
Uma Aplicação Médica para Classificação Automática de
Estágios de Câncer usando Anotações em Imagens de
Tumores e Ontologias

Edson Luque Mamani

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito: 01/08/2015

Assinatura: _____

Uma Aplicação Médica para Classificação Automática de Estágios de Câncer usando Anotações em Imagens de Tumores e Ontologias

Edson Luque Mamani

Orientador: Prof. Dr. Dilvan de Abreu Moreira

Monografia apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, para o Exame de Qualificação, como parte dos requisitos para obtenção do título de Mestre em Ciências - Ciências de Computação e Matemática Computacional.

USP – São Carlos
Março de 2015

Resumo

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

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

Sumário

1	Introdução	1
1.1	Contextualização	2
1.2	Motivação	3
1.3	Objetivo	3
1.4	Organização	4
2	Informática Biomédica	5
2.1	Sistemas de Informação Hospitalares	6
2.1.1	Picture Archiving and Communication System – PACS	6
2.2	Tecnologias para Imagens Médicas	7
2.2.1	Tomografia Computadorizada	8
2.2.2	Positron Emission Tomography	9
2.3	Anotação de Imagens	10
2.3.1	Anotação Semântica	10
2.3.2	Anotação de Imagens Médicas com Vocabulários Estruturados	11
2.4	Medical Background Knowledge	12
2.4.1	Estágio do Câncer	13
2.4.2	<i>Cancer Staging</i> : O Padrão TNM	13
2.5	Considerações Finais	13
3	Web semântica	17
3.1	Web Semântica	17
3.2	Ontologias	18
3.2.1	Web Ontology Language – OWL	19
3.2.2	Semantic Web Rule Language - SWRL	20
3.3	Bio Ontologies	20
3.3.1	Foundational Model of Anatomy – FMA	21
3.3.2	RadLex	21
3.3.3	Annotation Imaging Markup	22

3.4	Considerações Finais	23
4	Trabalhos Relacionados	25
4.1	Planeamento e Execução do Mapeamento Sistemático	25
4.2	Apresentação e Análise dos Resultados Obtidos no Mapeamento	26
4.3	Considerações Finais	32
5	Plano de Trabalho	33
5.0.1	Arquitetura proposta para desenvolvimento do trabalho	33
5.1	Atividades	34
5.1.1	Atividades Concluídas até o Momento	36
5.1.2	Resultados esperados	36
6	Referências	37

Lista de Figuras

2.1	Fluxos de dados e os principais componentes de um PACS. Fonte: (Serique, 2012)	7
2.2	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:(Arnold, 2012)	8
2.3	Estudo de paciente com metástase de cólon, pode ser visto que o aumento da captação de glicose no tecido canceroso é facilmente visível. Fonte:Ziegler (2005)	9
2.4	A imagem CT do fígado anotada com termos semânticos. Fonte : (Kurtz et al., 2014).	11
2.5	Um extrato da ontologia RadLex que é uma terminologia controlada em radiologia. Fonte: (Kurtz et al., 2014)	12
3.1	Arquitetura em camadas da Web Semântica, fonte: (Berners-Lee et al., 2001)	18
5.1	Arquitetura proposta.	34

Lista de Tabelas

2.1	Os critérios de T, N e M para câncer de pulmão. Fonte: Dameron et al. (2006a,b)	14
2.2	Definições para as diferentes estágios de tumores. Fonte: Dameron et al. (2006a,b)	15
4.1	Comparativo entre os trabalhos relacionados.	27
4.1	Comparativo entre os trabalhos relacionados.	28
5.1	Cronograma do projeto.	36

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
HIS	Hospital Information System
LIS	Laboratory Information System
NCI	National Cancer Institute
NGS	Nottingham Grading System
OWL	Web Ontology Language
PACS	Picture Archiving and Communication System
PET	Positron Emission Tomography
RadLex	Unified Language of Radiology Terms
RDF	Resource Description Framework
RIS	Radiology Information System
SWRL	Semantic Web Rule Language

Lista de Tabelas

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

1 Introdução

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

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

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

Embora esses algoritmos possam auxiliar a comunidade biomédica, quanto ao uso de imagens no tratamento de câncer, eles apresentam problemas quando uma consulta abstrata e ambígua é feita no contexto da classificação de pacientes com câncer. Por exemplo, quando um oncologista quer saber se um tumor já se encontra em estado avançado perto de se espalhar para alguma região próxima da origem do câncer, mas não para outras partes do corpo (Wennerberg et al., 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. Em geral, a semântica é o estudo do significado das coisas e, no contexto de Oncologia e Radiologia, a semântica é o estudo do significado do conteúdo dos laudos de radiologia, bem como o conteúdo e o significado das imagens usadas.

1.1 Contextualização

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

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

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

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

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

1.2 Motivação

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

Por outro lado, embora sistemas como o ePAD permitam a criação de anotações de imagem no formato AIM, facilitando assim a manipulação e consulta de metadados de imagens em AIM, eles não permitem representar anotações em um formato que seja diretamente adequado para o raciocínio (*reasoning*). O AIM fornece apenas um formato para transferência e armazenamento de dados. Um fato importante é que, atualmente, não há muitos métodos de *reasoning* semântico para se fazer inferências sobre lesões cancerígenas utilizando anotações sobre imagens médicas codificadas pelo AIM (Levy et al., 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 para *reasoning* semântico para fazer inferências sobre lesões cancerígenas a partir de anotações semânticas em AIM. Sendo assim, a principal motivação deste trabalho é a possibilidade de desenvolver métodos de *reasoning* baseados em sistemas de notação, como a *TNM Classification of Malignant Tumors* (TNM), para a classificação automática de pacientes com câncer, a partir de anotações AIM sobre imagens desses pacientes. e incorporá-los aos requisitos atendidos por sistemas de anotação de imagens, como ePAD.

1.3 Objetivo

Gerar automaticamente o *cancer staging* (estágio do câncer) de lesões cancerígenas, presentes em imagens médicas, utilizando tecnologias de *reasoning* e web semântica para processar metadados sobre anotações dessas imagens gerados por programas como o ePAD, para fornecer aos médicos uma segunda opinião sobre a classificação dos pacientes. Essa

1 Introdução

geração automática de informação, poderá aumentar a eficiência de radiologistas e oncologistas e garantir a qualidade uma melhor qualidade e uniformidade na interpretação de imagens por especialistas.

O cancer staging é uma tarefa de classificação onde é necessária a representação de conhecimento. No caso deste trabalho, esse conhecimento será codificado em ontologias (OWL) e regras (SWRL). Atualmente, existem sistemas de *cancer staging*, no entanto eles não são de código aberto e seus métodos de classificação não podem ser analisados ou reutilizados abertamente. Além disso, eles também apresentam problemas de precisão, como contradições nos resultados do *staging* (Zillner und Sonntag, 2012).

1.4 Organização

Este trabalho está estruturado da seguinte forma:

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

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

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

Capítulo 5: Esse capítulo mostra o cronograma a ser seguido e o plano de trabalho futuro.

