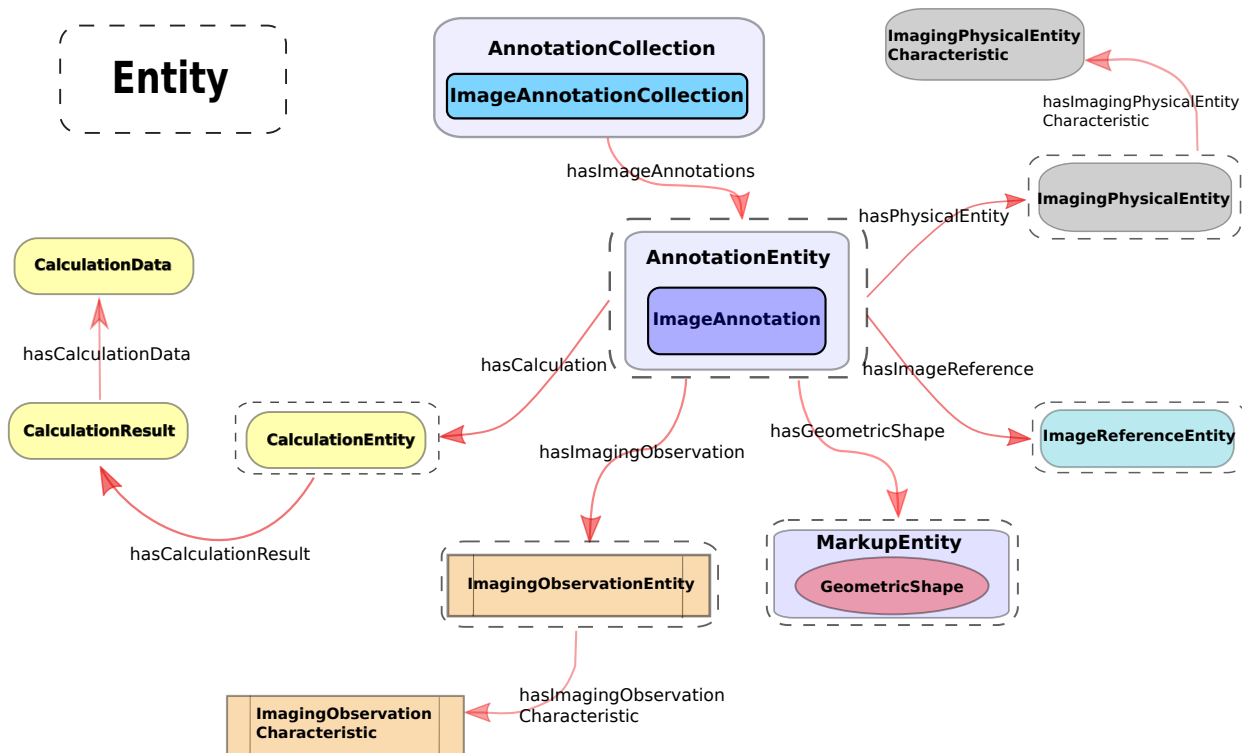
1. **AIM:Entity**: é uma classe abstrata que representa a existência de uma coisa, conceito, observação, cálculo, medição e desenho gráfico no AIM.
2. **AIM:AnnotationCollection**: é um conceito abstrato de um recipiente que reúne um tipo de anotação, como AIM:ImageAnnotation ou AIM:AnnotationOfAnnotation.
3. **AIM:ImageAnnotationCollection**: armazena instâncias de AIM:ImageAnnotation. Se associa com a classe Person que contém informação demográfica do paciente.
4. **AIM:ImagingPhysicalEntity<sup>2</sup>**: armazena uma localização anatômica como um termo codificado (por exemplo, RID2662 , fêmur, RadLex ) a partir de um vocabulário controlado reconhecido como RadLex®, SNOMED CT® ou Sistema Único de linguagem médica (UMLS).

<sup>2</sup> A única classe no modelo que teve seu nome totalmente alterado. Era **AnatomicEntity** na versão 3.0 do AIM.

5. **AIM:ImagingPhysicalCharacteristic**: descreve AIM:ImagingPhysicalEntity tais como termo codificado.
6. **AIM:MarkupEntity**: captura informações textuais e representação gráfica, como DICOM-SR e SCOORD3D, para coordenadas espaciais tri e bi-dimensionais.

As novas classes se relacionam entre si utilizando as propriedades mostradas na Figura 12.

Figura 12 – Neste diagrama, descrevemos 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 elas têm um sufixo "Entity", o que significa que essas classes são uma instância da classe abstrata principal “**Entity**”.



O exemplo na Listagem 5.1 ilustra os termos “**Expanded**” e mostra a codificação **OWL** (em sintaxe Manchester) de um indivíduo da ontologia **AIM4-O** representando uma anotação sobre uma imagem. Esse exemplo também mostra os novos recursos, de tipo **Expanded Term**, usados na criação da uma anotação.

A Listagem 5.1 mostra uma anotação e sua representação, na forma de indivíduo **AIM4-O**. Nela podemos ver que se 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 características mais salientes podemos citar o surgimento de dataProperties, como **typeCode**,

Listagem 5.1 – Representação em sintaxe Manchester de um individuo Sixmtfqbiqujoz.. da ontologia AIM4-O (indivíduo anônimo, sem URI). Ele 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: Sixmtfqbiqujozezep1r90t1n0po8wxyfy5svzs
4   Types: ImageAnnotationCollection
5   Facts:
6     aim:hasImageAnnotations :9gs43xqj1kyl13lmega0zhoenzvgeakkprft8fw8,.
7 Individual: 9gs43xqj1kyl13lmega0zhoenzvgeakkprft8fw8
8   Types: ImageAnnotation
9   Facts: aim:hasImageReference :sj19n9gf050ap9uzkc2h98ye4l3rxd71jac28g3g,
10     aim:hasMarkupEntity :jma0x979fa9y3k4cwypqjvtechwitqb3glvdzjyw,
11     aim:hasCalculationEntity :a4lpf1nnfcbvh5dljf6gfw6r6i3omiece6270nix,
12     aim:hasImagingObservation :hojggs5f543pzsffj8jbijgp11thm9qdjow4idd,
13     aim:hasPhysicalEntity :g08jnm9ow79tijgj1br3q8wd89nq69epjdnxxo30,
14     aim:dateTime "2014-10-05T22:15:43"^^xsd:dateTime,
15     aim:uniqueIdentifier "9gs43xqj1kyl13lmega0zhoenzvgeakkprft8fw8"^^
      xsd:string,
16     aim:comment "CT/THORAX 1.0 B45F/273"^^xsd:string,
17     aim:name "Liver2"^^xsd:string.
18 Individual: sj19n9gf050ap9uzkc2h98ye4l3rxd71jac28g3g
19   Types: DicomImageReferenceEntity
20   Facts: aim:type "DicomImageReferenceEntity"^^xsd:string,
21     aim:uniqueIdentifier "sj19n9gf050ap9uzkc2h98ye4l3rxd71jac28g3g"^^
      xsd:string.
22 Individual: jma0x979fa9y3k4cwypqjvtechwitqb3glvImageAnnotationCollectiondzjyw
23   Types: TwoDimensionMultiPoint
24   Facts: aim:type "TwoDimensionMultiPoint"^^xsd:string,
25     aim:uniqueIdentifier "jma0x979fa9y3k4cwypqjvtechwitqb3glvdzjyw"^^
      xsd:string
26     aim:imageReferenceUid "1.2.826.0.1.3680043"^^xsd:string,
27     aim:referencedFrameNumber 273.
28 Individual: a4lpf1nnfcbvh5dljf6gfw6r6i3omiece6270nix
29   Types: CalculationEntity
30   Facts: aim:hasCalculationResult :a4lpf1nnfnixExtendedCalculationResult,
31     aim:description "LineLength"^^xsd:string,
32     aim:typeCode "[{codeSystemName=private, codeSystem=
      LineLength,
33     aim:uniqueIdentifier "a4lpf1nnfcbvh5dljf6gfw6r6i3omiece6270nix"^^
      xsd:string.
34 Individual: a4lpf1nnfnixExtendedCalculationResult
35   Types: ExtendedCalculationResult
36   Facts: aim:dataType "[{codeSystemName=, codeSystem=}"^^xsd:string,
37     aim:type "ExtendedCalculationResult"^^xsd:string,
38     aim:unitOfMeasure "linear"^^xsd:string,
39     aim:values 1.7432112f.
40 Individual: hojggs5f543pzsffj8jbijgp11thm9qdjow4idd
41   Types: ImagingObservationEntity
42   Facts: aim:annotatorConfidence 0.0f,
43     aim:label "Lesion"^^xsd:string,
44     aim:typeCode "[{codeSystemName=RECIST-AMS, codeSystem=
      LesionBaseline,
45     Evaluation, code=S81, codeSystemVersion=}"^^xsd:string,
46     aim:uniqueIdentifier "hojggs5f543pzsffj8jbijgp11thm9qdjow4idd"^^
      xsd:string.
