Tailoring Online Information Retrieval to User’s Needs based on a Logical Semantic Approach to Natural Language Processing and UMLS mapping

Susan Kossman, RN, PhD1,3, Josette Jones, RN, PhD4 & Patricia Flatley Brennan, RN, PhD1,2
University of Wisconsin-Madison, School of Nursing1 and College of Engineering2
Illinois State University3, Indiana University School of Nursing and School of Informatics4

Abstract:
Depression can derail teenagers’ lives and cause serious chronic health problems. Acquiring pertinent knowledge and skills supports care management, but retrieving appropriate information can be difficult. This poster presents a strategy to tailor online information to user attributes using a logical semantic approach to natural language processing (NLP) and mapping propositions to UMLS terms. This approach capitalizes on existing NLM resources and presents a potentially sustainable plan for meeting consumers’ and providers’ information needs.

Problem and significance
Effective management of depression requires a knowledge and skill set to support behavior choice. Retrieving appropriate high quality, current, online health information at the point of need can be problematic. Tailoring healthcare information to users’ learning needs, abilities and preferences increases its utility (1). Retrieving tailored online information requires effective coding of information and algorithms matching content to users’ needs.

NLP methods code information in unstructured text, representing the knowledge in computable form for information retrieval. Common NLP methods include parsing text by word strings or noun phrases. Problems arise from ambiguity, redundancy and abbreviations. Attending to context can increase accuracy, precision and sensitivity (2). A logical semantic NLP approach parses at the sentence level, then uses string matching and probabilistic likelihood to map sentences to unique codeable propositions which represent their meaning (3). Propositions reside in a semantic knowledge base that represents the range of concepts in a particular domain. Mapping propositions to UMLS concepts and terms broadens their information retrieval potential.

Purpose:
To develop a system for retrieval of pertinent, current online healthcare information at the point of need tailored to consumer’s and healthcare provider’s information needs and learning preferences that is based on semantic propositions mapped to MeSH, SNOMED CT and selected UMLS terminologies.

Methodology
We are building a knowledge base of semantic propositions using a logical semantics approach to NLP of web-based information resources related to management of depression in adolescents. We identified a base set of high quality webpages through focused searching of PubMed and Medline plus resources. Propositions describe semantic meaning of page content. Propositions describing user attributes include role, reading level, learning style and preferred level of detail. Logical semantic software performs regular webpage searches to identify semantic propositions contained in online resources. The software presents hypothesized matches for researcher validation and addition to the knowledge base, increasing both the body of resources available for tailoring and the accuracy of proposition matching.

To increase accuracy, sensitivity and precision of information retrieval, propositions will be mapped to a set of UMLS terminologies relating to nursing and psychology. Both semantic propositions and mapped terms will be used in a tailoring algorithm. Propositions will also be mapped to MeSH and SNOMED CT and an algorithm developed to trigger automatic updates to the knowledge base based on matching key words in new articles published in PubMed and Medline plus.

Pilot testing of the tailoring algorithm with six mock users will compare pages retrieved with a test set of pages. Knowledge base updating will be tested by comparing triggered additions with manual searches. Iterative testing and refinement of the algorithms will occur to achieve high accuracy precision and sensitivity. The system will be used in the Blue Sky project aimed at improving adolescent depression management.

References

Acknowledgements:
NLM F37 LM009063 funding