Classificação automática de tumores cancerosos

usando anotações em imagens e ontologias

Dezembro, 2016

Aluno: Edson F. Luque Mamani, Orientador: Dilvan de Abreu Moreira {edluquem,dilvan}@icmc.usp.br,

Carlos, Brazil









Roteiro

- Introdução
 - Cenário motivador
 - Objetivos
- Web Semântica
 - AIM
 - Ferramentas de Anotação(ePAD)
- u TNM
- Metodologia
- Experimentos
- Conclusões



Introdução

- Imagens Radiológicas
 - Descoberta de evidencias relacionadas a doenças.
 - Fonte de grande quantidade de informações (implícitas).
 - Informação sem estrutura padronizada.
 - Solução → ter conteúdo semântico a essa informação.



Introdução

- Modelos de Informação
 - AIM (Annotation and Image Markup Project) NCI caBIG.
 - Radlex Terminology.

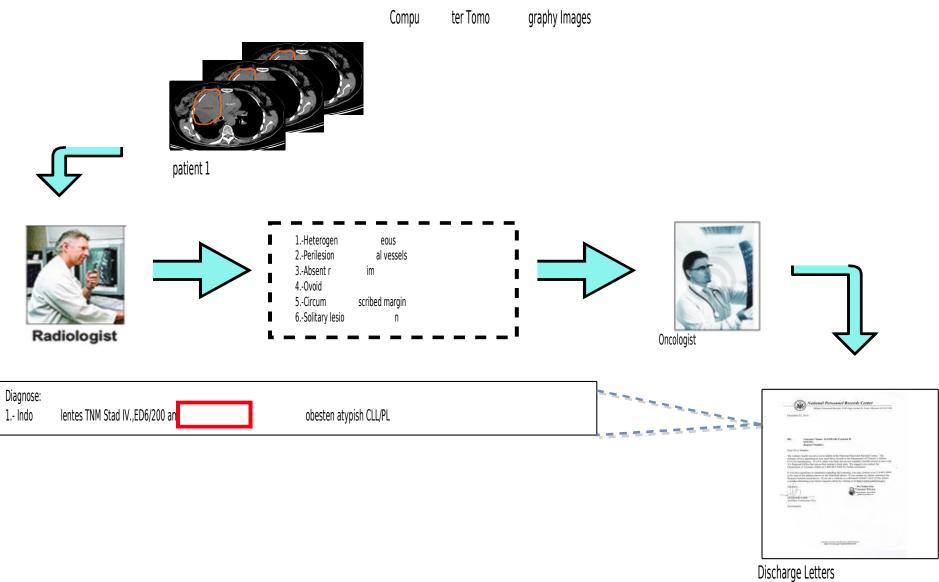


Cenário motivador

- □ Informações sobre o estágio de câncer são cruciais → avaliar o progresso do tratamento.
- Estágio de câncer
 - Descrição, localização, características e possível metástase de tumores cancerosos num paciente.
 - Padrão de classificação → TNM.

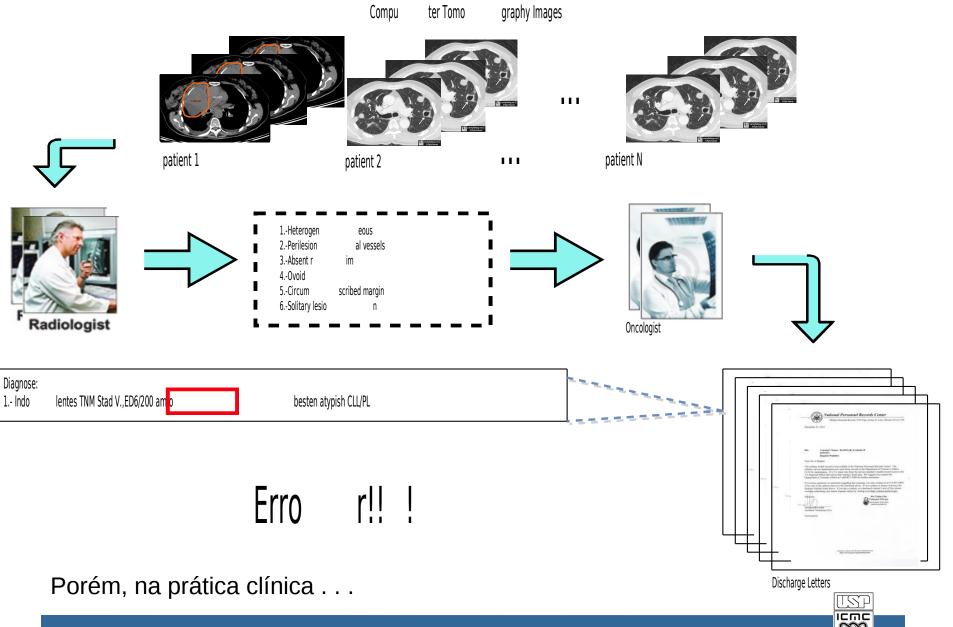


Fluxo de trabalho para avaliar o progresso individual de um paciente com câncer





Fluxo de trabalho para avaliar o progresso individual de um paciente com câncer



Introdução

- Processo de Classificação de câncer
 - Trabalho intensivo.
 - Exige → Precisão na interpretação da lesão cancerosa.
 - Variações nas interpretações.



Introdução

- Anotações sobre imagens
 - Ferramentas de Anotação (ePAD) → não adequadas para o raciocínio (reasoning).
 - AIM formato para transferência e armazenamento de dados.
 - Não há métodos para reasoning semântico usando AIM.



Objetivo

Gerar automaticamente, utilizando tecnologias da web semântica, o Cancer Staging de lesões cancerosas a partir de anotações (padrão AIM) em imagens feitas por radiologistas. A fim de fornecer aos médicos uma segunda opinião sobre o estágio de câncer em que um paciente se encontra.