47 Individual: g08jnm9ow79tijgj1br3q8wd89nq69epjdnxxo30
48   Types: ImagingPhysicalEntity
49   Facts: aim:annotatorConfidence 0.0f,
50     aim:label "Location"^^xsd:string,
51     aim:typeCode "[{codeSystemName=RadLex, codeSystem=liver,
      code=RID58, codeSystemVersion=}"^^xsd:string,
52     aim:typeCode "[{codeSystemName=RadLex.3.10, codeSystem=
      Couinaudhepaticsegment7, code=RID67}"^^xsd:string,
53     aim:uniqueIdentifier "g08jnm9ow79tijgj1br3q8wd89nq69epjdnxxo30"^^
      xsd:string.

```



que é um atributo que pode capturar informações sobre terminologias controladas em um esquema de codificação simplificado (codeSystemName=**RadLex**, codeSystem=**liver**, code=**RID58**, codeSystemVersion=).

Uma ilustração gráfica da Listagem 5.1 mostra como as classes de tipo **Extended** referem-se umas às outras (Figura 13).

Neste trabalho, além de adicionar as novas classes, também foram adicionadas propriedades de objetos (OWL ObjectProperties) e de dados (OWL DataProperties). Foi criado um conjunto de novas relações, como *isImageStudyOf*; *isImageReferenceOf*; *hasAnnotations*, que foram adicionadas à ontologia de Bulu 10. O objectivo dessas relações é adicionar à **AIM4-O** uma maior expressividade em tarefas de classificação complexas, tal como os critérios de classificação TNM. Cada uma dessas relações é uma propriedade funcional inversa das propriedades: **hasImageStudy**; **hasImageReference** e **hasPerson** respectivamente, de modo que eles só pode ter um valor. Mais formalmente, se estabelecermos que *hasAnnotations* é uma *inverseFunctionalProperty* de *hasPerson* não podem haver duas pessoas diferentes com a mesma anotação.

A **AIM4-O** é uma ontologia leve que foi desenvolvida com a intenção de ser adotada em uma gama de aplicações de outras disciplinas relacionadas ao processamento de metadados de imagem. A **AIM4-O** está de acordo com o perfil **OWL-DL**. Ela pode ser usada diretamente 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, este documento usa um cenário de exemplos, Figura 13.

## Gerar as condições para implementar o reasoning baseado em regras TNM.

A segunda tarefa, na construção de um classificador TNM, foi definir um mecanismo para transformar 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. Esse 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 baseadas na representação UML do modelo AIM 4.0. Posteriormente, a OWL-API (a API java usada para criar, manipular e serializar ontologias) foi usada para criar instâncias OWL dessas classes Java, permitindo assim ao script processar fluxos de documentos XML e preencher uma base de conhecimento OWL com seu conteúdo, Figura 15.

Figura 13 – Ilustração gráfica da Listagem 5.1. Classes de tipo **Extended** e seus indivíduos são mostrados. Ovais e retângulos representam entidades e as propriedades são mostradas em setas (cor vermelha). Entidades ovais, cercado por linhas pontilhadas, são classes abstratas de tipo **Entity**.

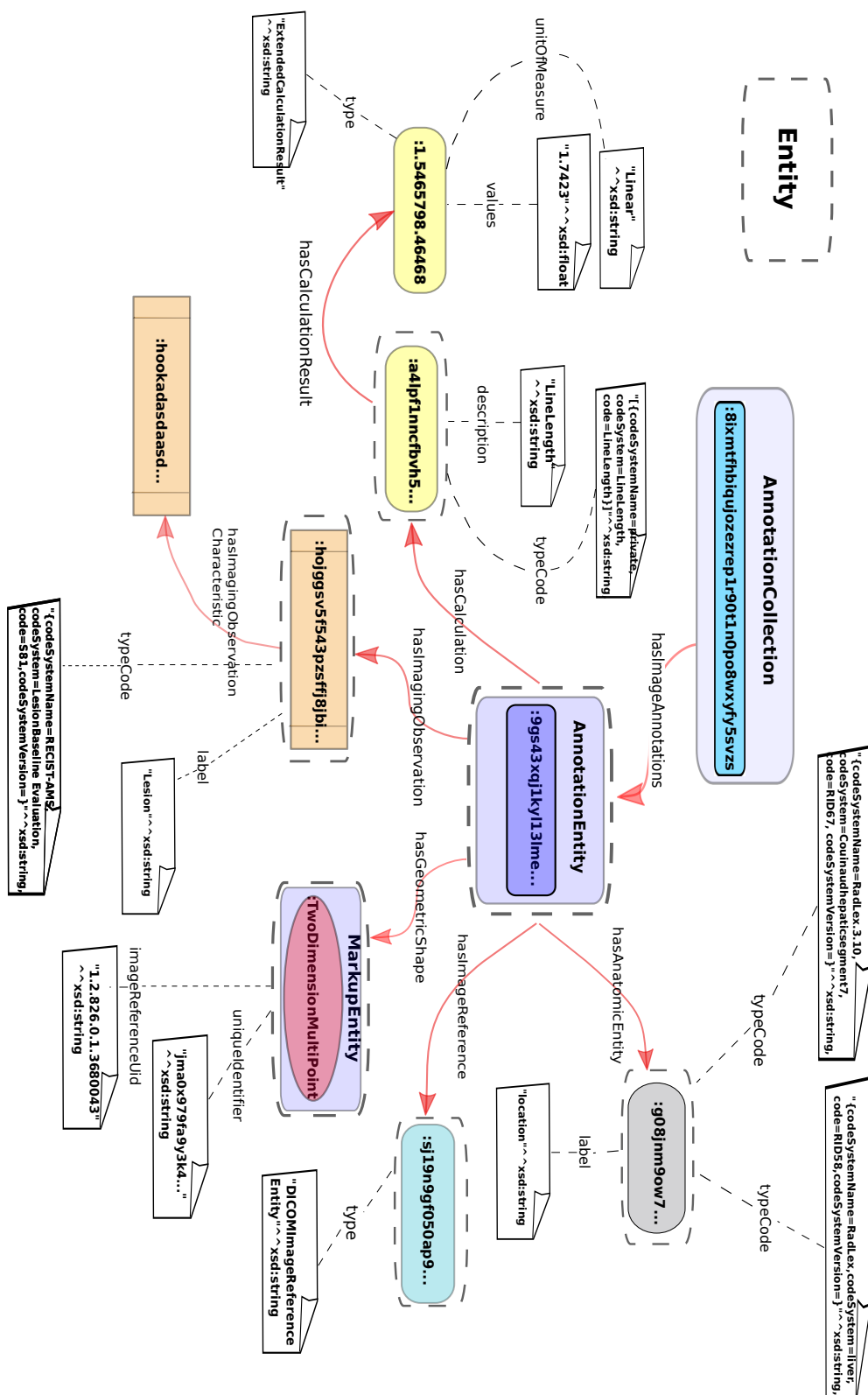


Figura 14 – OWL ObjectProperties adicionadas à ontologia de Bulu e que fazem parte da nova ontologia AIM4-O. Pode-se ver que a OWL ObjectProperty **hasAnnotations** é uma propriedade funcional inversa da propriedade **hasPerson**.

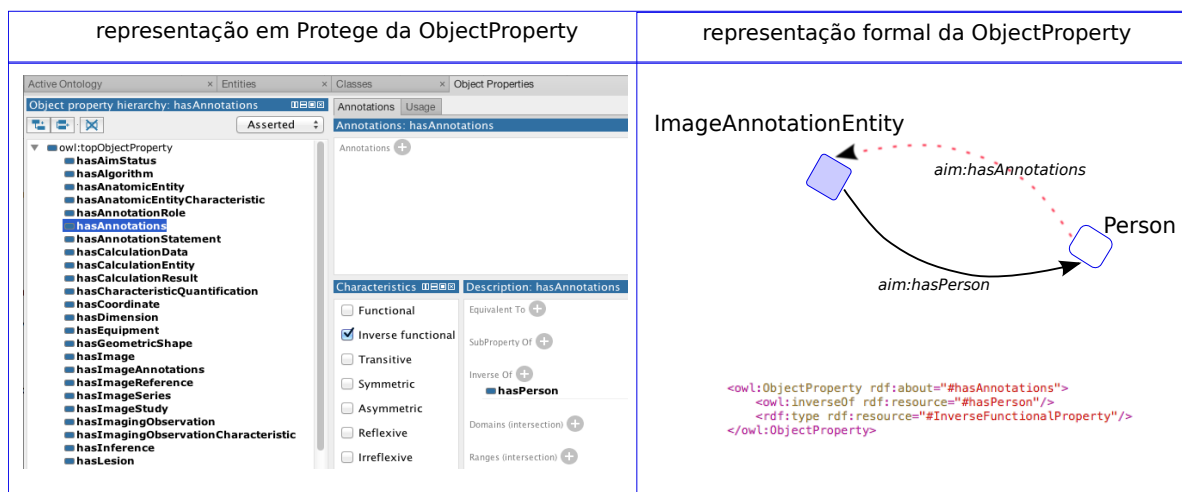
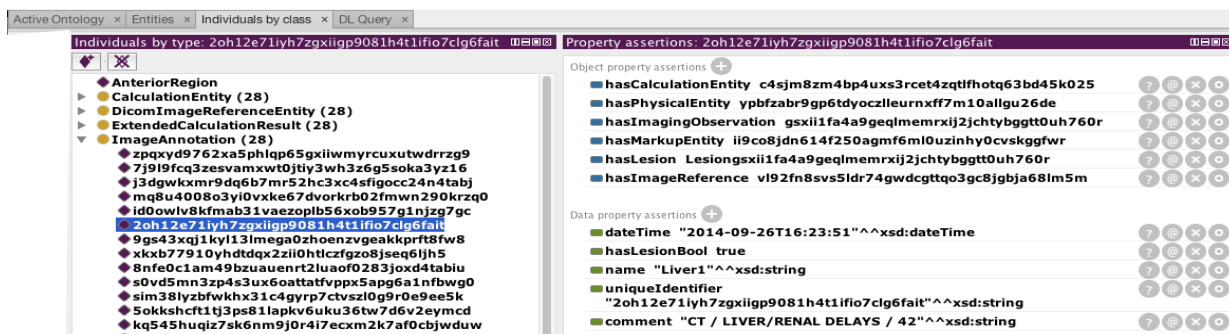


Figura 15 – Instâncias de anotações (indivíduos) AIM4-O, que são mostrados no painel esquerdo, e os detalhes e propriedades de uma instância são mostrados na direita.