2 Informática Biomédica

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

Embora em radiologia muitas coisas, incluindo a criação da imagem, sejam bastante padronizadas, existem muitas maneiras de se descrever o conteúdo de uma imagem, incluindo diferentes sinônimos de termos, vários níveis de descrição e foco em atores diferentes. Nesse contexto, o uso de terminologias visuais e estruturadas para anotações em imagens é uma abordagem promissora, pois permite a definição precisa de imagens médicas utilizando a semântica fornecida pelas anotações (Rubin, 2011).

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, não podem ser processadas por máquinas. Para facilitar o processamento dessas anotações por máquina, é necessário adicionar informações semânticas que sejam baseadas em fontes de conhecimento, como *staging systems*. Além disso, é preciso que essas informações sejam codificadas de forma 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* (Dasmahapatra et al., 2003).

Como o foco deste trabalho é a construção de um sistema para permitir o uso de *reasoning* em anotações contidas em imagens médicas. Este capítulo abordará os principais conceitos referentes a Informática Radiológica. Na Seção 2.1 são apresentados os HISs. Em seguida, na Seção 2.2, Imagens Médicas. Na seção 2.3, é discutido o processo de

Anotações em imagens. Por fim, na seção 2.4, serão descritos os sistemas de classificação que são a base de conhecimento para a classificação dos pacientes com câncer, a partir de imagens radiológicas.

2.1 Sistemas de Informação Hospitalares

Os Sistemas de Informação Hospitalares, ou Hospital Information Systems (HISs), surgiram por volta da década de 60 para atender às necessidades financeiras das instituições médicas (Medicine (2003)). Além disso, os HISs têm agregado várias funcionalidades como: atendimento clínico, protocolos clínicos, lembretes, alertas e gestão administrativa, que possibilitam incrementar o ensino e pesquisa realizados em instituições hospitalares.

A pesquisa realizada nas instituições hospitalares permite a descoberta de novos conhecimentos e a descoberta de novas tendências numa população, como, por exemplo, por meio da mineração de dados. Outros HISs, relacionados às informações médicas, são ainda mais específicos, lidando com particularidades de um determinado departamento clínico. Dessa forma, esses sistemas atuam como subconjuntos das funcionalidades de um HIS, como exemplos destes sistemas temos o Laboratory Information System (LIS) e o Radiology Information System (RIS).

Os LISs trabalham com o fluxo de trabalho dos serviços laboratoriais e de patologia, oferecendo interfaces de comunicação com os diversos dispositivos utilizados para a realização das avaliações químicas e de imunologia, microbiologia, genética e marcadores histopatológicos. Os LISs são vistos como um modelo de monitoramento, validação de resultados e divulgação eletrônica dos mesmos aos médicos solicitantes.

Já os RISs atendem ao fluxo de trabalho de um departamento de radiologia (Huang, 2010b). Onde eles têm funcionalidades de agendamento de pacientes, gerenciamento de imagens e estudos de relatórios, que podem ser revisões de resultados feitos para estudos anteriores.

Embora os RISs sejam similares aos PACSs. Eles se destacam dos PACSs por, normalmente, serem sistemas restritos aos dados das imagens, enquanto os PACSs estão preocupados com os aspectos de armazenamento e manipulação dos dados das imagens.

2.1.1 Picture Archiving and Communication System – PACS

Um Picture Archiving and Communication System (PACS) é um conjunto formado por dispositivos de captura de imagens, unidades de armazenamento, processamento computacional e banco de dados de imagens médicas integrados em rede (Huang, 2010b).



Figura 2.1: Fluxos de dados e os principais componentes de um PACS. Fonte: (Serique, 2012)

Desenvolver um sistema PACS envolve diversos desafios, como, por exemplo, a natureza e o tamanho dos estudos das imagens adquiridas, podendo variar de 2 até mais de 1000 MB (Alex A.T. Bui, 2010). Isso ocorre porque os novos equipamentos permitem gerar uma grande quantidade de imagens de alta resolução (Philbin et al., 2011).

A Figura 2.1 apresenta os componentes principais de um PACS e o fluxo de dados entre eles. Um desses componentes são os dispositivos de imagem e computadores de aquisição. Exemplos desses dispositivos digitais são: tomografia computadorizada (CT), ressonância magnética (RM), radiografia computadorizada (CR), tomografia por emissão de pósitrons (PET), raios-X digital, mamografia e ultrassom. Em muitos casos, os dispositivos de imagens são interligados a computadores de aquisição, pois os dispositivos de imagens fornecem apenas sinais brutos que precisam ser fatiados ou volumetrizados. Os computadores de aquisição armazenam uma pequena quantidade de dados localmente e, depois, enviam esses estudos (imagens) para os PACSs, através de um gateway, para o armazenamento definitivo de longo prazo (Serique, 2012).

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

2.2 Tecnologias para Imagens Médicas

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

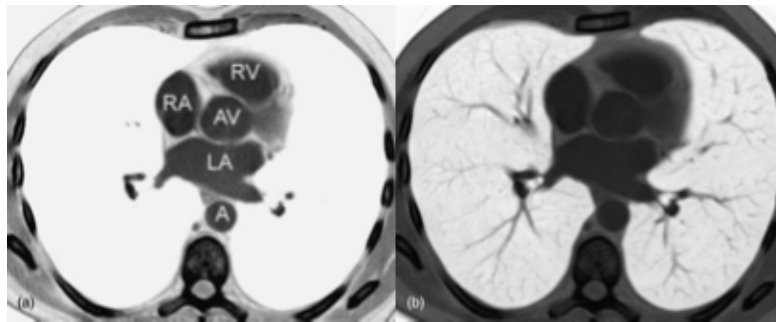


Figura 2.2: CT do peito, (a) anatomia do mediastino: Átrio direito (RA), Ventriculo direito (RV), Válvula aórtica (AV), Aorta (A), Átrio esquerdo (LA). (b) Anatomia do pulmão. Fonte:(Arnold, 2012)

ou outros métodos invasivos (Imaging, 2007).

Antes de analisar como é que a semântica pode melhorar ainda mais o processo de monitoramento do progresso dos pacientes com câncer, é importante compreender as tecnologias de imagem primárias com que iremos trabalhar em nossa proposta e o que elas fazem. Essas tecnologias de imagem abrangem Computed Tomography (CT) e Positron Emission Tomography (PET).

2.2.1 Tomografia Computadorizada

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

Por exemplo, no CT do peito uma vasta gama de densidades de tecido está presente, uma boa imagem das estruturas do mediastino são mostradas em detalhes do pulmão, mostrados na Figura 2.2.

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(Imaging, 2007).

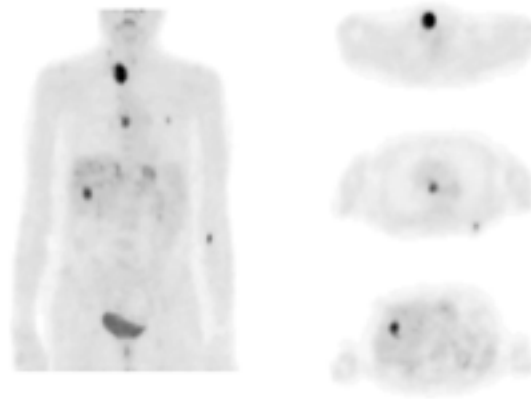


Figura 2.3: Estudo de paciente com metástase de cólon, pode ser visto que o aumento da captação de glicose no tecido canceroso é facilmente visível. Fonte:Ziegler (2005)

2.2.2 Positron Emission Tomography

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

Nesses estudos oncológicos PET, o paciente recebe uma injeção de uma substância que consiste de uma pequena quantidade de um composto marcado radioativamente semelhante ao açúcar comum, glicose. O açúcar radioativo pode ajudar na localização de um tumor, porque as células cancerosas utilizam glicose mais avidamente do que outros tecidos no corpo(Ziegler, 2005).

