

Mapping Free-Text Anatomical Descriptions to UMLS and to the Visible Human Dataset

Jason Sarnoski, Craig Morioka, Ricky Taira, and Hooshang Kangarloo

Medical Informatics, UCLA, Los Angeles, CA, USA

We propose the use of an anatomical model to map free-text terms to UMLS concepts, based on the individual mapping of each component onto the Visible Human Project model. We processed 3000 radiology reports covering the urinary system to extract relevant, descriptive, anatomical phrases utilizing our statistical based natural language processing (NLP) algorithm. The anatomical locations and spatial modifiers were linked to Unified Medical Language System (UMLS) concept codes. Corresponding UMLS terms identified as anatomical locations were also independently identified as 3-D coordinates within the Visible Human dataset. The goal is to link the free-text information from a clinical document to both a standard nomenclature and an image-based model.

Materials and Methods

The steps required to accomplish the linkage of a textual anatomical description from a biomedical document to the relevant images involve the following:

1. Natural Language Processing (NLP) Processing of document: identification of the relevant biomedical phrases in the document, in particular the anatomical terms and their modifiers.
2. Parse NLP extracted phrases into anatomical locations, and their modifiers to narrow search, and allow UMLS code to be more specific.
3. UMLS to CUI Finder: Search UMLS for concept unique identifier (CUI) for key anatomical structure.
4. Map CUI to 3-dimensional coordinate: Find center of mass coordinate for segmented anatomical structure.
5. Refinement of CUI to 3-D coordinate: Based on the descriptive phrase parsed by NLP, coordinate locations can be more precise than just the location of the organ

Results

For our test set of 150 NLP extracted phrases, 100% of the organs identified by the parser were correctly linked to an appropriate UMLS CUI. In addition, spatial modifiers such as right, left, upper, lower, superior, inferior were also correctly identified. In Table 1 column 1 is an example of the anatomy phrases extracted from NLP. The anatomy phrase is parsed into semantic classifications, primarily spatial direction modifiers and organs. These anatomical structures are then mapped to unique UMLS CUI codes shown in column 3. This provides a standard nomenclature to refer to the textual information. In Table 2 the NLP expressions and associated UMLS codes are given a three-dimensional coordinate in column 3. This coordinate adds a spatial model to both the UMLS code and the text that pointed to it, bridging textual information and imaging

data. These results demonstrate that free-text data can be structured semantically and spatially. The segmentation and center of mass program was written in IDL version 5.4 (Research Systems Inc. Boulder, CO)

Table 1 – Sample of NLP parsed Anatomical Phrases from Free-text Radiology Report and mapping to UMLS concept string and unique identifier

	Anatomy Phrases	UMLS Concept String	UMLS CUI
1	spleen	spleen	C0037993
2	left ureter proximally	left ureter	C0227683
3	portions of the right ureter	right ureter	C0227862
4	region of the kidneys	kidneys	C0022646
5	area of the left kidney	left kidney	C0227614
6	mid right kidney	right kidney	C0227613
7	right and left kidneys	right kidney, left kidney	C0227613, C0227614
8	right and left ureter	right ureter, left ureter	C0227862, C0227863

Table 2 Mapping of UML Concept String Phrase to 3-dimensional coordinate of center of mass for the kidney and various descriptions

Phrase	UMLS CUI	Coordinates – Center of Mass
right kidney	C0227613	(333,989,633)
left kidney	C0227614	(567,1016,614)
kidneys	C0022646	(333,989,633), (567,1016,614)
superior pole of the right kidney	C0734758	(350,1001,605)
superior pole of the left kidney	C0734759	(542,1022,588)
inferior pole of the left kidney	C0734760	(564,984,640)
inferior pole of the right kidney	C0734761	(337,963,661)

Acknowledgments

This research is being supported in part by the National Institute of Health Grant P01-EB00216-12