O objetivo da representação OWL, Figura 15, é dar as condições para se desenvolver mecanismos de reasoning, baseados em regras, para trabalhar com anotações AIM. Para atingir o objetivo da classificação da lesões de câncer, usamos relações de subclasse, interseção, equivalência e disjunção entre classes, OWL objectProperties e algumas regras como mecanismos de raciocínio. Este processo será descrito na terceira tarefa.

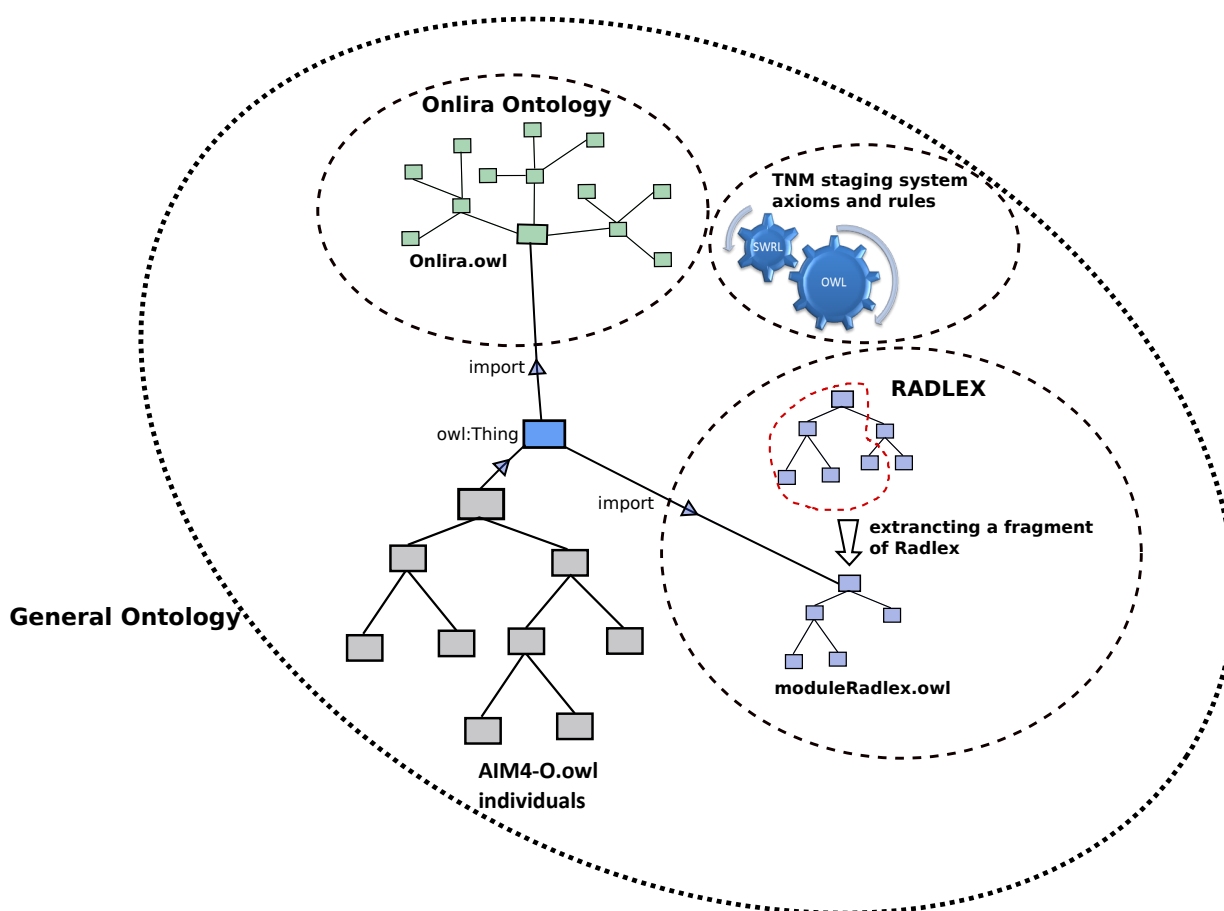
A fim de fazer automaticamente o staging de um tipo de câncer, precisamos ter o apoio de uma ontologia capaz de especificar um domínio particular. Neste caso, o órgão baseline afetado pelo câncer neste trabalho é o fígado. Além disso, considerou-se necessário a inclusão de uma representação OWL de parte do vocabulário Radlex, a fim de facilitar o trabalho de classificação e tratamento de indivíduos da ontologia AIM4-O (já que eles usam terminologias Radlex em sua estrutura). Finalmente, foi adicionado à nova ontologia um 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 dividida em 4 arquivos, como mostra a Figura 16. Eles contêm as seguintes informações:

1. A ontologia AIM4-O, com os indivíduos citados na seção acima.
2. Os conceitos gerais do TNM (regras SWRL e axiomas).
3. Onlira.owl: a topografia do fígado (com base na Ontologia Onlira (KOKCIYAN et al., 2014)).
4. ModuleRadlex.owl: Uma parte do vocabulário Radlex <sup>3</sup>.

Figura 16 – O diagrama da Ontologia Geral, seus componentes TNM (regras e axiomas) e importações necessárias a fim de se ter os mecanismos para classificar imagens de câncer de fígado.



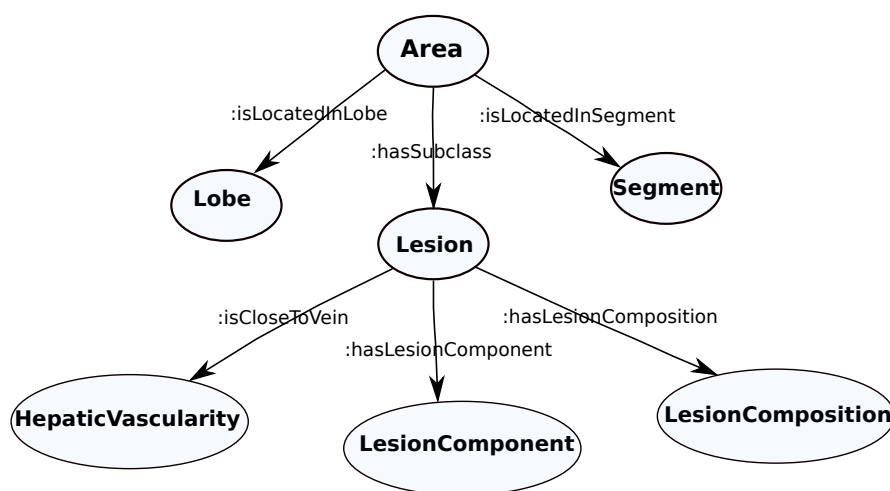
As seções a seguir fornecem uma descrição das ontologias utilizadas como parte da **Ontologia Geral**.

## ONLIRA (Ontologia de fígado para Radiologia)

A ontologia ONLIRA foi desenvolvida como parte do projeto CaReRa (KOKCIYAN et al., 2014). ONLIRA<sup>4</sup> e visa modelar observações de imagem no domínio do fígado com ênfase em propriedades e relações entre o fígado, as veias hepáticas e lesões hepáticas. O desenho de ONLIRA baseou-se em sessões de elicitación com radiologistas.

Partes dessa ontologia foram utilizadas para representar a topografia do fígado, descrita nas anotações AIM-XML geradas pelo ePAD.

Figura 17 – Relações entre as classes e outros conceitos da ontologia ONLIRA utilizados neste trabalho.