Por exemplo, a Figura 2.3 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(Imaging, 2007).

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

Todas as modalidades de análise acima mencionadas fornecem uma visão detalhada sobre a anatomia humana, suas funções, 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 und Bretschneider, 2013).

Embora imagens não processadas (raw images) possam ser úteis em algumas aplicações

de processamento por computador, grande parte do conteúdo semântico nas imagens e dos relatórios radiológicos associados a elas, não é explícito e disponível para computadores. Isso dificulta a descoberta de novas informações num grande volume de informação. Com base nisso, as aplicações médicas avançadas devem possuir descrições semânticas sobre dados clínicos, como as imagens médicas.

2.3 Anotação de Imagens

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

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

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

As anotações semânticas estão diretamente ligados ao alto nível de compreensão e descrições do usuário sobre as características da imagem (Rubin, 2011). Com base nessas considerações, incorporar características semânticas em sistemas de decisão sobre diagnóstico pode ser uma tentativa promissora para preencher a lacuna semântica entre a descrição visual de uma imagem e seu significado (Ma et al., 2010), Figura 2.4.

2.3.1 Anotação Semântica

De acordo com (Nagarajan, 2009), 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

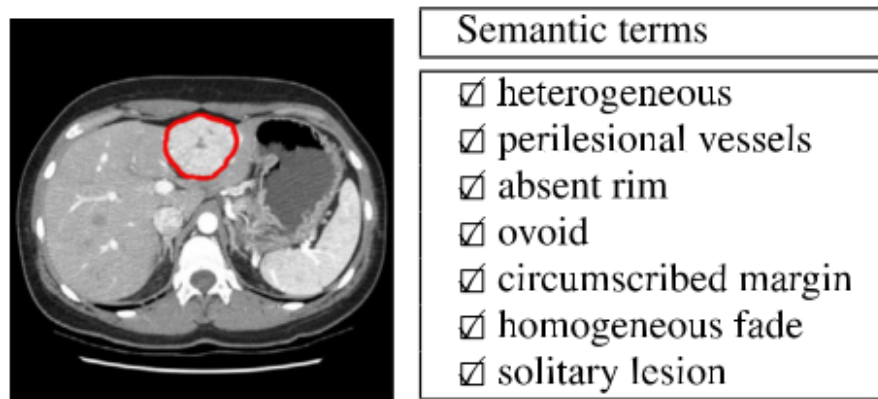


Figura 2.4: A imagem CT do fígado anotada com termos semânticos. Fonte : (Kurtz et al., 2014).

da biomedicina. esse processo é conhecido como anotação semântica de imagens médicas.

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

2.3.2 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 os de ajuda a decisão em diagnóstico. Para lidar com esse problema, os últimos trabalhos no domínio semântico usam vocabulários controlados para anotar as imagens (Korenblum et al., 2011).

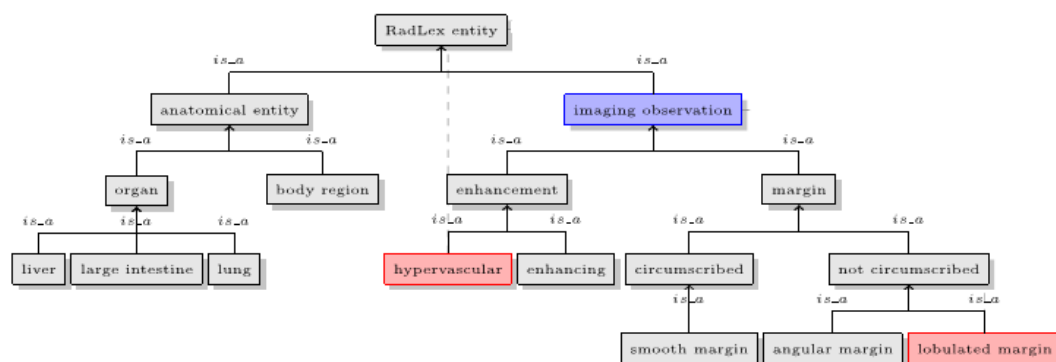


Figura 2.5: Um extrato da ontologia RadLex que é uma terminologia controlada em radiologia. Fonte: (Kurtz et al., 2014)

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 recentes, 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 trabalho indica que as anotações semânticas, usando um vocabulário controlado, podem levar a diagnósticos mais precisos.

2.4 Medical Background Knowledge

As ontologias estão relacionadas com vocabulários controlados. Elas fornecem uma maneira formal para modelar conhecimento (*Background Knowledge*) e geralmente são construídas a partir de um consenso entre especialistas de um domínio (Guarino, 1995). As ontologias representam uma poderosa forma de estruturar termos semânticos pertencentes a uma fonte de conhecimento particular.

No contexto da medicina, várias ontologias estão sendo desenvolvidas para organizar conceitos biomédicos de uma forma abrangente, por exemplo: *Medical Subject Headings (MeSH)*, *International Classification of Diseases (ICD Taxonomy)*, *Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT)*, *Unified Language of Radiology Terms (RadLex)* (Kurtz et al., 2014). Na Figura 2.5, vemos uma parte da ontologia RadLex.

Para classificar pacientes com câncer é necessário representar o conhecimento geral

do domínio (*background knowledge*). Nesse caso, esse conhecimento é chamado de *staging system*. No entanto, ter representações formais das definições de um *staging system* numa ontologia não é suficiente para fornecer ajuda automática na classificação de tumores de câncer. Esse conhecimento deve ser representado também em regras de raciocínio (*reasoning*). Portanto, representar o conhecimento sobre oncologia em ontologias e regras parece uma contribuição desejável para a realização da classificação automática de tumores. Conceitos como ontologias, regras e *reasoning* serão melhor detalhados no próximo capítulo.

2.4.1 Estágio do Câncer

O estágio do câncer categoriza a progressão de um câncer no corpo, em termos da extensão do tumor primário e de se espalhar para alguma região próxima da origem do câncer. A rotina para determinar o estágio de um câncer (*cancer staging*) em pacientes tem uma série de vantagens reconhecidas por organizações relacionadas a câncer em todo o mundo. O *cancer staging* permite que os médicos possam determinar o tratamento mais apropriadamente, avaliar os resultados de forma mais confiável e comparar as estatísticas a nível local, regional e nacional com mais confiança (Mccowan et al., 2000).

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*).

2.4.2 Cancer Staging: O Padrão TNM

O *cancer staging* no padrão *TNM Classification of Malignant Tumors* (TNM) é um processo de duas etapas. A primeira etapa consiste em dar notas para três conceitos: descrição do tumor (T), difusão em nós linfáticos (N) e possível metástase (M). Esses conceitos estão resumidos na Tabela 2.1.

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., 2006a). Esses conceitos estão resumidos na Tabela 2.2.

2.5 Considerações Finais

Este capítulo forneceu uma base teórica para compreensão de conceitos como vocabulários controlados, anotações de imagem, anotação semântica e representação do conhecimento.

