Towards Automatic Augmentation of Electronic Medical Records with MEDLINE® citations.

Vinod Chacko, MD
PGY2, Internal Medicine Residency, Easton Hospital, Easton, PA. Elective in Clinical Informatics, National Library of Medicine (NLM), NIH, DHHS, Bethesda, Maryland.

Abstract

The adoption of Electronic Medical Records (EMR) is increasing every year even though the rate of adoption is slow. MEDLINE® is adding thousands of new citations every day. It may be possible to enrich an EMR system with relevant information from NLM resources. During my elective at NLM I was assigned to evaluate the preliminary stages in designing and building a system, which automatically augment a patient’s EMR with pertinent information from NLM resources. Clinical informatics experiments using content from real EMR is necessary to achieve the ultimate goal of automatic augmentation of the EMR with MEDLINE®.

Objective: (i) To evaluate a preliminary methodology for extracting query terms from free text clinical data (ii) To evaluate the relevancy and quality of MEDLINE® search results generated by these search terms using Essie(1) search engine. (iii) To evaluate the summary created by RIDE-M(2).

Methodology: As a surrogate for real EMR data, this project utilized free text clinical scenarios and questions recorded by residents and a medical student at two teaching hospitals. Each clinical scenario consists of one or two sentence summary of the clinical case and a one sentence clinical question. As a preliminary approach all medical terms in the clinical scenario was collected to create query string and is ANDeD when searching MEDLINE®. Search terms from the clinical scenarios are used by RIDE-M to search MEDLINE® and fetch search results. These search results are then evaluated and ranked according to the clinical relevancy and quality of articles on a scoring scale of one to five. An evaluation of RIDE-M was also integrated into the query evaluator to answer whether the RIDE-M summary represented the MEDLINE® abstract. The clinical scenario and the associated question writer’s clinical expertise might have biased the clinical terms generated for searching MEDLINE®.

Results for the preliminary model:
Out of 476 clinical scenarios/question combinations 338 queries were created. (227) 67% queries generated no results and (111) 32.8% of queries generated one or more result. Out of the 111 queries generated, (71) 64% had results where one or more result answered the clinical question. 417 citations were extracted for the 111 queries with results. (Average 3.7 citation per query with result). Out of the 417 citations (47%) 195 of them answered the clinical question. 185 of these citations were ranked more than three in relevancy score.

Conclusions and research strategies:
In this experiment an average of 3.7 citations were generated per query. Out of the total citations 47% answered the clinical question. Unlike an EMR, the surrogate clinical scenarios we used do not represent real patient records either in content or volume. The proportion of total search terms created using these clinical scenarios to total number of words in the respective clinical scenario is not evaluated in this experiment. This type of evaluation will help to formulate projections when extracting search terms from EMR. Filters need to be devised to extract, standardize and classify diagnostic data in the EMR as well as to generate queries. Clinical informatics experiments using content from real EMR is necessary to achieve the ultimate goal of automatic augmentation of the EMR with MEDLINE®. Contextual information and user defined search terms should be used to assist in attaining accuracy in the initial stages of research. Experiments using EMRs and MEDLINE® resources have the potential to unravel new associations and patterns among different clinical entities separated in time and space. This process can be refined to utilize as a decision support tools using MEDLINE® resources enabling clinicians to use NLM resources at the point of care.

Acknowledgement:
Communication and Engineering Branch, Lister Hill National Center for Biomedical Communication, National Library of Medicine, Bethesda, Maryland.

References: