MetaMap Transfer (MMTx)

An Overview
Presented December 8, 2005
MMTx
Outline

• What is it?
• What is it good for?
• Who uses it?
• What does it look like?
• Where can I get it?
• Requirements, Terms and Conditions
• Customization, Common Options
• Team Members and Points of Contact
Tool for discovering controlled medical vocabulary terms (via UMLS Concepts) within documents

- Tokenizes documents into sentences, phrases, terms, words
- Matches phrases to closest UMLS Concepts
- Java Implementation of MetaMap
- It is Middleware. There is no GUI
MMTx
What is it Good for?

- Information Extraction tasks
- Classification/Categorization tasks
- Text Summarization/Question & Answer tasks
- Data-mining tasks
- Knowledge discovery tasks
- Text Understanding tasks
- UMLS Concept based indexing and retrieval
- NLP Tasks
Medical Problem Extraction from Clinical Documents
Extraction of Concepts from Chief Complaints
Extraction of Concepts from Clinical Narratives
Extracting Diagnoses from Discharge Summaries
Concept-Value Pair Extraction from Semi-Structured Echocardiogram Reports
Enzyme class Annotation
Term Identification in the Biomedical Literature
Extracting Molecular Binding Relationships from Biomedical Text
What is retinoblastoma?

Retinoblastoma is a rare type of eye cancer that develops in the retina, which is the part of the eye that detects light and color. Although this disorder can occur at any age, it usually develops in young children.
MMTx
Why would you want to use it?
Inverse_isa
Adjacent_to
Connected_to
Constitutes

Anatomical space
Uveal Diseases
Uvea
Iris structure

Anatomical conduit
Iris (Eye)

Pupil C0034121

Pupil of left eye
Pupil of right eye

MMTx
Why would you want to use it? (2)
### Phrase: "of eye cancer"

<table>
<thead>
<tr>
<th>Meta Candidates (7)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 Eye Cancer (Malignant neoplasm of eye, unspecified) [Neoplastic Process]</td>
<td></td>
</tr>
<tr>
<td>861 Cancer (Malignant Neoplasms) [Neoplastic Process]</td>
<td></td>
</tr>
<tr>
<td>861 Cancer (Malignant neoplasm, primary (morphologic abnormality)) [Neoplastic Process]</td>
<td></td>
</tr>
<tr>
<td>861 Cancer (Cancer Genus) [Invertebrate]</td>
<td></td>
</tr>
<tr>
<td>694 Eye [Body Part, Organ, or Organ Component]</td>
<td></td>
</tr>
<tr>
<td>694 Eye (Entire eye) [Body Part, Organ, or Organ Component]</td>
<td></td>
</tr>
<tr>
<td>638 Ophthalmic [Spatial Concept]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Meta Mapping (1000)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 Eye Cancer (Malignant neoplasm of eye, unspecified) [Neoplastic Process]</td>
<td></td>
</tr>
</tbody>
</table>
MMTx
What does the API look like?

Sections

Final Mappings

UMLS Concepts

Phrases

Sentences

Document

Java Container Classes
## MMTx API Methods and Constructors

<table>
<thead>
<tr>
<th>MMTxAPI</th>
<th>MMTxAPI(String[] args)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document</td>
<td><strong>processDocument</strong></td>
</tr>
<tr>
<td>void</td>
<td><strong>processSentence</strong></td>
</tr>
<tr>
<td>Sentence</td>
<td><strong>processSentence</strong></td>
</tr>
<tr>
<td>Sentence</td>
<td><strong>processString</strong></td>
</tr>
<tr>
<td>Phrase</td>
<td><strong>processTerm</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document</td>
<td><strong>processDocument</strong> (File pFile)</td>
</tr>
<tr>
<td>Document</td>
<td><strong>processDocument</strong> (String pDocumentText)</td>
</tr>
<tr>
<td>void</td>
<td><strong>processSentence</strong> (Sentence pSentence)</td>
</tr>
<tr>
<td>Sentence</td>
<td><strong>processSentence</strong> (String pSentenceText)</td>
</tr>
<tr>
<td>Sentence</td>
<td><strong>processString</strong> (String pString, boolean pTrmPrcsng)</td>
</tr>
<tr>
<td>Phrase</td>
<td><strong>processTerm</strong> (String pTerm)</td>
</tr>
</tbody>
</table>
MMTx
How does it Work?

Medical Text

Tokenization

POS Tagger Client

Lexical Lookup

Noun Phrase Parser

Variant Generation

Candidate Retrieval

Evaluation

Final Mapping

Document
Sections
Sentences
Phrases
UMLS Concepts
Final Mappings
MMTx
Where can I get it?

- The download is password restricted
- Preconditions:
  - Signed the UMLS license agreement
  - UMLS Knowledge source server account name and password
MMTx
Minimum Requirements

• OS’s supported:
  – Solaris/Windows NT/2000/XP, Linux, Mac OS/X
• Java 1.4 or better
• 400 mb disk space for software
• 3GB disk space for each Year’s Data per model
• 600+ MB or more RAM/Swap space
• Must be a UMLS Signatory
• Must customize the data to the vocabularies that you have rights to use
• The MMTx software soon to be under an open source agreement:
  Attribution, redistribution in-total, no NLM endorsement, indemnity
Can I customize the data?

• You will need to customize it because …
  – Need to extract only those vocabularies you have rights to
  – Extract only those vocabularies that make sense for your application
How can I customize the data?
MMTx
How do I tune it?

- Limit matches to conservative variation
  - Noun/adjective derivations, unique acronyms and expansions
- Filter by semantic type(s)
- Filter by vocabulary source(s)
- Match on (longer) composite phrases
MMTx
Team Members and Points of Contact

mmtx@nlm.nih.gov