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

Tabela 2.1: Os critérios de T, N e M para câncer de pulmão. Fonte: Dameron et al. (2006a,b)

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

Tabela 2.2: Definições para as diferentes estágios de tumores. Fonte: Dameron et al. (2006a,b)

Foi mostrado também como esses conceitos podem ser utilizados na área de informática biomédica para melhorar processos como a interpretação de imagens médicas em câncer.

3 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".

Conceitos relacionados a Web Semântica e ontologias foram mencionados no capítulo anterior, neste capítulo esses conceitos vão ser apresentados e discutidos. O capítulo está organizado da seguinte maneira, na Seção 3.1 a Web Semântica é discutida. Na Seção 3.2 serão expostos conceitos referentes à Ontologias, a Web Ontology Language (OWL) é exposta na seção 3.2.1. Em seguida, na seção 3.3 apresentaremos as principais bio-ontologias e terminologias envolvidas nas pesquisas deste trabalho.

3.1 Web Semântica

Berners-Lee et al. (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 de conceitual sobre a Web semântica, Berners-Lee propôs a arquitetura da Web Semântica conhecida como “bolo de noiva”, que está ilustrada na Figura 3.1. A idéia central dessa arquitetura é de que cada camada vai gradativamente trazendo uma nova contribuição como maior expressividade, possibilidade de se realizar inferências e autenticação. A arquitetura da Web Semântica está categorizada nas seguintes camadas: camada URI/Unicode, camada sintática, camada de dados, camada de ontologia, camada lógica, camada de prova, camada de validação e camada de assinatura digital.

A base da arquitetura de camadas da Web Semântica é formada pelos padrões Unicode e URIs. Eles facilitam o intercâmbio de dados: o Unicode permite aos computadores representar e manipular caracteres em quase todas as línguas existentes e URIs (*Uniform*

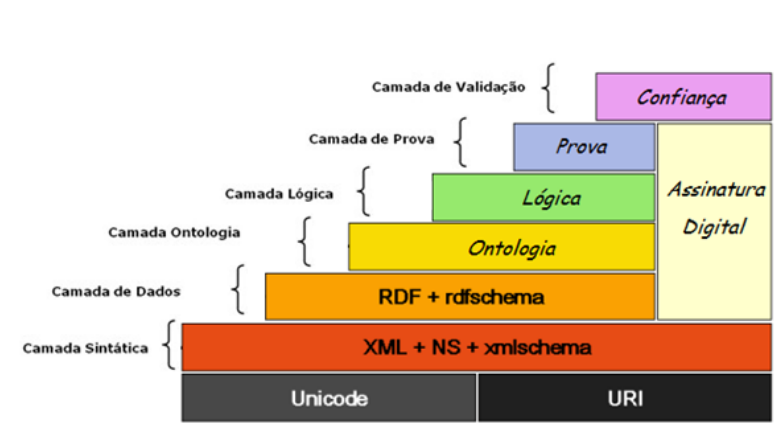


Figura 3.1: Arquitetura em camadas da Web Semântica, fonte: (Berners-Lee et al., 2001)

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.

3.2 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 und 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. Noy und McGuinness (2001) afirmam que as ontologias surgiram como artefatos para representar, organizar e compartilhar conhecimento 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).

3.2.1 Web Ontology Language – OWL

Embora a proposta inicial da Web Semântica fosse utilizar o RDF, essa linguagem demonstrou limitações em poder de expressividade (Horrocks et al., 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, surgiram a OWL 1.1, uma revisão incremental da linguagem e, depois de extensivas discussões, a OWL 2 (abril de 2008)¹. A OWL 2 foi considerada um passo substancial para a evolução da linguagem (Grau et al., 2008).

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 (Bechhofer et al., 2004).

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

¹<http://www.w3.org/TR/owl-features/>

UML, diagramas de Entidade de Relacionamento, e esquemas de banco de dados (Bechhofer 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 (Bechhofer et al., 2004).

3.2.2 Semantic Web Rule Language - SWRL

Dado que a interoperabilidade é uma das metas principais da Web Semântica e que as regras são uma parte fundamental dessas metas, um passo fundamental para fazer essa interoperação na Web é a Semantic Web Rule Language (SWRL)², que foi projetada para ser o idioma de regras da Web Semântica(Connor et al., 2005).

A SWRL é baseada numa combinação das sub-linguagens OWL-DL e OWL-Lite da OWL e as sub-linguagens Unary/Binary Datalog da *Rule Markup Language*³.

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

A especificação da SWRL⁴ não impõe restrições sobre como o raciocínio deve ser realizado com regras SWRL. Assim, pesquisadores são livres para usar uma variedade de mecanismos de regras para raciocinar com regras SWRL armazenadas em uma base de conhecimento OWL. Dessa forma, a SWRL fornece um ponto de partida conveniente para a integração de sistemas de regras para trabalhar com a Web Semântica(Connor et al., 2005).

3.3 Bio Ontologies

Grandes volumes de dados digitais sobre conhecimento médico e biológico sobrecarregam os pesquisadores dessas áreas. Eles têm a difícil tarefa de gerenciar terabytes de informação, com conteúdos semânticos, em grandes repositórios de dados que usam tecnologias de gerenciamento cada vez mais sofisticadas. Pesquisadores da área biomédica podem

²<http://www.daml.org/rules/proposal/>

³<http://www.ruleml.org/>

⁴<http://www.w3.org/Submission/SWRL/>

trabalhar com uma enorme quantidade de dados, que não pode ser interpretada sem o auxílio de computadores (Serique, 2012).

Os pesquisadores também devem agregar e integrar toda estas informação. Para tanto, eles precisam de ferramentas que permitam a descoberta de conhecimento nesse paradigma rico em dados. Portanto, eles começaram a utilizar ontologias para descrever a estrutura de seus domínios complexos e relacionar seus dados para compartilhar e interoperar informações Biomédicas(Rubin et al., 2008).

Muitas ciências, 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 que este trabalho se foca, radiologia e oncologia, as duas mais importantes ontologias usadas são a FMA e o RadLex.

3.3.1 Foundational Model of Anatomy – FMA

Uma bio-ontologia importante no escopo deste estudo é a Foundational Model of Anatomy (FMA). De modo geral, a anatomia é o estudo das estruturas físicas dos organismos biológicos. Descrever e capturar de forma computável as terminologias e ontologias que envolvem entidades anatômicas é uma tarefa não trivial. A falta de uma representação generalizada da anatomia leva os desenvolvedores de terminologias e ontologias, tanto da área de medicina clínica quanto de biologia, a representar a anatomia a partir dos seus próprios pontos de vista. Isso gera um obstáculo para correlacionar os termos anatômicos entre si, devido a heterogeneidade das representações anatômicas(Serique, 2012).

A estrutura ontológica da FMA fornece uma modelo de inferência, permitindo que ferramentas computacionais possam inferir informações sobre dados anatômicos (Rosse und Mejino, 2008). Especificamente, a FMA representa um corpo coerente de conhecimentos declarativos sobre a anatomia humana (Rosse und Mejino, 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*.

3.3.2 RadLex

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

unificada para organizar e recuperar esses dados.

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

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

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

O RadLex está disponível para download no *National Center for Biomedical Computing's Biportal site* ⁵.

