Exploring the Semantic Relations of Radiological Terms using UMLS Resources

David A. Campbell, Donald P. Harrington M.D., James J. Cimino M.D.
Department of Medical Informatics, Columbia University

INTRODUCTION

Natural Language Processing (NLP) is recognized as an option for transforming medical text into a codable form[1,2]. NLP systems use a lexicon of the terms expected in the target text. Terminologies, however, are concept driven; many terminology entries have little relationship to the written or spoken usage of the concept. The fullest effect of NLP will only be realized when the medically relevant lexical terms used in real text are mapped to a controlled terminology. Although work has been done mapping terminology concepts to one another[3], the derivational difference between lexica (lexical) and terminologies (conceptual) makes similar mappings more complicated.

The Unified Medical Language System (UMLS) has been used in various data creation, indexing and encoding systems. The UMLS accomplishes this by conjoining the sets of synonyms and concept relationships in its multiple constituent terminologies. The large number of terms associated with a concept increases the opportunity for successfully identifying a term lexically. We utilized the UMLS to map radiological lexical terms from the MedLEE language processor to concepts in the Medical Entities Dictionary (MED) terminology.

METHODS

A total of 3,436 MedLEE lexical terms of the semantic type 'problem' in the radiology domain were not paired with MED concepts and were lexically analyzed to determine a best-matching concept within the UMLS. All terms were broken up into their component words and matched to the Metathesaurus (MRCON) using the UMLS word index (MRXW.ENG) as a reference resulting in possible Concept Unique Identifiers (CUIs) for the term. Mappings of the CUIs to the MED were then constructed. If the CUI was found already in the MED, then the MedLEE term was mapped directly to that MED concept. If no exact match was found a recursive search of the UMLS relationship graph (MRREL) was used following the 'parent' relation to generate all the possible parents of the matched CUI.

A randomly generated set of MedLEE lexical terms that generated at least one MED concept were reviewed by an expert radiologist (DPH). These concepts were evaluated for their appropriateness in classifying the terms.

RESULTS

Of the 3,436 MedLEE terms used in the experiment, 2,812 (82%) were lexically mapped to the UMLS. The accuracy of lexical mapping was not measured. A total of 2,459 (87%) UMLS concepts were assigned MED concepts for an overall matching of 72%. Manual review of these matches showed that 90% of suggested matches were appropriate for the term, being either an exact match or a parent.

CONCLUSION

We have described a process to map lexical terms to terminology concepts using UMLS resources. The UMLS was used to address both lexical and logical issues. The UMLS's wide range of synonymy ameliorated the assignment of concepts to the MedLEE lexical items. Likewise, the UMLS relationships allowed for a general placement of the concept within the terminology framework when exact conceptual matches were not found. It is likely that the process would be successful in conceptually placing any set of lexical terms to any target terminology as it was independent of both the lexicon and the terminology. The question of how radiology terms should ultimately be represented in a controlled terminology is not addressed by this paper, although perhaps it is the most intriguing one. The process described here does suggest appropriate locations for radiology concepts in the context of the existing MED ontology.

Acknowledgments

This work was supported by the National Library of Medicine Training Grant LM07079

References

[1]Sager N, Friedman C, et al. Natural Language Processing and the Representation of Clinical Data. JAMIA. 1(2): 142-60, 1994.

[2] Friedman, C. Natural Language Processing in an Operational Clinical Information System. Natural Language Engineering. 1(1): 1995. 83-108.

[3] Dolin, R. H.; Huff, S. M.; Rocha, R. A.; Spackman, K. A., and Campbell, K. E. Evaluation of a "lexically assign, logically refine" strategy for semi-automated integration of overlapping terminologies. JAMIA. 1998 Mar-1998 Apr 30; 5(2):203-13.