- □ Quantidade crescente de informações na Web.
- ■Objetivo:
 - Tornar o conteúdo da Web mais facilmente processável por máquinas com técnicas inteligentes (Knowledge Management).



Ontologias

- Termos e relações explícitas para o compartilhamento de informações entre agentes (sistemas).
- Representação formal (RDF-S e OWL).



OWL

- Linguagem do W3C para ontologias na Web:
 - Não é uma linguagem de programação.
 - Não é uma linguagem para conformidade sintática.
 - Não é um banco de dados!.
- Infelizmente → não suporta sentenças de implicação.



SWRL

- Linguagem de regras para Web Semântica.
- Novos conhecimentos → conhecimento em OWL.
- Não impõe restrições!
- Ponto de partida → integração de sistemas de regras com a Web Semântica

```
Regra Pessoa(?x)^temLesion(?x, ?y)^
SWRL localizadoEm(?y, ?z)^figado(?z) \rightarrow TemCancer(?x)
```

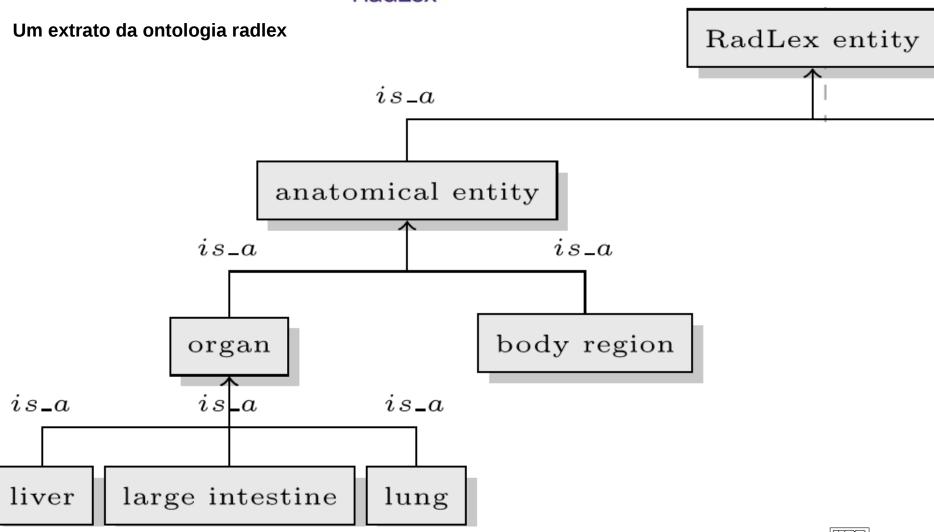
Para a execução de regras → Reasoners





- RadLex
 - Criado pela Radiological Society of North America (RSNA).
 - Termos utilizados em Radiologia.
 - Único e completo.
 - Usado no modelo AIM.
- Exemplos de termos:
 - "heterogeneous", "hypodense", "mass", "enhancing", "right lobe", "liver".
 - "heterogeneous hypodense mass enhancing in right lobe of the liver".





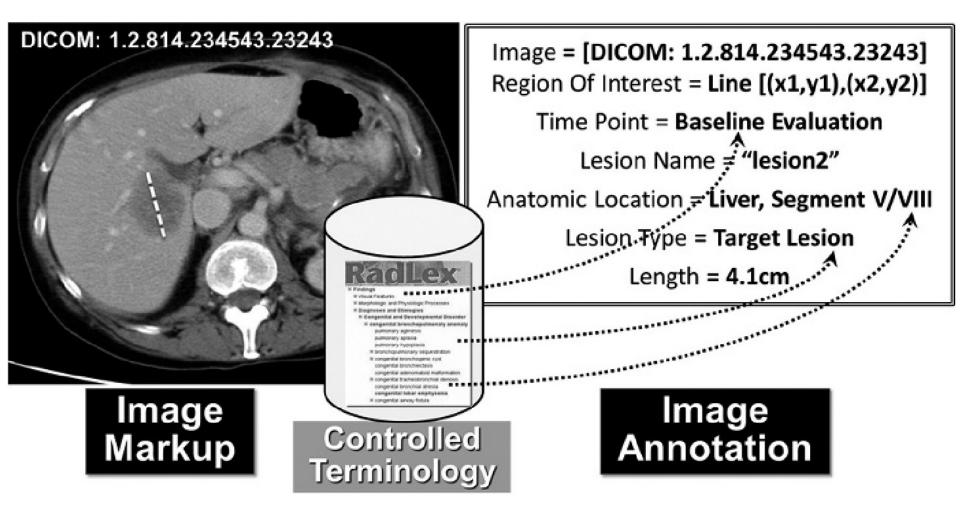
- Annotation and Image Markup Project (AIM)
 - Infra-estrutura de apoio → criação de coleções de anotações médicas.
 - Anotações: Objeto de informação da imagem.
 - não estão ligadas fisicamente as mesmas.



- Annotation and Image Markup Project (AIM)
 - ◆AIM Schema → Classes UML (disponível online).
 - Permite a mineração de dados semânticos: não cria nenhuma nova ontologia.
 - •O modelo passou por 47 iterações com SMEs e três revisões públicas.