3.3.3 Annotation Imaging Markup

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

O projeto Annotation Imaging Markup (AIM), que atualmente encontra-se em sua quarta versão, trata de um modelo de informação que oferece uma infra-estrutura de apoio para a criação de uma coleção de anotações médicas de imagens (que podem ser lidas por máquinas). Ele é resultado do cumprimento de novas exigências da comunidade de imagens médicas e vem conseguindo atingir os requisitos estabelecidos pela *caBIG In Vivo Imaging Workspace* (Rubin et al., 2008). O projeto é centrado no modelo AIM de anotação e marcação de imagens DICOM. As anotações e marcações de imagens

⁵<http://biportal.bioontology.org/ontologies/40885>

são consideradas como um objeto de informação da imagem, porém não estão ligadas fisicamente as mesmas. Esse modelo de informação, no entanto, é compatível com outros formatos de imagens além do DICOM. O modelo AIM não cria nenhuma nova ontologia para descrever anotações, ele usa termos do RadLex para descrever, lesões e junta a eles informações numéricas, como volume ou diâmetro.

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

Marcações podem ser usadas para descrever uma informação textual e as regiões de interesse visual. 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.

3.4 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 (reason) anotações de imagens médicas. O AIM vem sendo desenvolvido e elaborado, tanto para ajudar na estruturação dos relatórios radiológicos, principalmente imagens, como também para ser entendido por máquinas. Esses avanços vem contribuindo no entendimento, interoperabilidade e compartilhamento de conhecimento biomédico na área de radiologia.

4 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.

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.

4.1 Planeamento e Execução do Mapeamento Sistemático

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

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

O método utilizado para o levantamento de fontes primárias compreendeu a realização de buscas em bibliotecas digitais, como ACM Digital Library¹, IEEE Xplore Digital Library², SciELO (Scientific Electronic Library Online)³, DBLP (Computer Science Bibliography)⁴, BDBComp (Biblioteca Digital Brasileira de Computação)⁵, Google Acadêmico (Scholars), Springer⁶, PubMed (US National Library of Medicine National Institutes of Health)⁷.

¹<http://dl.acm.org/>

²<http://ieeexplore.ieee.org/>

³<http://www.scielo.org/>

⁴<http://dblp.uni-trier.de/>

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

⁶<http://www.springer.com/gp/>

⁷<http://www.ncbi.nlm.nih.gov/pubmed>

4 Trabalhos Relacionados

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 2014. 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, 11 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.

4.2 Apresentação e Análise dos Resultados Obtidos no Mapeamento

Nesta seção foi realizada a catalogação bibliográfica dos 11 estudos selecionados no mapeamento sistemático, a qual é apresentada na Tabela 4.1. 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).

4.2 Apresentação e Análise dos Resultados Obtidos no Mapeamento

Tabela 4.1: Comparativo entre os trabalhos relacionados.

Title (year)	Cancer Type	Data Input	Knowledge source – Grading/staging method	Results
Zillner und Sonntag (2012)	Lymphoma	Automatic tagging of medical image using OWL ontologies	Ontology-based representation of AAS ⁸ – Staging (OWL-DL reasoning)	No
Zillner (2009)	Lymphoma	Semantic annotation of patient records by tool support (mapped to OWL)	Ontology-based representation of AAS – Staging (OWL-DL reasoning)	No
Tutac et al. (2008)	Breast	Image recognition (pathology images) - automatic tagging	Ontology-based representation of NGS ⁹ – Grading (OWL-DL, SWRL reasoning)	Yes
Tutac et al. (2010)	Breast	Automatic tagging (pathology images)	Ontology-based Spatial representation of NGS – Grading (OWL-DL, SWRL reasoning)	No
Meriem et al. (2012)	Breast	WFML annotations mapped to OWL (pathology images)	Ontology-based representation of NGS – Grading (OWL-DL, SWRL reasoning)	Yes
Levy et al. (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

⁸Ann Arbor Staging

⁹Nottingham Grading System

Tabela 4.1: Comparativo entre os trabalhos relacionados.

Title (year)	Cancer Type	Data Input	Knowledge source – Grading/staging method	Results
Dameron et al. (2006a)	Lung	Simulated patient conditions data represented using classes	Ontology-based representation of TNM ¹⁰ – Staging (OWL-DL reasoning)	Yes
Oberkampff et al. (2012)	Lymphoma	Semantic annotations (mapped to OWL)	No grading system	Yes
Kurtz et al. (2014)		Manual semantic annotations from ePAD	No representation of Staging System – Classification (Ontological similarity between terms)	Yes
Racoceanu und Capron (2014)	Breast	WFML annotations mapped to OWL (pathology images)	Ontology-based representation of NGS – Grading (OWL, N3Logic reasoning)	No

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

- Dameron et al. (2006a) mostram que a classificação de tumores pode ser executada automaticamente usando OWL-DL. Ele se centra na classificação dos tumores de pulmão. Os autores usam a ontologia Foundational Model of Anatomy (FMA) para descrever entidades anatômicas a fim de representar alguns critérios da *TNM Classification of Malignant Tumors* (TNM). Além disso, os autores destacam algumas das limitações da Descrição baseada em Lógica (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. O sistema de referência de

¹⁰TNM Classification of Malignant Tumors Staging

4.2 Apresentação e Análise dos Resultados Obtidos no Mapeamento

classificação utilizado foi o da *World Health Organization* (WHO). Em seguida, *reasoning* baseado em classes foi utilizado, a partir de relatórios de patologia com descrições histológicas de pacientes (usando essa ontologia). Mas, em comparação com o projeto proposto, esses relatórios de pacientes foram preenchidos manualmente, os autores não apresentam resultados explícitos usando dados reais.

- Em Levy et al. (2009), uma metodologia e as ferramentas foram desenvolvidas para transformar o modelo de informações *Annotation and Image Markup* (AIM) em instâncias OWL. Em seguida, um mecanismo de *reasoning* baseado em ontologias e SWRL foi desenvolvido para trabalhar com essas instâncias a fim de classificar e calcular a carga tumoral. Embora Levy et al. (2009) utilizem AIM como modelo de informação, eles não fazem *staging* de câncer.
- 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*), como, por exemplo, *Cells Cluster* e *Lumina*, para conceitos de visão computacional: *Cell Cluster* \rightarrow *union of cells* e *Lumina* \rightarrow *white compact segments*). Os autores criaram um método baseado em regras para gerar automaticamente o *NGS* e modelar todos esses conceitos em OWL usando o Protégé (uma ferramenta para a edição de ontologias), com a finalidade de definir regras que são utilizadas para a indexação semântica das imagens de pacientes com câncer da mama. Embora a proposta de Tutac et al. (2008) seja desenvolvido no campo da análise de imagens histológicas para a indexação e não exatamente para uma classificação de pacientes, a abordagem apresentada é importante no desenvolvimento desta proposta, já que descreve claramente as regras utilizadas para a classificação do câncer com base num sistema de classificação, como o *NGS*.
- Em Zillner (2009) uma ontologia OWL-DL foi projetada para realizar o *staging* automático de pacientes com linfoma usando anotações semânticas. Essas anotações são armazenadas numa ontologia pela ferramenta de anotação que faz parte de projeto *MEDICO-ontology*. Além disso, para a classificação dos pacientes com linfoma, os autores capturaram as regras do sistema de classificação *Staging Ann Arbor* para linfoma numa ontologia. Mas, em contraste com a nossa proposta, o trabalho de Zillner (2009) só melhorou as anotações semânticas usadas para descrever imagens médicas, através da representação do sistema *Staging Ann Arbor* numa ontologia. A classificação dos pacientes, a partir dessas anotações semânticas me-

4 Trabalhos Relacionados

lhoradas (com o conhecimento clínico adicional), não é descrita em Zillner (2009). Além disso, não são mostrados, de forma explícita, os resultados da avaliação com dados reais.

