

Semantic Processing to Enhance Retrieval of Diagnosis Citations from Medline

Charles Sneiderman, MD PhD,¹ Dina Demner-Fushman, MD PhD,¹
Marcelo Fiszman, MD PhD,² Graciela Rosemblat, PhD,¹ François-Michel Lang MS,¹
Daphne Norwood, MD,² and Thomas C. Rindflesch, PhD¹

¹National Library of Medicine, Bethesda, MD

²Graduate School of Medicine, University of Tennessee, Knoxville, TN

ABSTRACT

We investigate the use of natural language processing (NLP) for enhancing precision when retrieving Medline citations on diagnostic procedure. Inter-annotator agreement is described as part of the evaluation method, and preliminary results are presented.

INTRODUCTION

Diagnosis is the cornerstone of the clinical process, and clinicians can benefit from convenient access to the research literature when considering diagnosis options [1]. We investigate the use of SemRep [2] for enhancing the results of PubMed searches on diagnostic procedures. We evaluate our method based on a test collection of citations relevant to diagnostic questions.

BACKGROUND

SemRep is an NLP system for representing part of the content of biomedical text with semantic predications. The program relies on medical knowledge in the Unified Medical Language System and produces, for example, the predication “Pulmonary Arteriogram DIAGNOSES Pulmonary Embolism” for the text *Pulmonary angiography is the gold standard for diagnosis of segmental pulmonary embolism*.

METHODS

A small test collection was created by using PubMed Clinical Queries for diagnosis to retrieve citations to answer questions about diagnostic procedures. Five questions were obtained from Family Practice Inquiries Network (FPIN). The first 50 citations for each question were evaluated by the second author. The first and the sixth authors, and a surgeon with 15 years experience provided additional judgments for 10 citations for each question. Evaluation was done on a three point scale: containing an answer, topically relevant, not relevant. Pairwise inter-annotator agreement was evaluated using Cohen’s kappa. Relevance judgments for 50 citations, with 10 reconciled between the first and second authors, were used to evaluate ranking of retrievals results based on semantic processing.

Citations were processed using SemRep retargeted to address diagnosis text. Citations with semantic

predications involving procedures and disorders from the query, identified in the document, were promoted to the top of the ranked list, in the order of frequency of the relations.

Evaluation was conducted using the treceval package and mean average precision (MAP), an official Text Retrieval Conference (TREC) [3] evaluation metric.

RESULTS

Table 1 presents retrieval results evaluated under two conditions: soft relevant means both topically relevant documents and documents answering the question were used as relevant in computing MAP; whereas only documents answering the question were used as hard relevance judgments.

Table 1. SemRep results compared to PubMed alone

MAP	Soft Relevant	Hard Relevant
PubMed	0.2051	0.2930
SemRep	0.2537	0.3026

DISCUSSION

Initial results are promising, in that SemRep predications were able to support somewhat better results than PubMed alone. The method needs to be tested on a larger collection of diagnosis questions and citations. Despite some variations, agreement between annotators is fairly uniform, probably indicating relevance judgments are interchangeable.

Table 2. Inter-annotator agreement

	a1-a2	a1-a3	a1-a4	a2-a3	a3-a4
3 point	0.44	0.50	0.51	0.50	0.50
binary	0.60	0.66	0.66	0.73	0.53

References

1. Wilczynski NL, Haynes RB. Hedges Team. EMBASE search strategies for identifying methodologically sound diagnostic studies for use by clinicians and researchers. BMC Med. 2005;3(1):7.
2. 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. 2003;36:462-477.
3. <http://trec.nist.gov/>