Anotações AIM em uma imagem





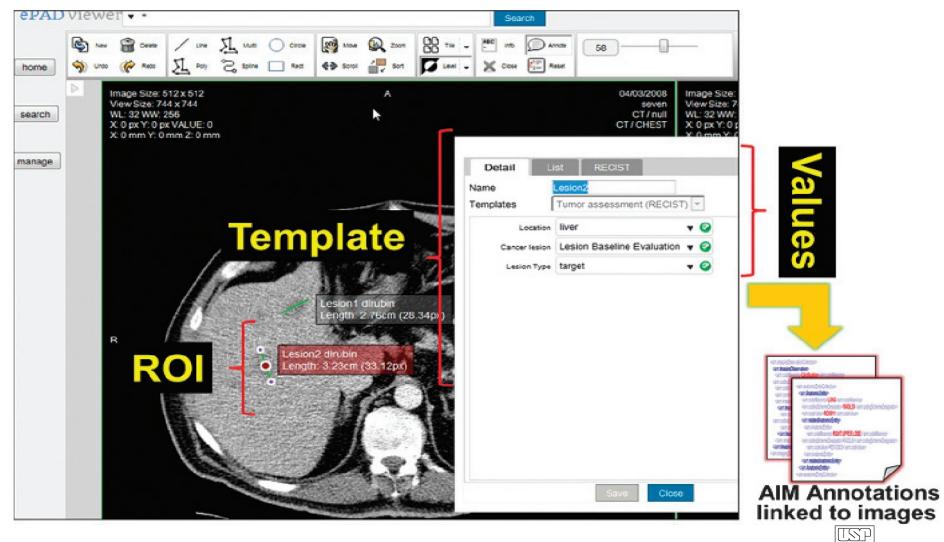
☐ Ferramentas para Anotação (AIM)

	AIM XML annotation	Platform Desktop / Web	3D Markup
AIM ClearCanvas Workstation	yes	Desktop	No
OsiriX-iPad	yes	Desktop	No
3D Slicer	No	Desktop	yes
ePAD 1.0	yes	Web	No



ePAD

Desenvolvido pelo Rubin Lab da Stanford University e o Intermídia-USP



icmc

Estágio do Câncer (TNM Staging System)

- Categoriza a progressão de um câncer no corpo
- Determina o tratamento mais apropriado
- Usado para avaliar resultados de forma mais confiável



Estágio do Câncer (TNM Staging System)

- Padrão internacional
 - **TNM** é um dos sistemas de *staging* de câncer mais amplamente utilizados.
- Baseado em:
 - T → No tamanho/extensão do tumor primário.
 - N →Quantidade de propagação para gânglios linfáticos próximos.
 - M → Presença de metástase.



TNM Staging System

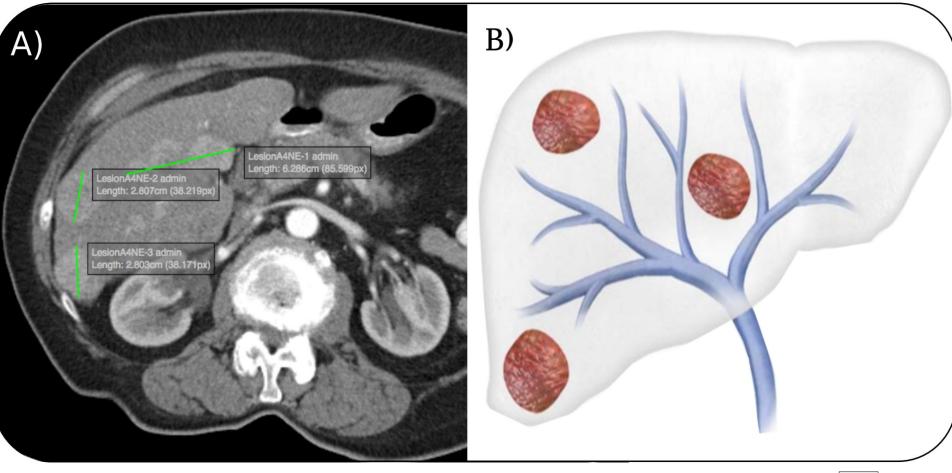
Critérios de T para câncer de fígado

	Tis	Primary Tumor cannot be assessed		
T: Primary Tumor	T0	No evidence of primary tumor		
	T1	Solitary Tumor without vascular invasion		
	T2	Solitary tumor with vascular invasion or multiple tumors, none > 5cm		
	T3a	Multiple tumors > 5cm		
	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 others than gallbladder or with visceral peritoneum		



Exemplo: Critério T3a

- A) Imagem com múltiplos tumores (Anotados utilizando o ePAD)
- B) Critério do T3a: múltiplos tumores com pelo menos um > 5 cm.





TNM Staging System

Critérios de N para câncer de fígado

N: Regional Lymph Nodes **NX** Regional lymph nodes cannot be assessed

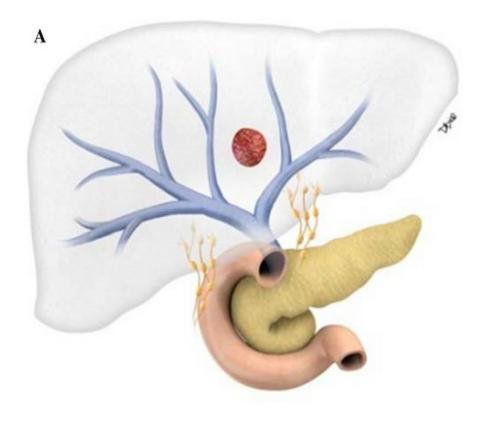
NO The cancer has not spread to lymph nodes

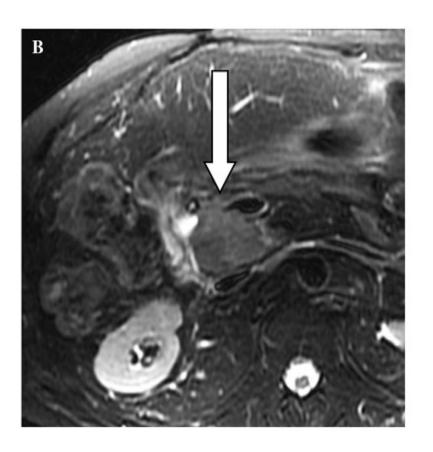




Exemplo: Critério N1

- B) Paciente com um linfonodo afetado(seta), mas sem metástase distante.
- A) Diagrama do critério N1.







