1-1-2002

**MedTextus: An Intelligent Web-Based Medical Meta-Search System**

Bin Zhu
*University of Arizona*

Gondy Leroy
*Claremont Graduate University*

Hsinchun Chen
*University of Arizona*

Yongchi Chen
*University of Arizona*

---

**Recommended Citation**

MedTextus: An Intelligent Web-Based Medical Meta-Search System

Bin Zhu¹, Gondy Leroy¹, Hsinchun Chen¹, and Yongchi Chen²
Management Information Systems Department
University of Arizona
Tucson, AZ 85721 USA
+1 520 626 9239
¹{bzhu, gleroy, hchen}@bpa.arizona.edu, ²mark@ai.bpa.arizona.edu

ABSTRACT
We propose to demonstrate a web-based prototype system that integrates the meta-search approach with existing information analysis and visualization technologies to facilitate concept-based searching behavior over the medical domain. The system distinguishes itself from other meta-search engines through two features. It incorporates the co-occurrence analysis and existing ontology to understand user’s query. It also utilizes the self-organizing map (SOM) to categorize and visualize search results.

Categories & Subject Descriptors: H.3.3 Information Search and Retrieval

General Terms: Algorithms, Design

Keywords: Meta search, categorization, concept space, visualization

INTRODUCTION
The emergence of meta-search engines provides a credible resolution of divergence by triangulating output from several engines. By sending queries to multiple search engines and collating only the highest-ranking subset of the returns, meta-search engines can greatly improve search results. However, a meta-search engine faces two challenges faced by any search engine. The first is to solve the semantic ambiguity of a user’s query. Another challenge is to relieve the information overload caused by browsing the large amount of information returned for one query. The system proposed integrates existing information analysis and visualization technologies to facilitate the query specification and to summarize the search results.

MedTextus System
The MedTextus system consists of three parts: the Concept Mapper, the meta spider, and the summarizer.

- **Concept Mapper** utilizes both the Concept Space and ontology to help users specify their information needs. The Concept Space is an automatically generated index similar to a human generated thesaurus, but it is based on document term co-occurrence analysis (Chen & Lynch, 1992). The related terms it provides can be used for term suggestion or for query expansion. Ontology provides consistent vocabularies and word representations necessary for clear communication within knowledge domains. The Medical Concept Mapper processes a user’s query by presenting synonyms from two types of ontology and suggesting related term from the Concept Space. A user can generate a more specific query by selecting terms from the list suggested.

- **Meta Spider** sends the query to several popular and authoritative databases of medical literature including MEDLINE, CANCERLIT, PDQ, MERCK, ACP, NGC, and DARE. It also fetches all related documents returned by those databases and filters duplicates.

- **Summarizer** categorizes the documents returned and visualizes the category generated. It first uses Arizona Noun Phraser (AZNP) to represent the content of a document. AZNP is a natural language based index tool that extracts noun phrases from textual documents (Tolle & Chen, 2000). A subset of noun phrases extracted was selected to represent documents according to the phrase frequency and the document frequency. Then the Summarizer utilizes the SOM to categorize documents returned. SOM is defined as a mapping from a high dimensional input space into a two-dimensional array of output nodes, where spatial proximity represents the semantic proximity (Kohonen, 1995). Such 2-dimensional output makes SOM an ideal candidate for visualizing the categories generated. The map summarizes different aspects of the document collection related to a query.

REFERENCES