A Web Application to Support Consumer Health Vocabulary Development

Jon Crowell[1], Qing Zeng[1], Tony Tse[2]

[1]Decision Systems Group, Brigham & Women's Hospital and Harvard Medical School
[2]LHNCBC, National Library of Medicine, National Institutes of Health, DHHS

ABSTRACT

We describe a Web application that supports collaborative development of a consumer health vocabulary. It performs text analyses and enables distributed human review. It also provides on-the-fly summary reports and facilitates the generation of a final vocabulary based on the results of the review.

INTRODUCTION

Consumer health vocabularies (CHV) consist of expressions (i.e., words and phrases) commonly used by laypersons to refer to medical concepts (e.g., heart attack for myocardial infarction). Identifying and characterizing consumer expressions and their intended meanings is a tedious manual task. Generally, the steps include selecting and annotating candidate terms from a corpus, analyzing contextual information to discern the intended meaning, and reaching consensus among multiple reviewers representing different perspectives.

We designed a Web application, VocabTool, to support the above steps especially the analysis of term- and concept-related information by geographically distributed reviewers. In a recent study [1], six reviewers used VocabTool to review 2,193 candidate CHV terms from a corpus of 12.5 million MedlinePlus® queries.

METHODS & RESULTS

VocabTool consists of three parts:

Processing Raw Data: VocabTool extracts queries, time stamps, and de-identified IP addresses from raw Web logs. It converts queries into ngrams and stores them, along with their counts, in a database. VocabTool then maps the queries and ngrams to UMLS® concepts using lexical processes.

Presenting Data for Review: VocabTool provides three components for reviewing the processed data (Figure 1):

1. Term-Concept Mappings. Allows reviewers to assess and annotate the mapping quality, including noting more appropriate UMLS concepts. Each UMLS concept is dynamically linked to the Semantic Navigator, providing contextual information from the UMLS Semantic Network.

2. Ngram Review. Allows reviewers to browse and search for expressions not mapped to UMLS terms. For example, the candidate CHV bi-gram "heart doctor" may occur frequency but fail to map to C0175906 ("Cardiologist").

3. Concept, Term, and Relationship Creation. Some health-related concepts or terms used by consumers do not exist in the UMLS. Reviewers need to create such concepts, terms and relations for use in a CHV.

Vocabulary Finalization: Reviews are recorded as they are entered. Reports are generated to facilitate consensus building and finalization of the vocabulary.

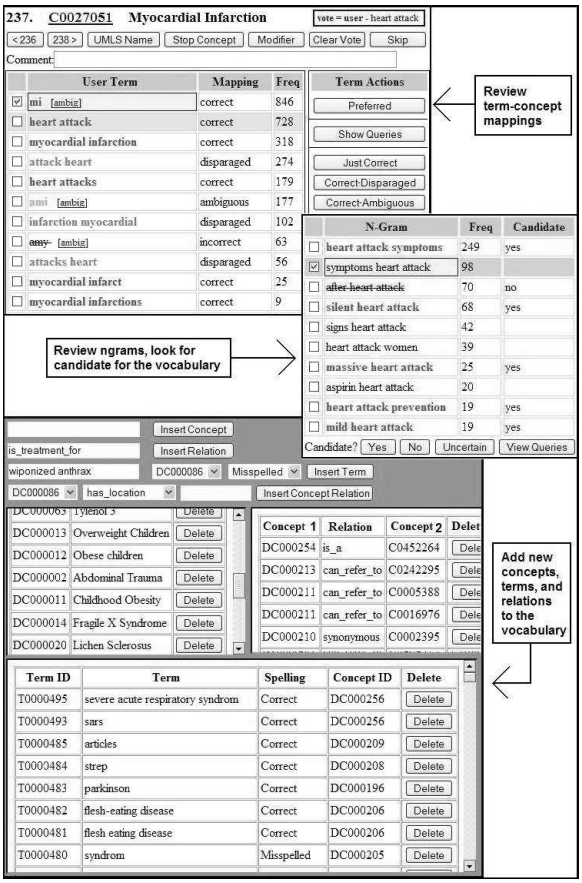

Figure 1. VocabTool screen shots

FURTHER WORK

An XML DTD for storing vocabulary terms and modules for reporting voting statistics are being developed.

1. Zeng QT et al. Identifying Consumer-Friendly Display (CFD) Names for Health Concepts. AMIA. 2005. Submitted.