TNM Staging System

Definições para os diferentes estágios de tumores

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



Metodologia

- A nossa abordagem é composta por três tarefas:
 - Representação ontológica do AIM.
 - 2) Implementar o reasoning baseado em regras TNM.
 - 3) Representação formal do Staging de câncer.



Metodologia

Representação ontológica do AIM

•Baseado na ontologia fornecida por Hakan Bulu.

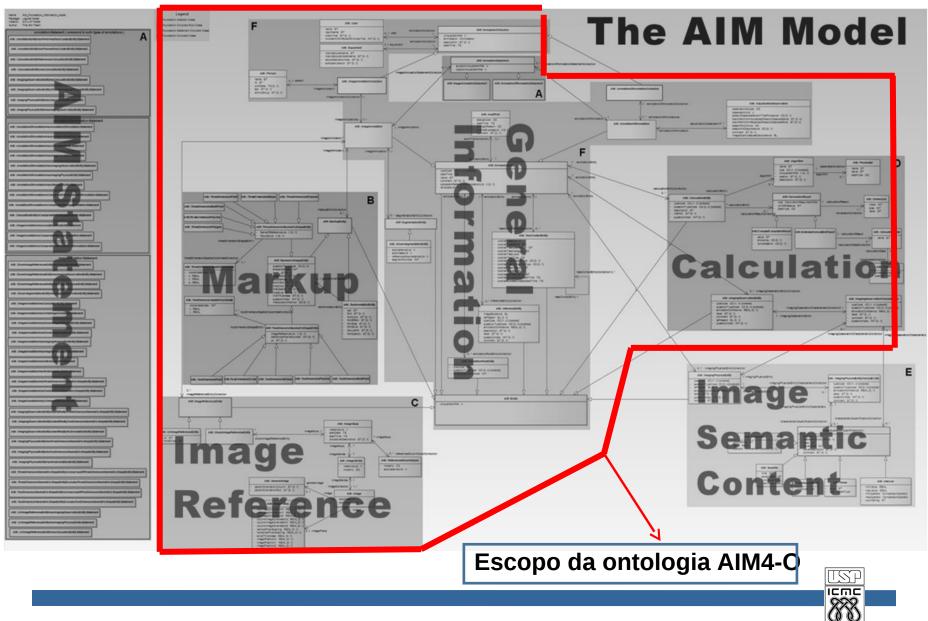
•AIM4-O → 6 classes adicionais.

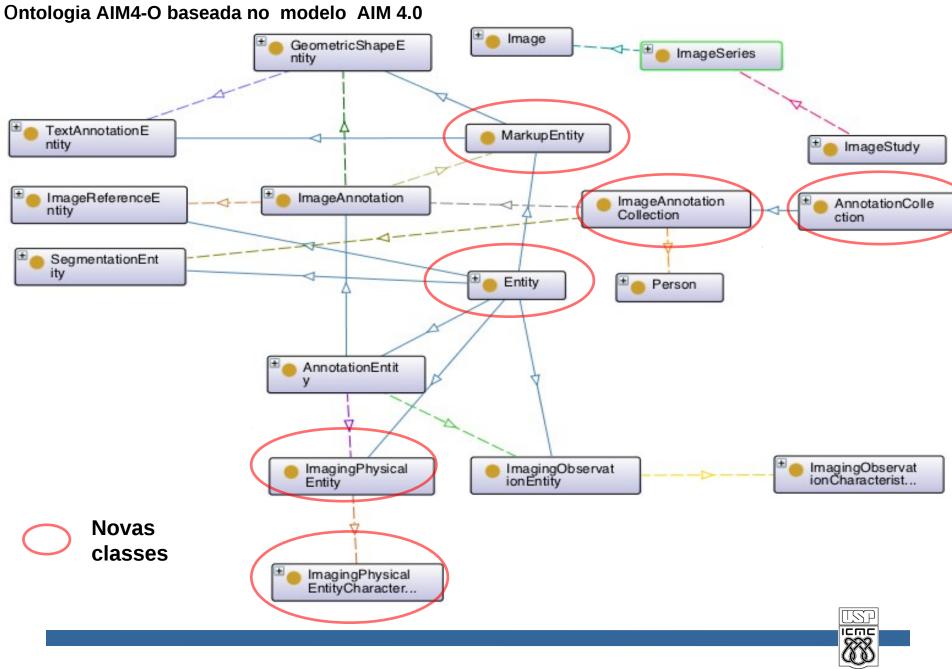


Ontologia de Bulu baseada em AIM 3.0 ImageStudy ImageReference Person GeometricShape Segmentation Image TextAnnotation ImageAnnotation Thing **ImageSeries** ImagingObservat Annotation ionCharacterist... Imaging Observat AnatomicEntity ion AnatomicEntityC haracteristic



