

Strategies for Mapping Concepts in Gastrointestinal Endoscopy Reports to the UMLS Metathesaurus

Michele Tringali^a, Thomas C. Rindflesch^b, Halil Kilicoglu^b, Marcelo Fiszman^b, Olivier Bodenreider^b

^aAzienda Ospedaliera Santa Maria della Misericordia, Udine, Italy

^bNational Library of Medicine, Bethesda, MD, USA

Introduction

Automatic access to information in clinical free text can provide support for medical decision support, outcome research, continuous quality improvement, and information retrieval. We discuss three strategies to improve results when using MetaMap [1] to map strings in gastrointestinal endoscopy (GIE) reports to concepts in the Unified Medical Language System® (UMLS)® Metathesaurus.® Due to word sense ambiguity, synonymy, and context-specific meaning, using MetaMap without modification produces limited results in this domain. For example, *esophagus* is incorrectly mapped to “Procedures of the esophagus.”

Materials and Methods

A sample of 80 (anonymized) esophago-gastroduodenoscopy reports (250 to 300 words) from Clarkson University Hospital, Omaha, Nebraska, was randomly selected and divided into training (55 reports) and testing (25 reports) sets. A gold standard was generated manually in both sets by selecting Metathesaurus concepts to represent GIE content. We used the training set to devise general methods to improve mapping of free text to the Metathesaurus. Baseline processing uses MetaMap without enhancements.

We use three strategies for enhancing baseline performance. The first of these is implemented as special processing for a small set of frequently occurring concepts in GIE. The string *fundus*, for example, is mapped to “gastric fundus” and not “Fundus of bladder.” Missing synonyms in the Metathesaurus (such as “mid esophagus” for “Middle third of esophagus”) were added to this set as they were discovered in the training set.

Two additional strategies depend on enriched knowledge sources. For the first, we isolated a list of gastroenterology concepts in the Metathesaurus by extracting all hierarchical descendants of a seed concept (e.g. “Gastrointestinal Diseases”). MetaMap processing is limited to this list. The second strategy depends on the identification of general medical terms that appear in GIE reports but are not specific to the GIE domain, such as *veins* and *abdomen*. These were identified in the training set and included as GIE terms.

Results

We tested the effectiveness (precision and recall) of these three methods by comparing output to both the training and test sets. The results are shown in Table 1, without enhancements to MetaMap (Base) and with the three strategies (After).

Table 1 – Effectiveness measures for training and testing sets

	Base		After	
	Recall	Precision	Recall	Precision
Training	65%	64%	71%	79%
Testing	60%	64%	64%	80%

Conclusion

We have developed and evaluated three general strategies for mapping free-text in gastrointestinal endoscopy reports to UMLS Metathesaurus concepts. These enhancements increase precision without degrading recall. Long term goals of this project based on concepts identified in these reports include automatic triggering of GIE-related guidelines, semantic annotation of GIE images, information retrieval in the GIE domain, and point of care decision support systems that depend on data from GIE reports.

Reference

- [1] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. Proc AMIA Symp 2001:17-21.

Address for correspondence

Thomas Rindflesch, National Library of Medicine,
8600 Rockville Pike, MS 43, Bethesda, MD 20894, USA.
Email: tcr@nlm.nih.gov. Phone: (301) 435-3191.