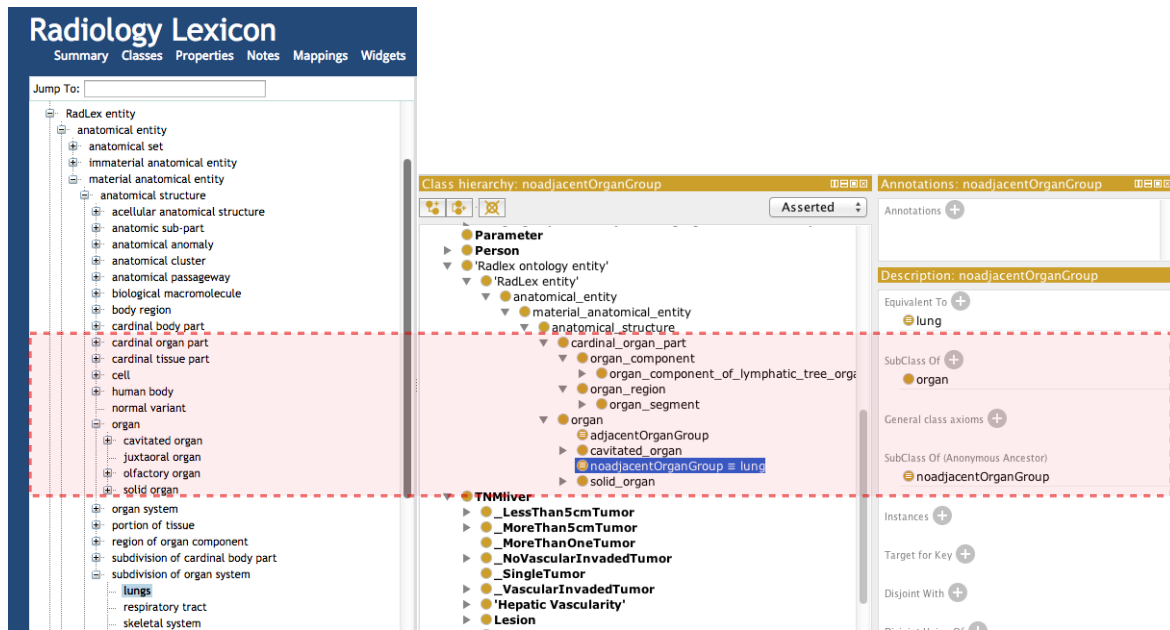
## Terminologia Radlex

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

Os critérios TNM referem-se a locais de entidades e estruturas anatômicas. É importante que a ontologia contenha, não apenas as estruturas anatômicas mencionados nos critérios de staging, mas também as partes ligadas direta e indiretamente a elas. Por exemplo, a fim de implementar os critérios N e M do TNM, adicionamos duas super classes: **adjacentOrganGroup**, que descreve o conjunto de órgãos adjacentes a um órgão principal (neste caso o fígado), e **noadjacentOrganGroup**, que descreve órgãos com base nos locais mais comuns de disseminação do tumor, conforme o trabalho de Faria et al. (2014). Como neste trabalho foi tomado, como caso de estudo, a classificação de câncer de fígado, incluímos pulmões e ossos como órgãos adjacentes, Figura 18.

<sup>4</sup> <https://bioportal.bioontology.org/ontologies/ONLIRA>

Figura 18 – Módulo da ontologia Radlex e classes que representam grupos de órgãos, tais como `adjacentOrganGroup` e `noadjacentOrganGroup`.



## Criando Classes para Representar 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 declaradas 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 tenha crescido em vasos sanguíneos (classe `T2_a`), e a segunda, representa um único tumor não maior do que 5 cm (classe `T2_b`).

## Representação Formal do *Staging* de Câncer

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

O staging de câncer é dividido em duas etapas principais. A primeira consiste em criar pontuações a partir da descrição do tumor (T); a propagação em gânglios linfáticos

(N) e possível metástase (M). A segunda 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 o sistema TNM exige usando-se apenas expressividade OWL + SWRL.

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

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

**Condição 2:** O staging deve considerar se os tumores são maiores ou menores que certos tamanhos "X" em cm.

**Condição 3 :** O staging deve considerar lesões em órgãos adjacentes.

### Condição 1: O staging 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 permitam inferir se um paciente tem tumores simples ou múltiplos explicitamente. No caso de tumores múltiplos, construímos a seguinte regra "**MoreThanOneTumor** " (em notação SWRL ) :

---


$$ImageStudy(?X) \wedge isImageStudyOf(?x, ?y) \wedge isImageStudyOf(?x, ?z) \wedge$$

$$DifferentFrom(?y, ?z) \rightarrow MoreThanOneTumor(?x)$$


---

Essa regra classifica um indivíduo como membro da classe **:MorethanOneTumor**, se um estudo de imagem "X" é referenciado por mais de uma anotação da imagem. Para classificar algo como **:MorethanOneTumor**, precisamos criar uma nova propriedade chamada **isImageStudyOf**. Ela é o inverso da OWL 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, 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 não se disse que algo existe, isso não significa que esse algo não existe". Por exemplo, pode-se descrever uma lesão de câncer em uma imagem, usando uma entidade de tipo **:ImagingObservation**, mas, a menos que se diga explicitamente que não há outras lesões, presume-se que podem haver outras que não foram mencionadas explicitamente.

Para resolver esse problema de mundo aberto, consideramos algumas alternativas como, por exemplo, reformular a ontologia AIM4-O para modelar **:hasImagingObservation** como uma classe e não uma relação. Mas, não achamos essa alternativa intuitiva. Em vez disso, decidimos indicar o número de lesões, de forma explícita, através da criação de um novo conceito, chamado **singleLesion**, como uma OWL ObjectProperty de entidades **:ImageStudy**. Esse conceito indica se o estudo descreve apenas uma lesão de câncer. Assumiu-se que o estudo descreve exatamente uma lesão câncer ("**singleLesion { true }**"), se ele é referenciado por uma única anotação na imagem. Mas, não é possível formular esse axioma usando apenas OWL. Em vez disso, essa informação é gerada como parte do processo de *parsing* das anotações de imagem no formato AIM para criar indivíduos da ontologia AIM4-O.

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: O staging deve considerar se os tumores são maiores ou menores que certos tamanhos "X" em cm

Esta condição é facilmente implementada, obtendo-se o valor da propriedade *values* pertencente à entidades do tipo **:CalculationResult**. Este tipo de 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 tomando "X=5" cm como tamanho para o TNM do fígado (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 este critério, o mais complicado para 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**<sup>17</sup>. A classe **:Lesion** lida com muitas características importantes de uma lesão, tais como a composição, densidade, tamanho e forma, mas infelizmente isso não é suficiente para satisfazer os critérios da classificação TNM. Por essa razão, foram adicionadas 3 propriedades a ela e, como resultado, definimos um novo conceito, a subclasse **:OutsideLesion**. As 3 novas propriedades são:

**hasLocation** (OWL ObjectProperty): indica a localização das lesões com base na taxonomia **Radlex**. A propriedade **hasLocation** relaciona as instâncias da classe **:Lesion**, da Ontologia Onlira, às instâncias da classe **:AnatomicalEntity**, do módulo Radlex (usado neste trabalho).

**isRegionalLymphNodeAffected** (OWL DataProperty): especifica se uma lesão está localizada em algum nódulo linfático. Esse conceito foi útil para critérios de classificação como N0 e N1.

**isAdjacentOrgan** (OWL ObjectProperty): Indica se a lesão, presente no local "X", está perto de órgão "Y". De acordo com o critério de classificação TNM do câncer de fígado, considerou-se como órgãos adjacentes ao fígado: o pâncreas, estômago, duodeno, fígado e cólon, Figura 19. Agrupamos esses conceitos (usando a taxonomia Radlex), criando duas novas classes **:AdjacentOrganGroup** e **:NoAdjacentOrganGroup**:

---


$$NoAdjacentOrganGroup \subseteq RadlexEntity.Organ$$

$$AdjacentOrganGroup \subseteq RadlexEntity.Organ$$

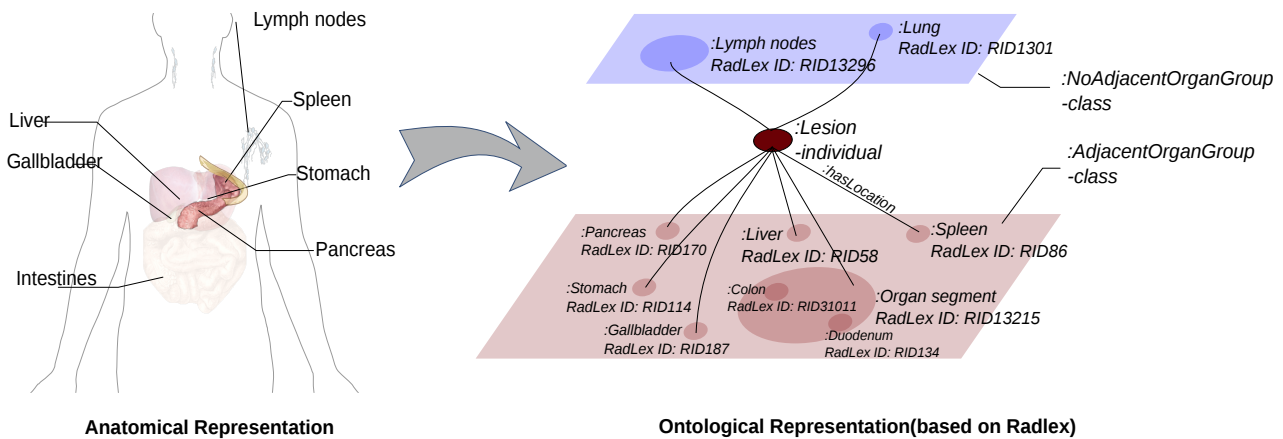

---

As classes **: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 câncer, neste trabalho, o órgão pai é o fígado. Finalmente, construímos a seguinte regra (em notação SWRL) para classificar se uma lesão fora do fígado (representada pelo conceito **:OutsideLesion**) está localizada em um órgão adjacente:

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

Neste ponto, uma vez que os critérios anteriores foram adequadamente codificados, usando **OWL** e **SWRL**, eles foram usados para construir os axiomas e regras a seguir. Eles foram capazes de classificar automaticamente lesões cancerosas, utilizando o sistema TNM (ver Tabelas 5, 6, 7). Notamos que a ordem como modelamos as coisas importava. Por exemplo, foi mais fácil definir as regras para os critérios de classificação **N1a**, **N0** e reutilizar suas definições para o **M0**, ao invés de começar com a definição de **M0**, através da manipulação de closures complexos.