AIM 4.0 Uml





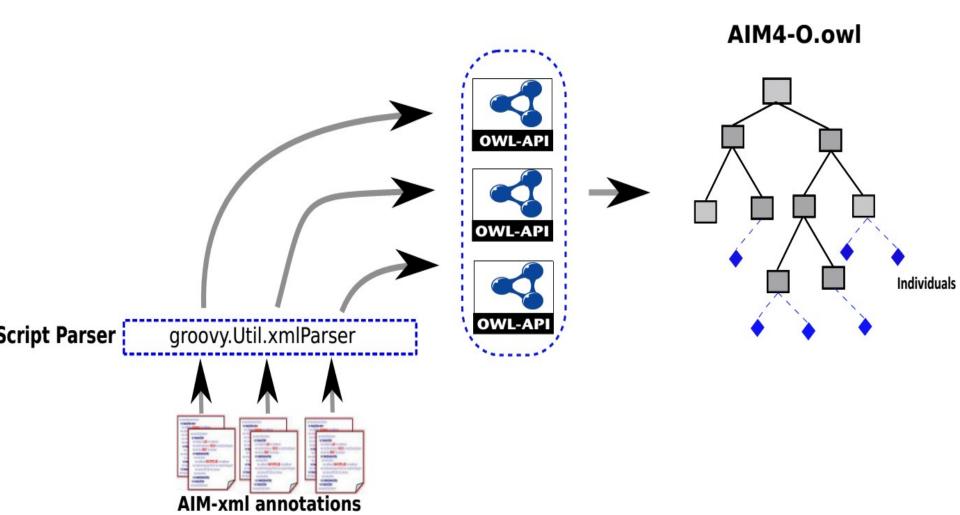
Metodologia

- A nossa abordagem é composta por três tarefas:
 - Representação ontológica do AIM.
 - 2) Implementar o reasoning baseado em regras TNM.



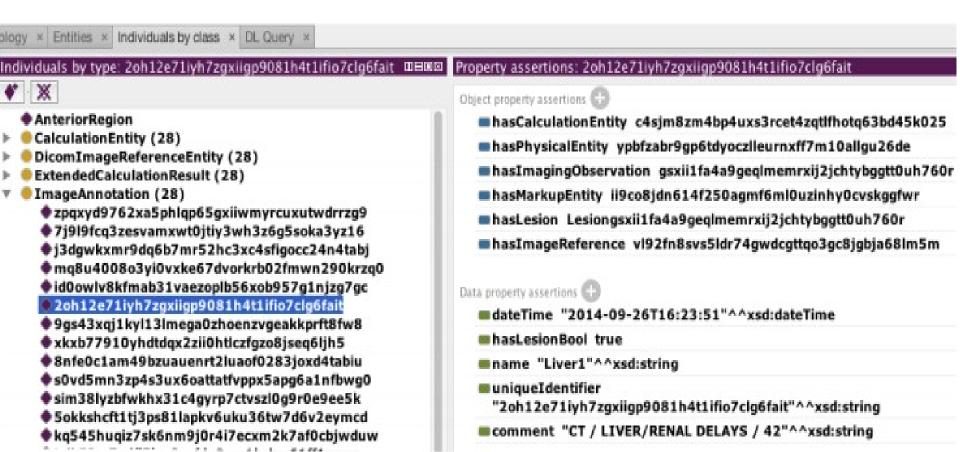
Mecanismo , processar XML \rightarrow OWL

linked to images





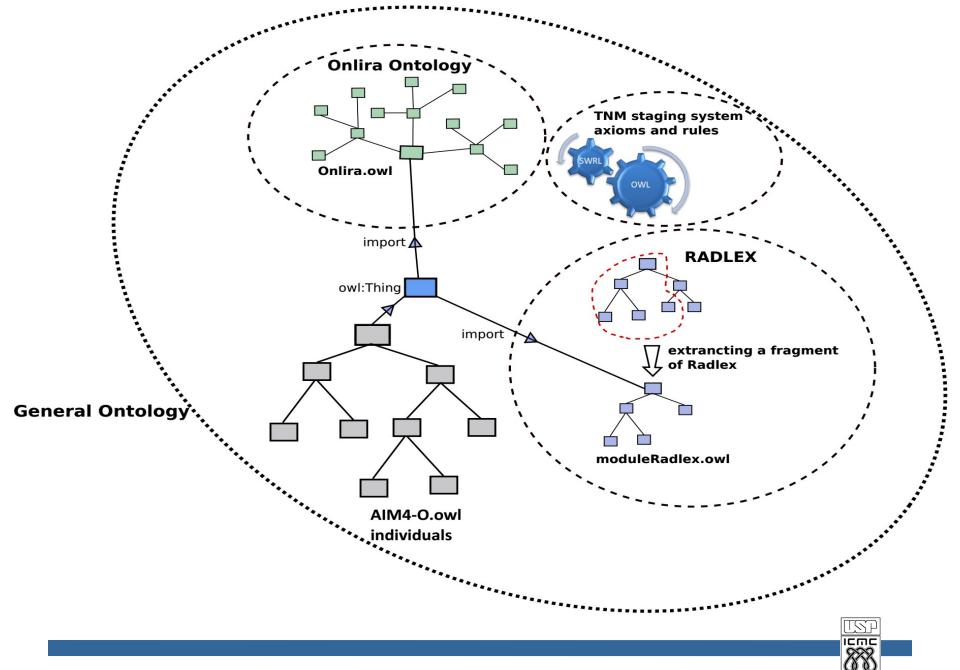
Instâncias de anotações (indivíduos) AIM4-O



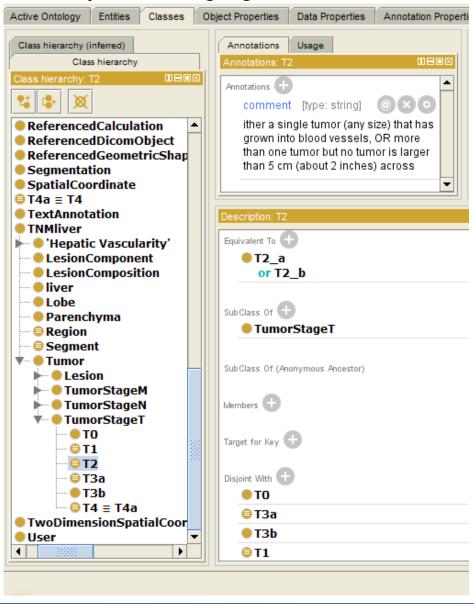


- Implementar o reasoning baseado em regras TNM.
- A ontologia AIM4-O, com os indivíduos.
- Onlira.owl: a topografia do fígado (com base na Ontologia Onlira).
- ModuleRadlex.owl: Uma parte do vocabulário Radlex.
- Os conceitos gerais do TNM (regras SWRL e axiomas).
- → ONTOLOGIA GERAL





criterio T2 representado pela Ontologia geral





- A nossa abordagem é composta por três tarefas:
 - Representação ontológica do AIM.
 - 2) Implementar o reasoning baseado em regras TNM.
 - 3) Representação formal do Staging de Câncer.



