Interpreting Hypernymic Propositions in an Online Medical Encyclopedia

Marcelo Fiszman MD PhD, Thomas C. Rindflesch PhD, and Halil Kilicoglu MS
National Library of Medicine, Bethesda, Maryland 20894

ABSTRACT

*Interpretation of semantic propositions from biomedical texts documents would provide valuable support to natural language processing (NLP) applications. We are developing a methodology to interpret a kind of semantic proposition, the hypernymic proposition, in MEDLINE abstracts. In this paper, we expanded the system to identify these structures in a different discourse domain: the Medical Encyclopedia from the National Library of Medicine's MEDLINE*plus*® Website*.

INTRODUCTION

The identification of semantic propositions in biomedical texts has the potential to support information retrieval and other applications in the biomedical domain. A specific kind of semantic proposition, the hypernymic proposition, occurs when two concepts, one more specific (hyponym) and the other more general (hypernym), are in a taxonomic relationship. This structure is important in the medical texts, particularly as a way of introducing new information. An example of a hypernymic proposition is illustrated by the relationship between "**Esophageal cancer"** and "**malignant tumor**" in "*Esophageal cancer is a malignant tumor of the esophagus*" The hypernymic proposition can be represented as: "Esophageal cancer-ISA-Malignant tumor."

We are developing a system called SemSpec[1] to interpret these structures in MEDLINE abstracts. In this paper, we extended the use of the system to a different discourse domain: the Medical Encyclopedia from The National Library of Medicine's MEDLINE*plus* Web site.

The medical encyclopedia has definitions for thousands of concepts, including diseases, procedures, medications, and medical diagnosis tests. The definitions are presented in specific sections of Web pages and are in free-text format. We evaluated the effectiveness of SemSpec to interpret hypernymic propositions in medical encyclopedia definitional sections.

METHODS

Interpretation of Hypernymic Propositions

To interpret hypernymic propositions SemSpec first identifies syntactic structures that potentially encode hypernymic propositions (verbs, appositive structures, and nominal modification). After syntactic arguments have been identified, MetaMap[2] matches them to concepts in the Metathesaurus.® Such concepts are then subjected to semantic validation. These must occur within the same semantic group as defined by McCray et al.[3] The current version of the system includes the semantic groups Disorders, Procedures and Chemicals & Drugs. The concepts themselves must be in hierarchical relationship in the Metathesaurus.

Evaluation

We conducted a preliminary evaluation to test the performance of SemSpec in extracting hypernymic propositions form free text definitional sections in the Medical Encyclopedia of MEDLINE*plus*. 430 definitional sections were randomly selected. These were then submitted to SemSpec and evaluated for correctness by a one of the authors (MF).

RESULTS

The 430 definition sections contained 590 sentences. SemSpec generated 159 hypernymic propositions. Of these, 129 were considered correct, resulting in precision of 81%. The majority of the propositions belonged to the semantic group Disorders (75%) and the majority were syntactically encoded by the verb *be*.

DISCUSSION

We extended a system that was developed to interpret hypernymic propositions in MEDLINE abstracts in a different a discourse domain. False positives were mostly related to incorrect argument identification of the verb *be* and coordination problems. The ability to interpret hypernymic propositions in online Web pages might be useful for information retrieval applications on the Internet.

References

1. Rindflesch TC, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. J Biomed Inform. Submitted, 2003.
2. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. Proc AMIA Symp 2001;:17-21.
3. McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. Medinfo 2001;10(Pt 1):216-20.