Figura 19 – Hierarquia de subclasses da ontologia estabelecidas usando-se Radlex.



Com a utilização da ontologia AIM4-O, os conceitos puderam ser facilmente relacionados uns com os outros. Além disso, os requisitos de integridade e de cardinalidade puderam ser especificados e checados para fins de *reasoning*.

## Considerações Finais

Neste capítulo, foram descritas as etapas necessárias para alcançar a implementação de um sistema de classificação de estágios de câncer baseado em anotações AIM contidas em imagens médicas. Para se conseguir isso, foi necessário uma representação

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

Primary tumor (T)	
Medical representation for T1 criterion:	<b>Solitary tumor (any size) without vascular invasion</b>
Formal representation for T1 in OWL-DL	$\equiv \text{NoVascularInvadedTumor} \wedge$ $\ni \text{hasImageReference}.\text{ImageReferenceEntity} \wedge$ $\text{ImageReferenceEntity} \wedge \ni \text{hasImageStudy}.\text{ImageStudy} \wedge \text{SingleTumor}$
Medical representation for T2a criterion :	<b>Solitary tumor (any size) with vascular invasion</b>
Formal representation for T2a in OWL-DL	$\equiv \text{VascularInvadedTumor} \wedge$ $\ni \text{hasImageReference}.\text{ImageReferenceEntity} \wedge$ $\text{ImageReferenceEntity} \wedge \ni \text{hasImageStudy}.\text{ImageStudy} \wedge \text{SingleTumor}$
Medical representation for T2b criterion :	<b>multiple tumors none &lt; 5cm</b>
Formal representation for T2b in OWL-DL	$\equiv \text{LessThan5cmTumor} \wedge$ $\ni \text{hasImageReference}.\text{ImageReferenceEntity} \wedge$ $\text{ImageReferenceEntity} \wedge \ni \text{hasImageStudy}.\text{ImageStudy} \wedge \text{MoreThanOneTumor}$
Medical representation for T3a criterion :	<b>Multiple tumors, with at least one tumor &gt; 5cm</b>
Formal representation for T3b in OWL-DL	$\equiv \text{MoreThan5cmTumor} \wedge$ $\ni \text{hasImageReference}.\text{ImageReferenceEntity} \wedge$ $\text{ImageReferenceEntity} \wedge \ni \text{hasImageStudy}.\text{ImageStudy} \wedge \text{MoreThanOneTumor}$
Medical representation for T4a criterion :	<b>Tumors with direct invasion of adjacent organs other than the gallbladder</b>
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 – The American Joint Committee on Cancer/International Union against cancer TNM classification system and formal representation.

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

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

Distant metastasis (M)	
Medical representation for M0 criterion :	<b>No distant metastasis</b>
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 :	<b>Distant metastasis</b>
Formal representation for M1 in OWL-DL	$\equiv ImageAnnotation.hasLesion \wedge \ni Onlira.Lesion \wedge$ $Lesion.hasLocation \wedge \ni Organ.noadjacentOrganGroup$

formal híbrida baseada em axiomas OWL dos critérios de classificação TNM e regras SWRL. Além disso, esse 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 com a presunção de mundo aberto (Open World) de OWL, embora houvessem muitas alternativas para resolver o problema, escolhemos utilizar a mais simples, usando o mecanismo de parsing de AIM para OWL.

O próximo capítulo descreve os resultados obtidos com a implementação do classificador TNM, ressaltando as principais contribuições e limitações da pesquisa.

---

# Experimentos

---

Neste capítulo, serão apresentados os resultados que foram alcançados. Em primeiro lugar, avaliamos a expressividade da ontologia AIM4-O. Depois, discutimos como nosso classificador é usado para classificar automaticamente *image findings* com base nos critérios TNM. o que é o objetivo principal deste trabalho. Finalmente, uma avaliação quantitativa do classificador TNM foi feita usando precisão e revocação. Todo o processo foi validado por radiologistas, usuários do Rubin Lab (Radiology Dept. – Stanford University).

## Dados utilizados

Nosso primeiro conjunto de dados é um conjunto de relatórios clínicos de pacientes com carcinoma hepatocelular (HCC) do Genomic Data Commons (GDC)<sup>1</sup> do National Cancer Institute (NCI). Neste trabalho, em todos os experimentos foram usados dados abertos e publicamente disponíveis do GDC. Mas mais importante para a classificação de metadados de imagens foi a exigência de se ter um conjunto de dados de imagens para validar os resultados de relatórios clínicos. O The Cancer Imaging Archive (TCIA)([CLARK et al., 2013](#))<sup>2</sup> é um serviço que de-identifica e hospeda um grande arquivo de imagens médicas de câncer, acessível para download público, e que está relacionado com os relatórios clínicos do GDC por meio de um *Subject ID*. A modalidade de imagem selecionada foi a tomografia computadorizada (CT). Essas imagens foram carregadas para a ferramenta de anotação ePAD.

Embora a avaliação do *cancer staging* possa ser aplicada a outros tipos de câncer, este trabalho enfoca o *staging* de câncer de fígado (HCC) por razões de disponibilidade de dados clínicos e de imagens. Para um dado paciente, a entrada para o classificador

---

<sup>1</sup> <https://gdc.cancer.gov/>

<sup>2</sup> <https://public.cancerimagingarchive.net/ncia/login.jsf>

consiste em arquivos AIM (anotações de imagem) e a saída consiste no estágio de câncer para esse paciente. Neste trabalho, a 7ª edição do TNM foi empregada.

## Avaliação quantitativa da ontologia AIM4-O

De acordo com [Blomqvist, Sepour e Presutti \(2012\)](#): *A avaliação ontológica é o processo de avaliação de uma ontologia com respeito a determinados critérios, usando certas medidas*. Neste trabalho, realizamos a avaliação da ontologia AIM4-O do ponto de vista funcional. Descrevemos a avaliação da ontologia, como uma avaliação centrada em tarefas contrastando suas necessidades ontológicas, tais como questões de competência (CQs) e requisitos de inferência.

Para avaliar a ontologia AIM4-O, estudamos como poderia ajudar na busca de relatórios clínicos que descrevam resultados de análise de imagens (relatos de câncer), para este fim, foram comparadas duas abordagens diferentes:

- Com base em ontologias (busca semântica): Se os relatórios sobre *image findings* são descritos como indivíduos da ontologia AIM4-O, os relatórios podem ser recuperados usando linguagens de consulta de lógica de descrição, como consulta DL ([KOKCIYAN et al., 2014](#)):
- Com base em palavras-chave (busca por palavra-chave): Relatórios clínicos de análise de imagens são geralmente escritos em linguagem natural, portanto, um método para busca pode ser a busca por palavra-chave.

Para destacar as diferenças entre essas duas abordagens, descrevemos quatro consultas expressas tanto em DL (DL query) como em palavras-chave (Tabela 8):

1. Q1 - Find all reports related to an image observation (tumor observation)
2. Q2 - Find all reports that describe multiple tumors
3. Q3 - Find all reports that contain a tumor observation that has a size greater than 8 cm
4. Q4 - Find all reports that contain a tumor observation with descriptors (i.e invasion, mass, vascular)



Query ID	DL query	Keyword query
Q1	hasAnnotations some (hasImageAnnotations some (hasImagingObservation some (ImagingObservationEntity and label value "Lesion type ")))	Tumor
Q2	hasAnnotations min 2	Multiple tumor
Q3	hasAnnotations some (hasImageAnnotations some (hasCalculationEntity some (hasCalculationResult some (some values float [> 8.0f]))))	Tumor size greater than 8 cm
Q4	hasImagingObservationCharacteristic 1 min	Vascular tumor invasion mass

Tabela 8 – Queries em Description Logic (DL query) e palavras chaves para as quatro queries usadas na avaliação quantitativa da ontologia AIM4-O.