- Representação formal do Staging de Câncer.
- TNM apenas usando OWL + SWRL.
- 3 condições refletem → staging desejável.



- Representação formal do Staging de Câncer.
- 3 condições refletem → staging desejável:
 - Condição: considerar a existência de tumor solitário ou múltiplo no mesmo site.
 - Condição: considerar se os tumores sao maiores ou menores que certos tamanhos "X".
 - Condição: considerar lesões em órgãos adjacentes.



- Condição 1 : considerar a existência de tumor solitário ou múltiplo no mesmo site.
- Problema de Open World !!!
- Solução:
 - ► Informação → parsing AIM-XML para OWL (estrutura de dados).
 - Novo conceito → singleLesion (dataProperty) de :ImageStudy.



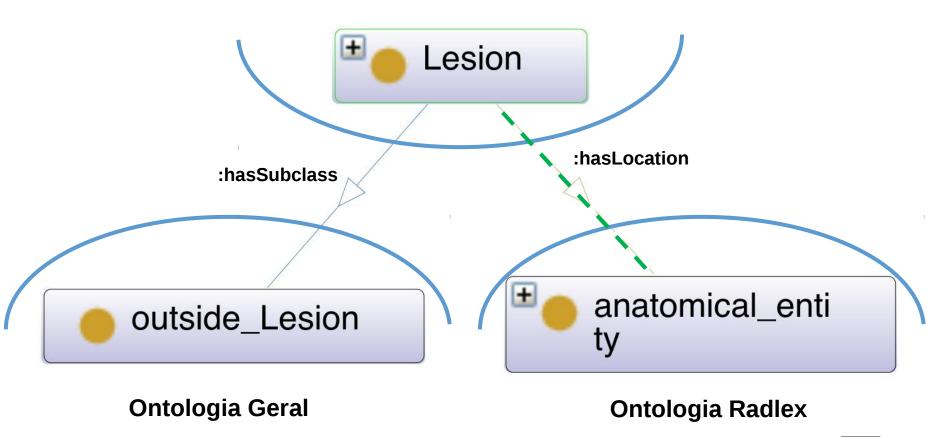
- Condição 1: considerar a existência de tumor solitário ou múltiplo no mesmo site.
 - Finalmente \rightarrow 2 regras SWRL.
- Condição 2: considerar se os tumores sao maiores ou menores que certos tamanhos "X".
 - Baseado → propiedade :value do :Calculation Result.
 - Fácilmente implementada → 2 regras SWRL.



- Condição 3: considerar lesões em órgãos adjacentes.
 - subClasse : OutsideLesion → :Lesion ONLIRA.
 - 3 novas propiedades:
 - ObjectProperty hasLocation
 - ONLIRA: Lesion → Radlex: Anatomical Entity



Ontologia ONLIRA

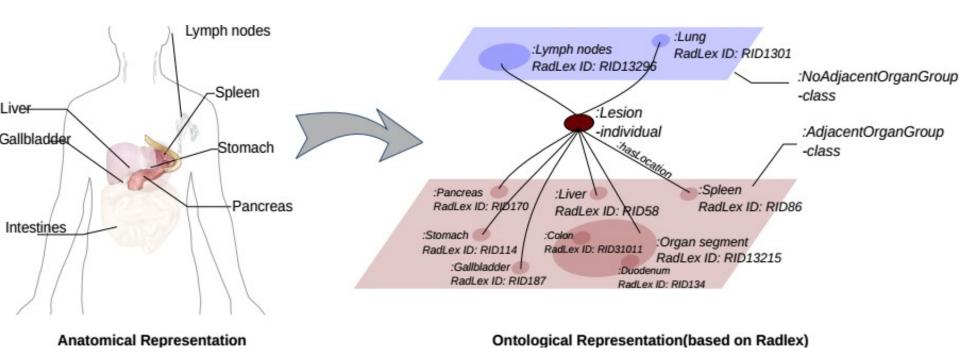




- Condição 3: considerar lesões em órgãos adjacentes.
 - subClasse : OutsideLesion → :Lesion ONLIRA.
 - 3 novas propiedades:
 - ObjectProperty : hasLocation
 - ONLIRA:**Lesion** → Radlex:**AnatomicalEntity**
 - 2. DataProperty :isRegionalLympNodeAffected
 - 3. DataProperty :isAdjacentOrgan



Hierarquia de subclasses da ontologia estabelecidas usando-se Radlex.





- Exemplo de formalização de regras (T)
- T1: Este critério é representado por uma classe definida com as seguintes condições necessárias e suficientes:
- Definição formal do T1:



Onde:

- NoVascularInvadedTumor → classe primitiva, afirma-se utilizando a regra SWRL 1
- _T1_singleTumor → classe definida representada pelo Axioma 1:

```
ImageAnnotation(?x), hasLesion(?x,?
Regra 1 y), Vasculature_Proximity(?y,?val), equal(?
val, "other") → _NoVascularInvadedTumor(?x)

Axioma 1 ImageReferenceEntity and (hasImageStudy some _SingleTumor)
```



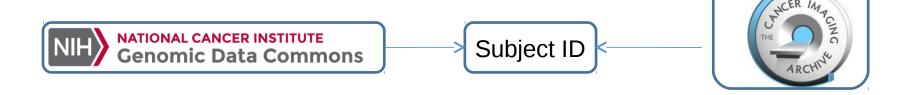
Onde:

 SingleTumor → classe primitiva. Ele é afirmado usando a regra SWRL 1.1.