- O principal objetivo em Tutac et al. (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 et al. (2008) não faz nenhuma avaliação do modelo com dados reais de pacientes com câncer de mama.
- O trabalho proposto em Zillner und 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. Para verificar ou falsificar as posições de linfoma, conhecimento baseado em ontologia foi utilizado. Esse processo é conhecido como *spatial-anatomical reasoning* Manuel et al. (????). 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*. Em contraste com a nossa proposta, as anotações que descrevem as imagens médicas (anotações semânticas) são realizadas automaticamente por algoritmos de processamento de imagem e não manualmente. Além disso, nesse trabalho foram encontradas inconsistências através da realização de uma avaliação clínica a partir de registros reais de pacientes.
- O trabalho em Meriem et al. (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)* e é necessária para realizar o *grading* do câncer da mama. Eles desenvolveram uma ontologia chamada *Breast Cancer Ontology (BCO)* com base em OWL-DL e SWRL. Além disso, a Wide Field Markup Language (WFML), que é específica para o campo de histopatologia, foi usada

4.2 Apresentação e Análise dos Resultados Obtidos no Mapeamento

para descrever imagens histopatológicas usando anotações. Por fim, a linguagem de consulta SPARQL é usada para consultar o sistema. Embora o trabalho de Meriem et al. (2012) utilize imagens histológicas, apenas são descritas algumas regras do *reasoning* que são necessárias para o *grading* do câncer de mama. Além disso, os autores mencionam apenas poucas possíveis aplicações para resolver problemas de ambiguidade na análise de imagens histológicas. Por exemplo, o uso de *reasoning* baseado em relações espaciais das anotações na imagem. Finalmente, são usados algoritmos de extração de características para descrever as imagens histológicas usando anotações.

- 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.
- Em Kurtz et al. (2014) e Seifert et al. (????) 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. (????) 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 und Capron (2014) 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 Racocanu und Capron (2014), 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.

4.3 Considerações Finais

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

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

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

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

5 Plano de Trabalho

Nesta capítulo seção será descrita a metodologia a ser seguida para o desenvolvimento deste trabalho. Ela será dividida nas seguintes etapas:

- Desenvolvimento de um mecanismo para transformar as anotações semânticas, a partir dos documentos XML gerados pelo AIM e ePAD, para instâncias no modelo OWL.
- Desenvolver (ou estender) uma ontologia com foco no grau de câncer com base no critério de classificação *TNM Classification of Malignant Tumors* (TNM).
- Integração de conhecimentos por meio de *reasoning*. O processo de *reasoning* permite a classificação de um paciente, integrando o conhecimento capturado pelas ontologias médicas. O modelo ontológico vai capturar o *cancer staging* individual de cada paciente.
- Informações *de staging* dos pacientes em banco de dados, pode ser consultado por meio de SPARQL
- Implementação de interfaces em parceria com o Dept. of Radiology, Stanford University.
- Avaliação dos resultados obtidos (stages) com radiologistas/oncologistas do Dept. of Radiology, Stanford University.

5.0.1 Arquitetura proposta para desenvolvimento do trabalho

Para o uso na classificação de pacientes com câncer a partir de imagens(*staging*), é necessário em primeiro lugar ter uma ontologia (ou ontologias) que descreve os padrões para o *staging* de câncer, como, por exemplo, o padrão *TNM*. Nesse sentido. Essa é uma ontologia ainda por desenvolver e contribuiremos em sua construção neste projeto. Por outro lado é necessário transformar as anotações da imagem que estão no formato AIM obtidas a partir do ePAD, para um formato como OWL. O formato AIM-OWL permitirá

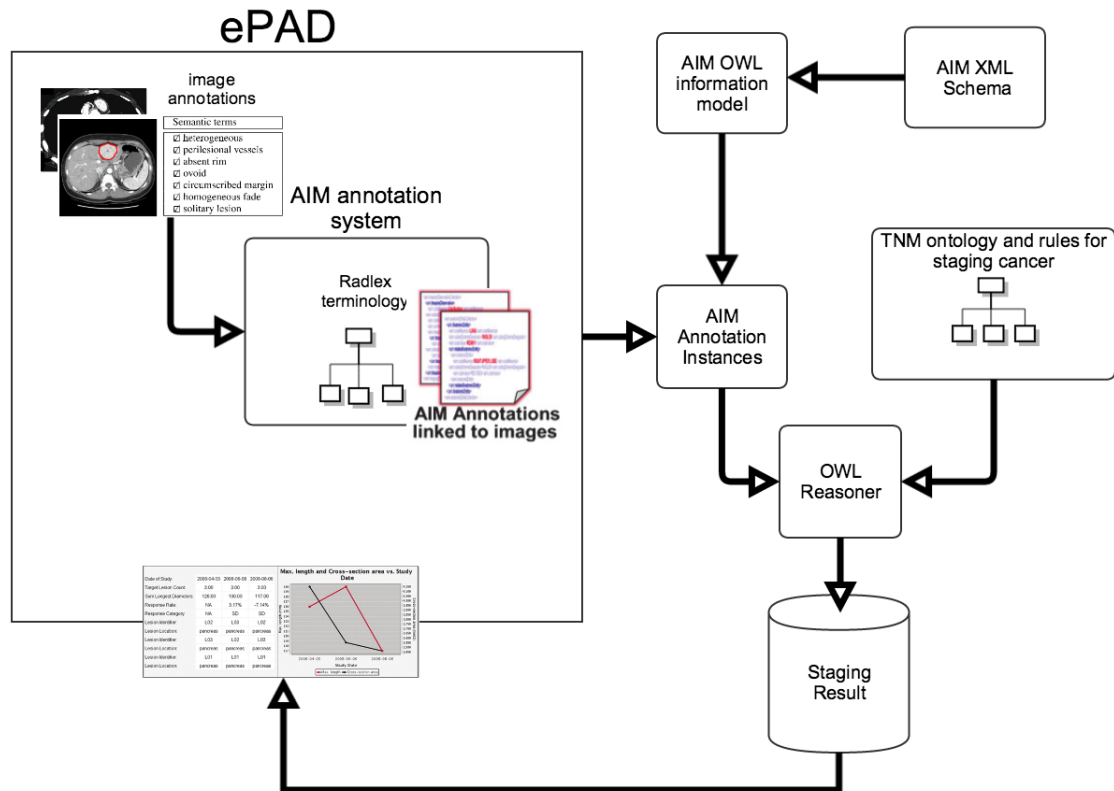


Figura 5.1: Arquitetura proposta.

finalmente fazer *reasoning* pelas regras que representam o padrão de TNM. Nos também vamos contribuir na construção dessas regras.