Foram considerados os seguintes pontos, a fim de avaliar as duas abordagens:

- A avaliação baseia-se num conjunto de relatórios clínicos do GDC: Foram usados 15 relatórios radiológicos aleatórios, de diferentes pacientes, escritos em linguagem natural e convertidos em instâncias AIM4-O.
- Um relatório é recuperado se ele satisfaz a consulta DL ou se contiver todas as palavras-chave na consulta de busca.
- Finalmente, comparamos a precisão e revocação com o *gold-Standard*. O *gold-Standard* foi determinado por um radiologista manualmente.

Os resultados de precisão e revocação para as quatro consultas são mostrados na Tabela 9.

Query ID	Q1		Q2		Q3		Q4	
	DL-Query	Key-Query	DL-Query	Key-Query	DL-Query	Key-Query	DL-Query	Key-Query
SemanticP	1		1		0.9		0.67	
SearchR	1		1		0.9		1	
KeywordP		1		0.4		0		1
SearchR		0.8		0.4		0		0.7

Tabela 9 – Resultados mostrando os valores de R=recall e P=precision para as 4 queries.

## Observações

Analisando as quatro consultas podemos ver que a busca semântica tem o maior número de documentos relevantes recuperados:

- Q1:** Com a abordagem semântica, 15 relatórios foram recuperados com precisão e revocação de 1 (15/15). Com a busca através de palavras-chaves, 12 relatórios foram recuperados com uma precisão de 1 (12/12) e revocação de 0,8 (12/15).
- Q2:** Com a abordagem semântica, 5 relatórios foram recuperados com precisão e revocação de 1 (5/5). Com a busca através de palavras-chaves, 5 relatórios foram recuperados com precisão e revocação de 0,4 (2/5).
- Q3:** Com a busca semântica, 10 relatórios foram recuperados com uma precisão de 0,9 (9/10) e revocação de 0,9 (9/10). Com a busca por palavras-chave, não foram obtidos relatórios, porque não havia relatórios que continham todas as palavras da consulta (ou seja, tamanho de lesão maior que 8 cm).
- Q4:** Com a busca semântica, 15 relatórios foram recuperados com uma precisão de 0,67 (10/15) e revocação de 1 (10/10). Com a busca de palavras-chave, 7 relatórios foram recuperados com uma precisão de 1 (7/7) e revocação de 0,7 (7/10).

A busca semântica obteve melhores resultados, com valores de revocação próximos de 1 (0,98 em média) e sempre melhores do que a busca através de palavras-chaves (0,48, em média). Além disso, em todos os casos, exceto um, os valores de precisão foram melhores para a busca semântica (em média 0.89, para semântica, e 0.52, para textual). A pesquisa usando palavras-chave teve precisão e revocação 0 para a query Q3, o que é um resultado não desejado.

Na próxima seção, uma vez que avaliamos a nossa ontologia podemos avaliar a capacidade do nosso Classificador baseado em TNM com casos clínicos reais.

## Determinação do estágio de câncer (cancer staging) baseado em TNM

Para avaliar a capacidade de nosso classificador TNM, primeiro criamos um template chamado "template TNM" para fornecer ao radiologista um conjunto de termos semânticos pré-especificados para anotação de imagens. Todas essas anotações de imagem foram armazenadas no formato padrão AIM e são compatíveis com o ePAD. Um exemplo de uma anotação é apresentado na Figura 20.

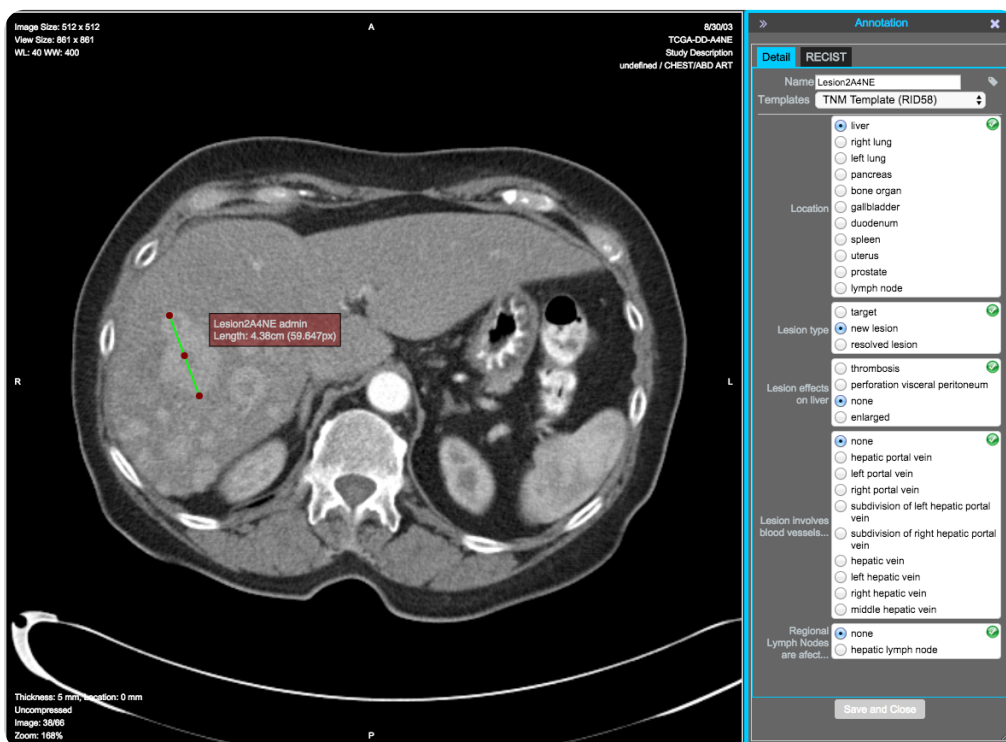


Figura 20 – Image CT do fígado sendo anotada usando o ePAD e a template TNM (mostrada no apêndice à direita da imagem).

Subseqüentemente, geramos anotações de imagem em AIM-XML e as classificamos automaticamente usando o nosso classificador TNM. Esse processo foi devidamente avaliado e considerado válido por um radiologista experiente. Ele analisou visualmente a validade das anotações geradas. O processo que seguimos foi o seguinte:

1. O conjunto de dados utilizado foi obtido dos seguintes bancos de dados abertos GDC e TCIA (os mesmos usados para a avaliação da ontologia). O TCIA contém apenas imagens, número de séries e estudos. Para a classificação TNM de câncer de fígado, buscamos por "LIHC - Liver hepatocellular carcinoma" no GDC. Separamos os pacientes com informações disponíveis tanto no GDC como no TCIA (relatórios médicos e imagens, respectivamente), achamos 52. A informação sobre o tamanho do tumor foi obtida por revisão manual dos relatórios médicos (disponíveis no GDC).
2. Depois de lidos os relatórios médicos, foi criada uma planilha Excel que forneceu informações sobre os *image findings* como: tamanho da lesão, invasão vascular entre outros.
3. Usando esse arquivo Excel e dados do GDC, as anotações em AIM foram criadas e integradas à nossa base de conhecimento (como indivíduos da ontologia AIM4-O).
4. As anotações AIM, sobre cada paciente, foram usadas como entrada do classificador TNM, que calculou o *cancer staging* para cada paciente.

n = 51		Actual Stages				
		I	II	IIIA	IIIB	IVA
Predicted Stages	I	24	2	0	0	0
	II	0	10	2	0	0
	IIIA	0	1	8	0	0
	IIIB	0	0	0	2	1
	IVA	0	0	0	0	1

Tabela 10 – Matriz de Confusão dos estágios de câncer calculados pelo classificador (predicted) versus os valores que os médicos colocaram nos relatórios (actual).

- Os resultados do classificador TNM foram comparados com os resultados do *cancer staging* que os médicos relataram nos relatórios médicos.

As anotações em formato AIM para os 52 pacientes foram geradas pelo autor deste trabalho, a partir dos relatórios médicos. Esses relatórios continham as *image findings* (tamanho da lesão, invasão vascular, etc) e o *cancer stage* para cada paciente, em linguagem natural. Para garantir que o autor entendeu corretamente o que estava escrito no texto dos relatórios, um radiologista experiente analisou relatórios médicos e respectivas anotações AIM de cinco pacientes escolhidos aleatoriamente. Em todos os cinco casos, as *image findings*, reportadas nos relatórios médicos, foram corretamente copiadas para as anotações AIM.

Foram classificadas automaticamente todas as anotações de imagens AIM geradas com base nos 52 diferentes relatórios clínicos, a fim de analisar a precisão da nossa abordagem de *staging* automático com TNM. Os cancer stagings gerados pelo classificador foram comparados aos descritos pelos médicos nos relatórios (nosso *gold standard*). Um paciente, Subject ID TCGA-DD-A1EJ, foi retirado dessa análise. O especialista considerou que a classificação TNM, reportada pelo médico, estava incorreta (mais informações adiante).