Dados utilizados:

- Genomic Data Commons (GDC) do National Cancer Institute (NCI).
- The Cancer Imaging Archive (TCIA)





- Experimento 1: Avaliação quantitativa da ontologia AIM4-O
- Como poderia ajudar na busca de relatórios clínicos → resultados de análise de imagens.
- Foram comparadas 2 abordagens:
 - Com base em ontologias (busca semântica).
 - Com base em palavras-chave (busca por palavra-chave).



- Experimento 1: Avaliação quantitativa da ontologia AIM4-O
- Para destacar as diferenças entre essas duas abordagens
- Descrevemos quatro consultas → em DL (DL query) é palavraschave



- Experimento 1: Avaliação quantitativa da ontologia AIM4-O
- Q1 → Find all reports related to an image observation (tumor observation).
- Q2 \rightarrow Find all reports that describe multiple tumors.
- Q3 → Find all reports that contain a tumor observation that has a size greater than 8 cm.
- Q4 → Find all reports that contain a tumor observation with descriptors (i.e invasion, mass, vascular).



- Experimento 1: Avaliação quantitativa da ontologia AIM4-O
- Q1 → Find all reports related to an image observation (tumor observation).
- Q2 \rightarrow Find all reports that describe multiple tumors.
- Q3 → Find all reports that contain a tumor observation that has a size greater than 8 cm.
- Q4 → Find all reports that contain a tumor observation with descriptors (i.e invasion, mass, vascular).



- Experimento 1: Avaliação quantitativa da ontologia AIM4-O
- A avaliação baseia-se → Foram usados 15 relatórios radiológicos
 GDC 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.
- Comparamos precisão e revocação → gold-Standard (radiologista).



Resultados de precisão e revocação

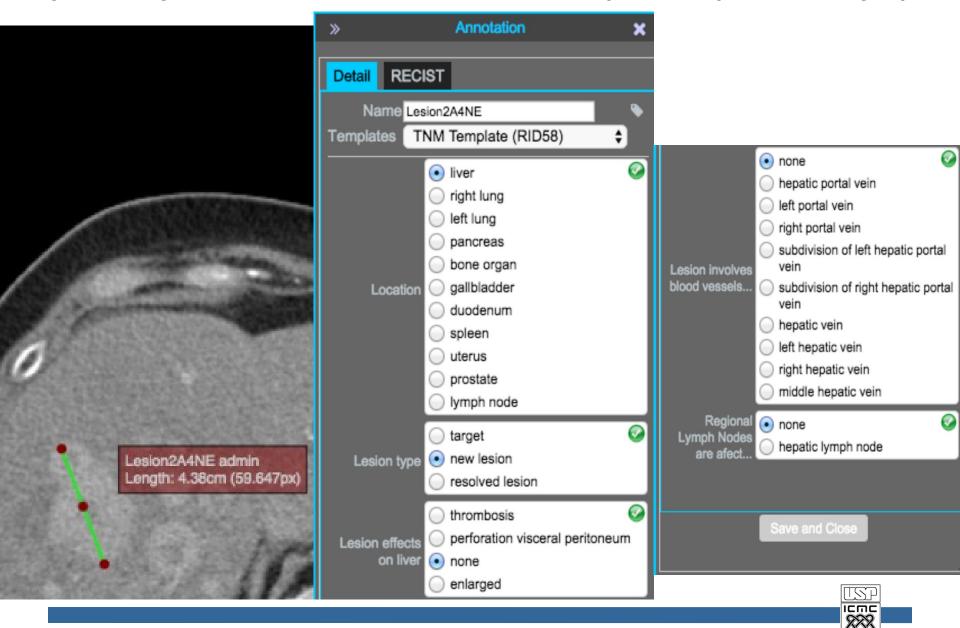
Query ID	Busca Sei	mântica- DL	Palavras-chave		
	Precisão	Revocação	Precisão	Revocação	
Q1	1	1	1	0.8	
Q2	1	1	0.4	0.4	
Q3	0.9	0.9	0	0	
Q4	0.67	1	1	0.7	



- Experimento 2: Determinação do estágio de câncer (cancer staging) baseado em TNM
- Primeiro criamos → TNM template.



Image CT do fígado sendo anotada usando o ePAD e a template TNM (direita da imagem)



- Experimento 2: Determinação do estágio de câncer (cancer staging) baseado em TNM
- Primeiro criamos → TNM template.
- Subseqüentemente → geramos anotações AIM-XML, classificamos usando o nosso classificador TNM.
- Processo:



- Experimento 2: Determinação do estágio de câncer (cancer staging) baseado em TNM
- Processo:
 - Separamos pacientes → informações no GDC como no TCIA=
 52
 - Lidos os relatórios médicos → planilha Excel
 - Excel + dados GDC → anotações AIM criadas e integradas (AIM4-O).
 - Anotações AIM → entrada do classificador TNM
 - Resultados do classificador vs staging dos relatórios médicos.