Um esboço da proposta atual para classificação, no domínio de câncer, é apresentado na Figura 5.1.

5.1 Atividades

A etapa inicial deste projeto foi o estudo detalhado do problema, sua pesquisa na literatura e a análise das tecnologias, padrões e ferramentas já usados no ePAD. O ePAD tem uma arquitetura voltada a plugins e o sistema proposto neste projeto pode ser desenvolvido como um plugin. Isso vai reduzir em muito o tempo de desenvolvimento já que muitas funcionalidades necessárias já estão implementadas no ePAD.

Um dos problemas centrais aqui será a representação formal do conhecimento necessário para classificar os estágios de câncer dos pacientes (*cancer staging*). Procuramos desenvolver uma solução que seja prática e confortável para os médicos (integrada ao ePAD) mas que, ao mesmo tempo, seja factível usando as tecnologias web que temos a disposição. Para isso contaremos com o feedback dos usuários do Rubin Lab (nossos parceiros na Stanford University).

Para o cumprimento dos objetivos estabelecidos e o alcance dos resultados esperados, são previstas as seguintes atividades:

1. Obtenção de créditos referente as disciplinas do programa de mestrado.
2. Exame de proficiência na língua portuguesa.
3. Levantamento bibliográfico sobre a área de pesquisa.
4. Levantamento dos novos requisitos para o ePad com usuários.
5. Análise dos sistemas de estadiamento de câncer.
6. Análise e planejamento de arquitetura do sistema e sua interação com os usuários.
7. Exame de proficiência na língua inglesa.
8. Qualificação: redação da monografia de qualificação, incluindo detalhamento do projeto.
9. Exame de Qualificação (apresentação da proposta à Comissão Examinadora).
10. Redação de artigos que promovam a divulgação dos resultados alcançados e a apresentação das contribuições científicas e tecnológicas..
11. Avaliação final do sistema e ajustes no projeto proposto
12. Redação da dissertação.
13. Redação e submissão de artigos com os resultados obtidos.
14. Defesa.

Essas atividades serão cumpridas de acordo com o cronograma da Tabela Table 5.1 on page 36.

5 Plano de Trabalho

Ativ.	2014				2015				2016
	1 Tri	2 Tri	3 Tri	4 Tri	1 Tri	2 Tri	3 Tri	4 Tri	1 Tri
A1					
A2						
A3						
A4							
A5							
A6							
A7					.				
A8					...				
A9							
A10					..				
A11						
A12							
A13							
A14						

Tabela 5.1: Cronograma do projeto.

5.1.1 Atividades Concluídas até o Momento

Quanto a metodologia proposta para desenvolvimento do nosso trabalho, os passos 1 ao 7 já foram concluídos. Além disso, a redação de artigos, já vem sendo realizada desde o início do 2015.

5.1.2 Resultados esperados

- Ao término desta pesquisa, o principal resultado será mostrar que, com a utilização de técnicas de *reasoning* e ontologias podemos ajudar os medicos a obter uma maior taxa de precisão na interpretação das lesões de câncer em imagens.
- Modelar o conhecimento médico relacionados com TNM em regras de raciocínio para classificação semântica.
- Este projeto também visa a escrita e publicação de artigos científicos nas áreas de Bioinformática, Ontologia, Web Semântica para permitir a disseminação dos resultados obtidos e da base de conhecimentos criada.

6 Referências

Alex A.T. Bui, Ricky K. Taira (2010): Medical Imaging Informatics.

Arnold, Hodder (2012): Introduction to medical imaging.

Bechhofer, Sean; Frank van Harmelen; Jim Hendler; Ian Horrocks; Deborah L. McGuinness; Peter F. Patel-Schneider und Lynn Andrea Stein (February 2004): *OWL Web Ontology Language Reference*. Techn. Ber., W3C, <http://www.w3.org/TR/owl-ref/>.

Berners-Lee, Tim; James Hendler und Ora Lassila (Mai 2001): The Semantic Web. *Scientific American*, 284(5):34–43.

Bulu, Hakan und Daniel L Rubin (2015): Java Application Programming Interface (API) for Annotation Imaging Markup (AIM). <http://web.stanford.edu/group/qil/wikiupload/4/41/AimApiManuscript.pdf>. [Online; accessed 02-March-2015].

Cardoso, D. S.; J. K. Serique; K. F. Amanqui; L. J. Campos dos Santos und A. D. Moreira (Jun 2014): A Gazetteer for Biodiversity Data as a Linked Open Data Solution. Web2Touch 2014 - Modelling the Collaborative Web Knowledge.

Connor, Martin O; Holger Knublauch; Samson Tu; Benjamin Grosof; Mike Dean; William Grosso; Mark Musen und Stanford Medical Informatics (2005): Supporting Rule System Interoperability on the Semantic Web with SWRL. S. 974–986.

Dameron, Olivier; Élodie Roques; Daniel Rubin; Gwenaëlle Marquet und Anita Burgun (2006a): Grading lung tumors using OWL-DL based reasoning. In: *9th International Protégé Conference - Presentation Abstracts*. S. 69–72.

Dameron, Olivier; Élodie Roques; Daniel Rubin; Gwenaëlle Marquet und Anita Burgun (2006b): Grading lung tumors using OWL-DL based reasoning Methods. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.137.5378&rep=rep1&type=pdf>. [Online; accessed 02-March-2015].