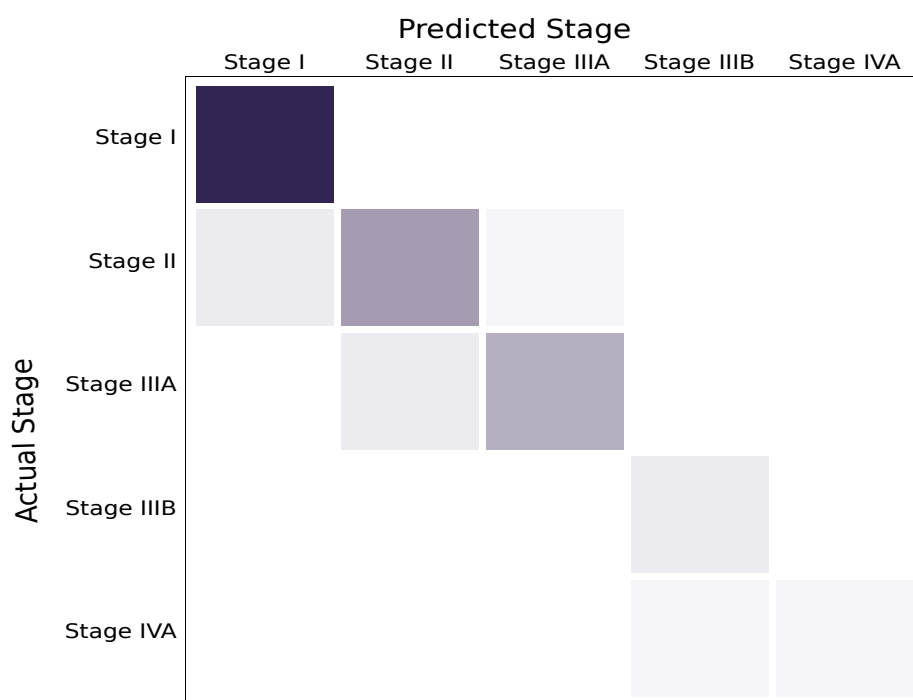
Para o cálculo de precisão e revocação, quando o *staging* automático coincide com o dado pelo médico, ele é considerado positivo. Na Tabela 10, é mostrada a Matriz de Confusão com os estágios de câncer calculados pelo classificador versus os valores que os médicos colocaram nos relatórios.

A precisão foi 85,7% e a revocação 81,0% (para os 51 pacientes). Isso significa que, para a precisão, em 85.7% das vezes o sistema concordou com o *staging* dado pelos médicos. Para a revocação, isso significa que, de todas as vezes que um dado *stage* foi informado por um médico, em 81.0% dos casos o sistema concordou com ele. É importante salientar que, mesmo quando o sistema divergiu dos médicos, a diferença máxima entre eles foi de apenas um estágio. Discutimos os cenários de *staging* dos pacientes com a ajuda do nosso especialista e parceiro do projeto do Radiology Department da Stanford University School of Medicine e ele achou os resultados muito bons.

Os resultados da avaliação são resumidos na matriz de escala de cores, Figura 21. Ela representa uma matriz de confusão para uma classificação de stage multicêntrico. A interpretação dos resultados é como se segue: quanto mais preto seja o pixel na diagonal da matriz, foi classificado melhor essa classe. pixels no Grays localizados na horizontal indicam que a classe atual foi confundido pelo classificador com a classe correspondente no seu eixo vertical.

Para *stages* iniciais de câncer, como I, II, IIIA, a porcentagem de classificações incorrectas (i.e., falsos positivos e falsos negativos) foi muito pequeno. Elas são representadas pela diagonal realçada da matriz (Figura 21). Para estágios de câncer mais avançados, como IIIB ou IVA, ela foi maior (Figura 22). Isso pode ter acontecido pelo fato desses estágios serem descritos por conceitos relativamente mais complexos ou, simplesmente, por termos poucos pacientes nesses estágios.

Figura 21 – Matriz de escala de cores para a Matriz de Confusão.



Além disso, o classificador revelou o fato de que existem alguns relatórios clínicos com estágios que não seguem o padrão do TNM. Por exemplo, no caso do paciente com Subject ID TCGA-DD-A1EJ, o resultado do classificador TNM foi considerado correto, após validação pelo nosso especialista. O estágio previsto pelo classificador foi *Stage I*, mas a informação presente no relatório médico <sup>3</sup> indicava *Stage III*. Como esse foi o único caso em que a diferença entre o estágio dado pelo classificador e os médicos diferiu em mais de um nível, o especialista quis analisar o relatório e declarou que o estágio dado pelo

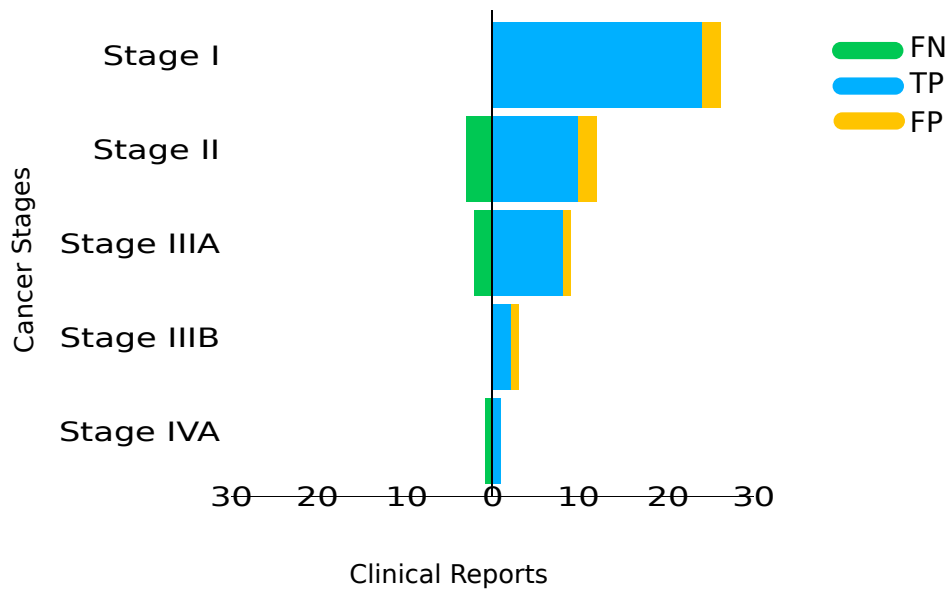
<sup>3</sup> <https://gdc-portal.nci.nih.gov/cases/52292ffc-0902-4d97-b461-20723987a177>

classificador estava correto. Ele recomendou que não usássemos os dados desse paciente e, por isso, esse relatório foi excluído dessa análise.

Tais incidentes não são excepcionais, porque o processo de *staging* dos pacientes com câncer através da análise de imagens, é uma tarefa com base na inspeção de imagens e podem ocorrer variações na interpretação de imagens entre especialistas.

Com ajuda de sistemas de suporte à decisão clínica, baseados em metadados de imagens (como classificadores TNM), casos clínicos com estagio errôneo podem ser detectados e, por conseguinte, erros de tratamento médico podem ser reduzidos. Isso mostra a importância de se melhorar sistemas de suporte à decisão clínica, pelo processamento automático de metadados de imagem na avaliação do câncer.

Figura 22 – Resumo dos Histogramas para cada estágio após da classificação TNM dos metadados dos 51 pacientes



## Considerações Finais

Neste capítulo foram descritos os resultados obtidos com o desenvolvimento do classificador TNM. Inicialmente usamos e comparamos duas abordagens, a busca semântica e a busca por palavras-chave, para avaliar a ontologia AIM4-O. Nessa análise, nossos resultados mostraram que a busca semântica obteve melhores resultados. Isto se deve ao fato de que em relatórios clínicos baseados em ontologias, máquinas tem acesso a mais informações sobre os dados e assim podem ter mais sucesso, com maior precisão e revocação, em tarefas de busca.

Após demonstrar como o uso da ontologia AIM4-O pode ser benéfico no processo de buscas por informação, testamos o classificador TNM com relatórios de 51 pacientes com câncer de fígado. Ele obteve resultados de precisão e revocação muito bons (85.7% e

81,0% respetivamente), quando comparamos suas predições de estágios de câncer com o que foi reportado por médicos.

O próximo capítulo conclui esta dissertação, discutindo os resultados obtidos. Além disso, ressaltam-se as principais contribuições e limitações da pesquisa, finalizando com trabalhos futuros.