Geração manual do AIM apartir dos relatorios

</ImagingPhysicalEntity>

Relatorios

```
Surgical date:
TISSUE DESCRIPTION:
Liver Segments I, II, III and IV (2280 grams, 20 x 19 x 11 cm), separately submitted
left portal lymph node (4.5 x 2.6 x 1.2 cm) and gallbladder (7.5 x 3.5 x 1 cm).
     A1, A2, A3, B1, B2, C1, C2, C3, C4, C5, C6
DIAGNOSIS:
Liver, Segments I, II, III and IV, resection: Grade 3 (of 4) hepatocellular carcinoma,
                                                                                   AIM-xml
       <dateTime value="2016-05-30T22:43:29"/>
       <name value="xxxxxxxx"/>
       <comment value="xxxx"/>
       <imagingPhysicalEntityCollection>
         <ImagingPhysicalEntity>
           <uniqueIdentifier root="2.25.21343316135144112868648893393525445856835"/>
           <typeCode code="RID58" codeSystem="liver" codeSystemName="RadLex" codeSystemVersion="1.0"</pre>
           <annotatorConfidence value="0.0"/>
           <label value="Location"/>
```



Patient Key:

Geração manual do AIM apartir dos relatorios

Relatorios

The tumor closely approaches the surgical margin (less than 1 mm) which microscopically is uninvolved. Vascular invasion is identified, involving hepatic veins microscopically. The tumor is not encapsulated. The

AIM-xml



Matriz de Confusão dos estágios de câncer calculados pelo classificador

N=51			Actual Stages					
			IIA	IIIA	IIIB	IVA		
Predicted Stages	I	24	2	0	0	0		
	П	0	10	2	0	0		
	IIIA	0	1	8	0	0		
	IIIB	0	0	0	2	1		
	IVA	0	0	0	0	1		

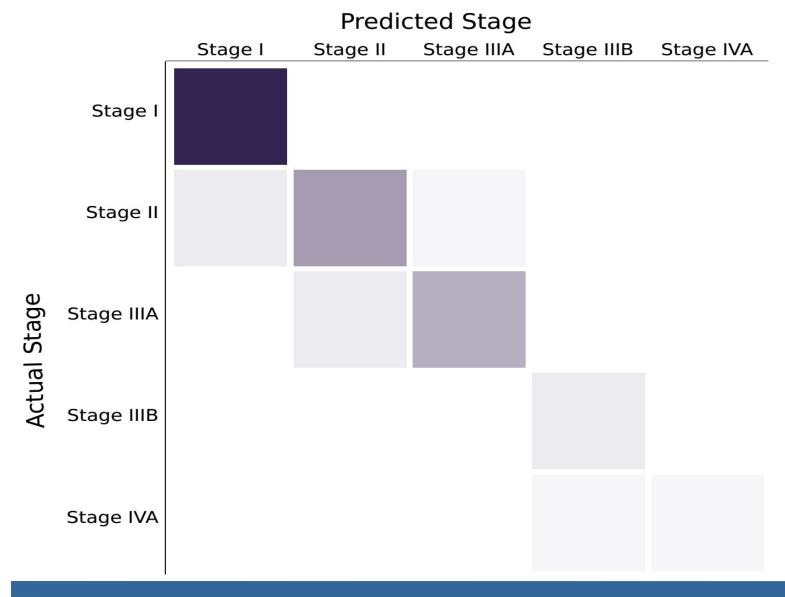


- A precisão foi 85,7% e a revocação 81,0% (para os 51 pacientes)→resultados muito bons (especialista).
- Classificador revelou casos clínicos com estagio errôneo paciente:
 Subject ID TCGA-DD-A1EJ.
- Tais incidentes não são excepcionais!.

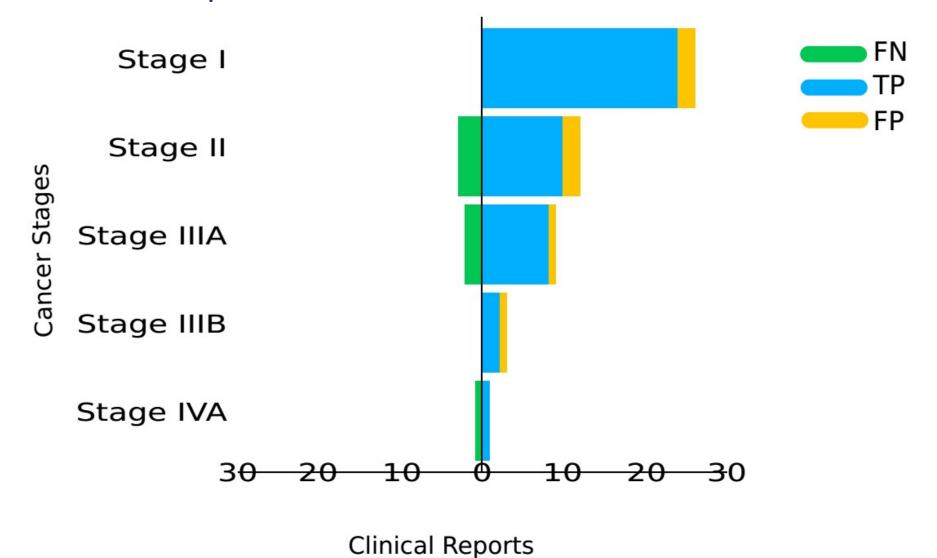


Matriz de escala de cores para a Matriz de Confusão.

67



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





Conclusão

Contribuições

- Classificador TNM → maior taxa de precisão e padronização na interpretação de estágios de lesões cancerosas.
- Além dos bons indicadores de precisão e revocação (85,7% e 81,0%), podem ser detectados casos clínicos com estagio errôneo → erros de tratamento médico podem ser reduzidos.



Conclusão

Contribuições

- Uma nova ontologia AIM4-O → pode ser adotada em aplicações relacionadas ao processamento de metadados de imagens.
- Avaliação quantitativa da ontologia AIM4-O → tarefa de busca por relatorios.

Precisão (89% x 52%) e revocação (98% x 48%).



Conclusão

Contribuições

- Método de mapeamento de anotações AIM-XML→ instâncias AIM4-O.
- Alinhamento do modelo AIM com ontologias médicas, como Radlex e Onlira.
- Representação formal de regras, axiomas e procedimentos de raciocínio automático >> para fazer cancer staging baseado em TNM.



Obrigado!!

