

JANUS - *A system for* annotating, vetting and testing Natural Language Processing Systems

Michael Tanenblatt, Anni Coden, Igor Sominsky
IBM, T.J. Watson Research Center

Creation of annotated corpus

The image displays a software interface for creating an annotated corpus. It consists of several windows:

- Text Editor:** Shows medical text with several words highlighted in red. The text includes:
 - GROSS DESCRIPTION
 - FINAL DIAGNOSIS
 - Sigmoid, ma
 - Sigmoid, po
 - Sigmoid and
 - Multiple lym
 - Sigmoid, mass at 22 cm, endoscopic biopsy (AE46-395890; 12/10/03): Invasive, grade 3 (of 4) ad
 - Sigmoid, polyp, polypectomy (AE46-39589H; 12/10/03): Tubulovillous adenoma with low-grade c
 - Sigmoid and rectum, resection specimen (AE46-32959; 12/23/03): Invasive, grade 3 (of 4) adenoc
 - Multiple lymph nodes (16 regional and 2 separately submitted inferior mesenteric artery lymph nod
- Annotation List:** A list of annotation types with checkboxes. The checked items are:
 - SCRAnatomicalSite
 - SCRCoreference
- Edit Panel:** A panel for editing the selected annotation type, 'SCRAnatomicalSite'. It contains:
 - begin: 35, end: 42
 - Fragments: Sigmoid
 - Negation: 0
 - Inference: 0
 - Terminology: ICD9

Visualization by concept

endoscopic biopsy (AE46-39589O; 12/10/03): Invasive, grade 3 (of 4) adenocarcinoma identified.

omy (AE46-39589H; 12/10/03): Tubulovillous adenoma with low-grade dysplasia, completely excised.

ction specimen (AE46-32959; 12/23/03): Invasive, grade 3 (of 4) adenocarcinoma forms a fungating, circumferential, polypoid mass (with regional and 2 separately submitted inferior mesenteric artery lymph nodes) are negative for tumor. Surgical

The interface includes a sidebar with a concept tree and a detailed view for the selected concept, SCRPrimaryTumorReading.

Files | **Types** | **Annotations** | **Diffs**

- ▶ SCRDimension
- SCRGenericReading
- ▶ SCRGradeValue
- SCRGrossDescription
- SCRGrossDescriptionPart
- ▶ SCRHistologicalDiagnosis
- ▶ SCRlymphNodes
- ▶ SCRlymphNodesReading
- SCRMetastaticTumor
- SCRMetastaticTumorReading
- SCRMultipleReadingModel
- ▶ SCRNamedEntity
- SCRNamedEntityBase
- ▶ SCRPrimaryTumor
- ▶ SCRPrimaryTumorReading
- ▶ SCRSize

Edit SCRPrimaryTumorReading

begin 368 end 382

Diagnoses adenocarcinoma

Size 7 x 3 x 3 cm

PrimarySites rectum
Sigmoid

GradeValues 3 (of 4)

Specification of comparison

```
• <?xml version="1.0" encoding="UTF-8"?>
• <tns:CFEConfig xmlns:tns="http://www.apache.org/uima/cfe/config" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
• nullValueImage="" xsi:schemaLocation="http://www.apache.org/uima/cfe/config CFEConfig.xsd ">
•   <tns:comparisonDescriptor>
•     <tns:truth className="truth" enclosingAnnotation="uima.tt.sentenceAnnotation">
•       <tns:targetAnnotationMatcher annotationTypeName="SCRAnatomicalSite"/>
•       <tns:featureAnnotationMatchers annotationTypeName="SCRAnatomicalSite" windowSizeInside="1">
•         <tns:groupFeatureMatchers>
•           <tns:featureMatchers refId="L1" featurePath="begin"/>
•           <tns:featureMatchers refId="L2" featurePath="end"/>
•           <tns:featureMatchers refId="L3" featurePath="code"/>
•           <tns:featureMatchers refId="L4" featurePath="coveredText"/>
•         </tns:groupFeatureMatchers>
•       </tns:featureAnnotationMatchers>
•     </tns:truth>
•     <tns:experiment className="experiment" enclosingAnnotation="uima.tt.sentenceAnnotation">
•       <tns:targetAnnotationMatcher annotationTypeName="AnatomicalSite"/>
•       <tns:featureAnnotationMatchers annotationTypeName="AnatomicalSite" windowSizeInside="1">
•         <tns:groupFeatureMatchers>
•           <tns:featureMatchers refId="R2" featurePath="begin"/>
•           <tns:featureMatchers refId="R3" featurePath="end"/>
•           <tns:featureMatchers refId="R4" featurePath="conceptCode"/>
•           <tns:featureMatchers refId="R5" featurePath="coveredText"/>
•         </tns:groupFeatureMatchers>
•       </tns:featureAnnotationMatchers>
•     </tns:experiment>
•   <tns:transformationRules/>
•   <tns:comparisonRules expression="$L1==$R1" refId="CR1"/>
•   <tns:comparisonRules expression="$L2==$R2" refId="CR2"/>
•   <tns:comparisonRules expression="$L3==$R3" refId="CR3"/>
•   <tns:comparisonRules expression="$L4==$R4" refId="CR4"/>
• </tns:comparisonDescriptor>
• </tns:CFEConfig>
```

Visualization of comparison

GROSS DESCRIPTION

FINAL DIAGNOSIS

Sigmoid, mass at 22 cm, endoscopic biopsy (AE46-39589O; 12/10/03): Invasive, grade 3 (of 4) ad

Sigmoid, polyp, polypectomy (AE46-39589H; 12/10/03): Tubulovillous adenoma with low-grade c

Sigmoid and rectum, resection specimen (AE46-32959; 12/23/03): Invasive, grade 3 (of 4) adenoc.

Multiple **lymph nodes (16 regional and 2 separately submitted inferior mesenteric)** artery lymph nod

Files Types Annotations Diffs

SCRAnatomicalSite

All Differing Matching

Source	Begin ▲	End	Covered Text
BOTH	35	42	sigmoid
BOTH	156	163	sigmoid
BOTH	279	286	sigmoid
BOTH	291	297	rectum
BOTH	521	524	fat
BOTH	557	567	peritoneal
REF	630	641	lymph nodes
	630	701	lymph nodes
REF	709	720	lymph nodes
	636	714	nodes (16 re
REF	709	720	lymph nodes
	682	720	inferior mese

Visual evaluation

GROSS DESCRIPTION

FINAL DIAGNOSIS
Sigmoid, mass at 22 cm, endoscopic biopsy (AE46-395890; 12/10/03): Invasive, grade 3 (of 4) ad
Sigmoid, polyp, polypectomy (AE46-395890; 12/10/03): Tubulovillous adenoma with low grade
Sigmoid and rectum, resection s
Multiple lymph nodes (16 region

SCRAnatomicalSite

begin 630 end 701

Fragments lymph nodes inferior mesenteric

Negation 0

Inference 0

Terminology ICDO

Code C77.2

Coreferences lymph nodes (16

AnatomicalSite

begin 630 end 641

Fragments lymph nodes

Negation 0

Inference 0

Terminology ICDO

Code C77.9

Coreferences lymph nodes

in	End	Covered Text
	42	sigmoid
5	163	sigmoid
9	286	sigmoid
1	297	rectum
1	524	fat
7	567	peritoneal
0	641	lymph nodes
0	701	lymph nodes
9	720	lymph nodes
5	714	nodes (16 re
9	720	lymph nodes
2	720	inferior mese