---

## Conclusão

---

Nesta dissertação de mestrado, geramos automaticamente o *Cancer Staging* (estágios do câncer) de lesões cancerosas presentes em imagens médicas, utilizando tecnologias de *reasoning* da web semântica. O processo de *staging* em pacientes por análise de imagem é uma tarefa realizada por especialistas, como oncologistas e radiologistas, e, muitas vezes, envolve trabalho intensivo que exige precisão na interpretação de lesões de câncer. Essa precisão de especialistas é alcançada através do treinamento e da experiência, mas as variações na interpretação de imagens são uma limitação das observações humanas. Nesse contexto, desenvolvemos uma abordagem de *staging* automático (classificador TNM), baseada em metadados (anotações) de imagens, que pode ajudar médicos a obter uma maior precisão e padronização na determinação de estágios de câncer em pacientes.

## Contribuições

As principais contribuições deste trabalho são detalhadas a seguir:

- Uma abordagem (ou classificador TNM) para gerar automaticamente o estágio de um paciente com câncer, baseado no standard TNM e anotações em imagens médicas (no formato AIM), a fim de fornecer aos médicos uma maior taxa de precisão e padronização na interpretação de estágio de lesões cancerosas.
- Teste da capacidade do classificador TNM desenvolvido, através da análise quantitativa e discussão com especialistas dos cenários reais de *cancer staging* que foram avaliados. Nessa discussão, podemos destacar, além dos bons indicadores de precisão e revocação obtidos pelo classificador (85,7% e 81,0%), que podem ser detectados casos clínicos com estagio errôneo e que, por conseguinte, potenciais erros de tratamento médico podem ser reduzidos.

- Uma nova ontologia AIM4-O, baseada no trabalho de [Bulu e Rubin \(2015\)](#), que fornece uma representação semântica do modelo de dados AIM 4.0. AIM4-O é uma ontologia leve que foi desenvolvida com a intenção de poder ser adotada em aplicações em outras disciplinas relacionadas ao processamento de metadados de imagens. AIM4-O está de acordo com o perfil OWL-DL. Ela pode ser usada diretamente em aplicações ou servir como modelo de referência, para a criação de ontologias de domínio específico, como, por exemplo, em radiologia.
- Avaliação quantitativa, centrada em tarefa, da ontologia AIM4-O. Buscas semânticas de relatórios sobre imagens médicas, usando a AIM4-O, apresentaram melhor precisão (89% x 52%) e revocação (98% x 48%) que buscas usando palavras-chaves no mesmo conjunto de dados.
- O método de mapeamento para transformar automaticamente anotações, em formato AIM-XML, para instâncias da ontologia AIM4-O. Ele permite processar fluxos de documentos AIM e preencher uma base de conhecimento em OWL com o seu conteúdo. As anotações podem ser obtidos de qualquer ferramenta de marcação que use o AIM padrão.
- Alinhamento do modelo AIM com ontologias médicas, como Radlex e Onlira, a fim de facilitar o trabalho de classificação e tratamento dos indivíduos da ontologia AIM4-O.
- Representação formal, com base na lógica de primeira ordem, de regras, axiomas e procedimentos de raciocínio automático a fim de automaticamente fazer *cancer staging* baseado em TNM.

É importante notar que o procedimento de *staging* automático não dá ao médico novas informações. Ele deve ser visto como uma segunda opinião para fins de qualidade no diagnóstico clínico, portanto o objetivo final é a melhoria da qualidade e padronização dos diagnósticos médicos.

## Dificuldades e Limitações

O presente trabalho apresentou algumas dificuldades e limitações durante seu projeto e implementação:

- Algumas limitações a este trabalho é que nosso conjunto de dados é pequeno. Necessitamos da cooperação de radiologistas, para interpretar os exames e validar resultados, e eles têm muito pouco tempo disponível para essa atividade,
- Um fator limitante encontrado diz respeito a pouca disponibilidade de tempo dos especialistas para realizar os testes.

- A distância dos nossos parceiros radiologistas, do Radiology Dept. da Stanford University, e a barreira linguística também foram dificuldades para este trabalho.

## Trabalhos Futuros

Como possíveis trabalhos futuros decorrentes dessa dissertação destacam-se:

- Algoritmos automáticos de segmentação de imagem já são atualmente utilizados para detectar lesões em imagens de CT ou MRI (por exemplo [Oberkampff et al. \(2016\)](#)). Assim, a extração automática de representações estruturadas de lesão e seu tamanho em imagens está ao alcance. Com representações estruturadas (por exemplo, na Figura 12), grande parte das anotações, baseadas no critério TNM, em imagens pode ser automatizada em trabalhos futuros.
- Embora não tenhamos realmente feito isso, assumimos que nossa abordagem poderia ser aplicada a outros tipos de câncer como cólon, pulmões, e outros critérios de classificação, como o Recist. Seria necessário somente a modificação das regras e axiomas que representam os critérios TNM, evitando assim criar uma nova ontologia.
- A integração de uma DSL, ou Linguagens de Domínio Específico, como ferramenta de comunicação entre os médicos e a representação formal dos critérios TNM.
- Trabalhos futuros incluirão conjuntos de dados mais variados para avaliação, expansão do classificador para outros órgãos e incorporação em sistemas de informação existentes. O classificador TNM tem potencial para ser integrado a sistemas de software maiores.



# 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 na página 53.)

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](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 nas páginas 49 e 50.)

BLOMQVIST, E.; SEPOUR, A. S.; PRESUTTI, V. Ontology testing - methodology and tool. In: \_\_\_\_\_. *Knowledge Engineering and Knowledge Management: 18th International Conference, EKAW 2012, Galway City, Ireland, October 8-12, 2012. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 216–226. ISBN 978-3-642-33876-2. (Citado na página 88.)

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 19, 37, 67, 68, 69, 86 e 98.)

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 na página 51.)

- CHANNIN, D. S. et al. The cabig<sup>TM</sup> 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.)
- CLARK, K. et al. The cancer imaging archive (tcia): Maintaining and operating a public information repository. *Journal of Digital Imaging*, v. 26, n. 6, p. 1045–1057, 2013. ISSN 1618-727X. Disponível em: <http://dx.doi.org/10.1007/s10278-013-9622-7>. (Citado na página 87.)
- 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](http://dx.doi.org/10.1007/11574620_69). (Citado nas páginas 52 e 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](http://protege.stanford.edu/conference/2006/submissions/abstracts/8.2_2006protege-dameron-short.pdf). Acesso em: 02 March. 2015. (Citado nas páginas 44, 45, 46, 60, 61 e 65.)
- 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:PMC1789017] [PubMed:4552594]. (Citado na página 38.)
- FARIA, S. C. et al. Tnm/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 nas páginas 45, 46 e 77.)
- 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 57.)
- 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 51.)
- 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 57.)

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 51.)

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 37 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](https://doi.org/10.1148/radiol.220001257)] [PubMed:[11526283](https://pubmed.ncbi.nlm.nih.gov/11526283/)]. (Citado na página 38.)

KITCHENHAM, B.; CHARTERS, S. *Guidelines for performing Systematic Literature Reviews in Software Engineering*. 2007. (Citado na página 59.)

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 76, 77 e 88.)

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 42.)

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, 43, 61 e 64.)

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 nas páginas 55 e 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](https://pubmed.ncbi.nlm.nih.gov/PMC2815449/)] [PubMed:[20351880](https://pubmed.ncbi.nlm.nih.gov/20351880/)]. (Citado nas páginas 31, 33, 61 e 62.)

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 41.)

MARQUET, G. et al. Grading glioma tumors using OWL-DL and NCI Thesaurus. *AMIA Annu Symp Proc*, p. 508–512, 2007. [PubMed Central:PMC2655830] [PubMed:18693888]. (Citado nas páginas 31, 61, 62 e 65.)

MASSICANO, F. et al. An ontology for tnm clinical stage inference. In: *ONTOBRAS*. [S.l.: s.n.], 2015. (Citado nas páginas 61, 64 e 65.)

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. I*, p. 1–6, 2012. Disponível em: <<http://www.cyberjournals.com/Papers/Mar2012/06.pdf>>. Acesso em: 19 September. 2016. (Citado nas páginas 61, 63 e 65.)

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](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](http://dx.doi.org/10.1007/978-3-642-29764-9_3)>. (Citado na página 63.)

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](http://dx.doi.org/10.1007/978-0-387-34685-4_2)>. (Citado na página 41.)

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 42.)

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. (Citado na página 62.)

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 61, 63 e 64.)



- OBERKAMPF, H. et al. Semantic representation of reported measurements in radiology. *BMC Medical Informatics and Decision Making*, v. 16, n. 1, p. 5, 2016. ISSN 1472-6947. Disponível em: <<http://dx.doi.org/10.1186/s12911-016-0248-9>>. (Citado na página 99.)
- 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 61 e 64.)
- 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](http://dx.doi.org/10.1007/978-1-84628-885-2_4)>. (Citado na página 55.)
- 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 41.)
- 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, 56 e 57.)
- 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 64.)
- 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 64.)

- 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, 54 e 55.)
- 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] [DOI:10.1259/bjr/33620087] [PubMed: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 61, 63 e 65.)
- 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 61, 62, 63 e 65.)
- 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 39 e 40.)
- 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 61, 62 e 65.)
- 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 61, 63, 65 e 66.)