- Dasmahapatra, S.; P. Lewis und N. Shadbolt (2003): Ontology-based medical image annotation with description logics. *Proceedings. 15th IEEE International Conference on Tools with Artificial Intelligence*, S. 77–82.
- Depeursinge, Adrien; Camille Kurtz; Christopher Beaulieu; Sandy Napel und Daniel Rubin (2014): From Radiological Image Data : Preliminary Results With Liver Lesions in CT. 33(8):1669–1676.
- Gimenez, Francisco; Jiajing Xu; Yi Liu; Tiffany Ting Liu; Christopher F. Beaulieu; Daniel L. Rubin und Sandy Napel (Juli 2011): On the Feasibility of Predicting Radiological Observations from Computational Imaging Features of Liver Lesions in CT Scans. *2011 IEEE First International Conference on Healthcare Informatics, Imaging and Systems Biology*, S. 346–350.
- Grau, Bernardo Cuenca; Ian Horrocks; Boris Motik; Bijan Parsia; Peter Patel-Schneider und Ulrike Sattler (2008): OWL2: The Next Step for OWL. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4).
- Gruber, Thomas R. (Juni 1993): A Translation Approach to Portable Ontology Specifications. *Knowl. Acquis.*, 5(2):199–220.
- Guarino, Nicola (Dez. 1995): Formal Ontology, Conceptual Analysis and Knowledge Representation. *Int. J. Hum.-Comput. Stud.*, 43(5-6):625–640.
- Horrocks, Ian; Peter F. Patel-Schneider und Frank Van Harmelen (2003): From SHIQ and RDF to OWL: The Making of a Web Ontology Language. *Journal of Web Semantics*, 1:2003.
- Huang, H.K. (2010b): *PACS and Imaging Informatics: Basic Principles and Applications*. Wiley.
- Imaging, Medical (2007): Medical Imaging in Cancer Care : CHARTING How Innovation in.
- Korenblum, Daniel; Daniel Rubin; Sandy Napel; Cesar Rodriguez und Chris Beaulieu (Aug. 2011): Managing biomedical image metadata for search and retrieval of similar images. *Journal of digital imaging*, 24(4):739–48.
- Kundu, Sanjoy; Maxim Itkin; Debra a. Gervais; Venkataramu N. Krishnamurthy; Michael J. Wallace; John F. Cardella; Daniel L. Rubin und Curtis P. Langlotz (2009): 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*, 20(November 2006):433–435.
- Kurtz, Camille; Adrien Depeursinge; Sandy Napel; Christopher F Beaulieu und Daniel L Rubin (Juli 2014): On combining image-based and ontological semantic dissimilarities for medical image retrieval applications. *Medical image analysis*, 18(7):1082–1100.
- Langlotz, Curtis P. (2006): RadLex: A New Method for Indexing Online Educational Materials. *RadioGraphics*, 26(6):1595–1597. PMID: 17102038.
- Levy, Mia a; Martin J O'Connor und Daniel L Rubin (Jan. 2009): Semantic reasoning with image annotations for tumor assessment. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, 2009:359–63.
- Ma, Hao; Jianke Zhu; Michael Rung-tsong Lyu; Irwin King und Senior Member (2010): Bridging the Semantic Gap Between Image Contents and Tags. (August):462–473.
- Manuel, M; Sven Regel und Michael Sintek (2009): RadSem : Semantic Annotation and Retrieval for Medical Images.
- Manuel, M; Daniel Sonntag und Patrick Ernst (????): A Spatio-Anatomical Medical Ontology and Automatic Plausibility Checks.
- Marquet, Gwenaëlle; Olivier Dameron; Stephan Saikali; Jean Mosser und Anita Burgun (Jan. 2007): Grading glioma tumors using OWL-DL and NCI Thesaurus. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, S. 508–12.
- Mccowan, Iain; Darren Moore und Mary-jane Fry (2000): Automated Cancer Stage Classification from Free-text Histology Reports.
- of Medicine, Institute (2003): *Key Capabilities of an Electronic Health Record System: Letter Report*. The National Academies Press.
- Meriem, Benmarouf; Tlili Yamina und A Digitized Pathology (2012): Interpretation breast cancer imaging by using ontology. (I):1–6.
- Nagarajan, Meenakshi (2009): semantic annotations in web services.
- Napel, Sandy A; Christopher F Beaulieu; Cesar Rodriguez und Daniel L Rubin (2010): Automated Retrieval of CT Images of Liver Lesions on the Basis of Image Similarity : Method and Preliminary Results 1 Methods : Results :. (July):243–252.

- Noy, Natalya F. und Deborah L. McGuinness (2001): Ontology Development 101: A Guide to Creating Your First Ontology. Online.
- Oberkampff, Heiner; Sonja Zillner; Bernhard Bauer und Matthias Hammon (2012): Interpreting Patient Data using Medical Background Knowledge. S. 2–6.
- Philbin, James; Fred Prior und Paul Nagy (2011): Will the Next Generation of PACS Be Sitting on a Cloud? *Journal of Digital Imaging*, 24(2):179–183.
- Racoceanu, D. und F. Capron (Okt. 2014): Towards semantic-driven high-content image analysis: An operational instantiation for mitosis detection in digital histopathology. *Computerized Medical Imaging and Graphics*.
- Rosse, Cornelius und Jr Mejino, Jos  L.V. (2008): The Foundational Model of Anatomy Ontology. In: *Anatomy Ontologies for Bioinformatics*, Hg. Albert Burger; Duncan Davidson und Richard Baldock, Springer London, Bd. 6 von *Computational Biology*, S. 59–117.
- Rubin, Daniel L. (2011): Finding the Meaning in Images: Annotation and Image Markup. *Philosophy, Psychiatry, & Psychology*, 18(4):311–318.
- Rubin, Daniel L; Cesar Rodriguez; Priyanka Shah und Chris Beaulieu (Jan. 2008): iPad: Semantic annotation and markup of radiological images. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, S. 626–30.
- Rubin, Daniel L.; Debra Willrett; Martin J. O’Connor; Cleber Hage; Camille Kurtz und Dilvan a. Moreira (Febr. 2014): Automated Tracking of Quantitative Assessments of Tumor Burden in Clinical Trials. *Translational Oncology*, 7(1):23–35.
- Seifert, Sascha; Adrian Barbu; S. Kevin Zhou; David Liu; Johannes Feulner; Martin Huber; Michael Suehling; Alexander Cavallaro und Dorin Comaniciu (Febr. 2009): <title>Hierarchical parsing and semantic navigation of full body CT data</title>. S. 725902–725902–8.
- Seifert, Sascha; Michael Kelm; Manuel Moeller; Saikat Mukherjee; Alexander Cavallaro; Martin Huber und Dorin Comaniciu (????): Semantic Annotation of Medical Images. S. 3–10.
- Serique, Kleberson Junio Amaral (2012): *Anotação de imagens radiológicas usando a web semântica para colaboração científica e clínica*. Mestrado.

- Smith, Barry; Michael Ashburner; Cornelius Rosse; Jonathan Bard; William Bug; Werner Ceusters; Louis J Goldberg; Karen Eilbeck; Amelia Ireland; Christopher J Mungall; Neocles Leontis; Philippe Rocca-Serra; Alan Ruttenberg; Susanna-Assunta Sansone; Richard H Scheuermann; Nigam Shah; Patricia L Whetzel und Suzanna Lewis (nov 2007): The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech*, 25(11):1251–1255.
- Smith, David A. und Gideon S. Mann (2003): Bootstrapping Toponym Classifiers. In: *Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References - Volume 1*. Association for Computational Linguistics, Stroudsburg, PA, USA, HLT-NAACL-GEOREF '03, S. 45–49.
- Tutac, a. E.; V. I. Cretu und D. Racoceanu (2010): Spatial representation and reasoning in breast cancer grading ontology. *2010 International Joint Conference on Computational Cybernetics and Technical Informatics*, S. 89–94.
- Tutac, Adina Eunice; Daniel Racoceanu; Thomas Putti; Wei Xiong; Wee-Kheng Leow und Vladimir Cretu (Mai 2008): Knowledge-Guided Semantic Indexing of Breast Cancer Histopathology Images. *2008 International Conference on BioMedical Engineering and Informatics*, S. 107–112.
- Wennerberg, Pinar; Klaus Schulz und Paul Buitelaar (Febr. 2011): Ontology modularization to improve semantic medical image annotation. *Journal of biomedical informatics*, 44(1):155–62.
- Ziegler, Sibylle I. (2005): Positron emission tomography: Principles, technology, and recent developments. *Nuclear Physics A*, 752:679–687.
- Zillner, Sonja (2009): Towards the Ontology-based Classification of Lymphoma Patients using Semantic Image Annotations. In: *SWAT4LS'09*. S. –1–1.
- Zillner, Sonja (2010): Reasoning-Based Patient Classification for Enhanced Medical Image Annotation. S. 243–257.
- Zillner, Sonja und Claudia Bretschneider (2013): Grammar-Based Lexicon Extension for Aligning German Radiology Text and Images. (September):105–112.
- Zillner, Sonja und Daniel Sonntag (März 2012): Image metadata reasoning for improved clinical decision support. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 1(1-2):37–46.