

Customizable and Ontology-Enhanced Medical Information Retrieval Interfaces
G. Leroy, K. M. Tolle, H. Chen
Management Information Systems Department, University of Arizona, Tucson, USA

Abstract

This paper describes the development and testing of the Medical Concept Mapper as an aid to providing synonyms and semantically related concepts to improve searching. All terms are related to the user-query and fit into the query context. The system is unique because its five components combine human-created and computer-generated elements. The Arizona Noun Phraser extracts phrases from natural language user queries. WordNet and the UMLS Metathesaurus provide synonyms. The Arizona Concept Space generates conceptually related terms. Semantic relationships between queries and concepts are established using the UMLS Semantic Net. Two user studies conducted to evaluate the system are described.

Keywords: Medical Information Retrieval, Ontologies, UMLS, Deep Semantic Parsing

1 Introduction

The Medical Concept Mapper described in this paper is a new tool which combines the powerful Unified Medical Languages System (UMLS) developed by the National Library of Medicine (NLM) with automated computational search space tools developed by the AI Lab at the University of Arizona. The Medical Concept Mapper consists of the Arizona (AZ) Noun Phraser with Specialist Lexicon, WordNet, the UMLS Metathesaurus and Semantic Net, and Concept Space. This system is innovative because it is an in-depth integration of manually created ontologies and computer generated tools, the intertwining of which allows a synergy to surface that surpasses the weaknesses and strengths of each tool when used on its own.

The structure of this paper is as follows: First, the components used in our new system are discussed. Their implementation, structure and the difficulties encountered when using them are described. Then the actual integration of all components, the algorithms and two user studies, are presented. The last section gives an overview of how this system will be integrated with other existing components into a medical MetaSpider.

2 Background

2.1 *The Arizona Noun Phraser*

Noun phrasing is the extraction of noun phrases from free text. It has been used in information retrieval to capture a “richer linguistic representation” of document content [1]. It has the potential to improve precision over other document indexing techniques, since it allows for multi-word queries to be matched with words present in the text documents [2].

The AZ Noun Phraser was developed at the University of Arizona AI Lab to extract high-quality phrases from textual data [3]. It has three components: a tokenizer, a tagger, and a noun phrase generator. The tokenizer module is designed to take raw text input and create output that conforms to the UPenn Treebank word tokenization rules. Its task is to separate all punctuation and symbols from text without interfering with textual content. The tagger module is a significantly revised version of the Brill tagger [4]. Our initial implementation included two existing multi-domain corpora: the Wall Street Journal Corpus and the Brown Corpus. The version used for this experiment also includes the UMLS Specialist Lexicon, which has previously been shown to improve the extraction of medical phrases from text [3]. The third module is the phrase generator, which converts the words and associated part-of-speech tags generated by the tagger into noun phrases. The phraser utilizes a pattern-matching algorithm to isolate phrases from the tagged text [3].

2.2 *Ontologies*

Ontologies provide consistent vocabularies and world representations necessary for clear

communication within knowledge domains. A frequently used definition of 'ontology' is that it is '*an explicit representation of a conceptualization*' [5-7]. Ontologies range from very general, e.g. WordNet [8], EuroWordNet [9], Cyc [10], to very domain-specific, e.g. the Enterprise Ontology [11] for business communication. The UMLS is an example of an extensive and specific ontology. Both the UMLS and WordNet will be discussed, since they will be used by the Medical Concept Mapper.

2.2.1 WordNet

WordNet is an on-line accessible lexical database that contains approximately 95,600 different word forms that are organized into word senses [8]. The word senses are very detailed. For each of them, a descriptive gloss and a set of synonyms is provided. The total lexicon contains four word categories: nouns, verbs, adjectives, and adverbs. There are about 57,000 nouns. The lexicon can be accessed online or downloaded free at <http://www.cogsci.princeton.edu/~wn/>. It covers different relations such as synonymy, antonymy, meronymy, and hyponymy. The relations are returned for all word senses of a term.

It is difficult to select the correct word sense in a given instance and incorrect word sense disambiguation will interfere with information retrieval [12-14]. The noun "head", for example, has 30 different senses. However important it may be to add synonyms to a query, including those that are irrelevant can have a very negative effect.

2.2.2 UMLS

The UMLS is a long-term project of the NLM [15] developed to enable new information technologies to take advantage of controlled medical vocabularies [15-17]. The knowledge it contains, its definitions, concepts and structure are used in a variety of applications. One possible use is mapping user queries to relevant retrieved information [18]. It is sometimes used purely as a knowledge base for other medical tools. Carenini and Moore [19] extracted the knowledge contained in the relations of the Semantic Net and used it for their patient education system. Others use its concepts for Web searches [20].

The UMLS consists of four components: the Metathesaurus, the Semantic Net, the Specialist Lexicon, and the Knowledge Sources Server. Only the Metathesaurus and Semantic Net will be described here, since they are used by the Medical Concept Mapper.

The goal of the Metathesaurus is to link terminology and underlying concepts [15]. It combines the vocabulary of more than 50 different sources - including Mesh terminology - into one consistent set [21]. This is one of the main advantages of the Metathesaurus, since retrieval is higher when terminology is not strictly limited to Mesh terms [22]. Aronson and Rindflesh [23] used the UMLS Metathesaurus for query expansion. They mapped queries to concepts in the Metathesaurus which improved retrieval.

The 1999 release of the Semantic Network contains 134 semantic types and 54 relationships [21]. It contains relations between concepts and can be of help with query interpretation [16]. Semantic types are also used in a system currently under development by Pratt [24], who uses them to relate the terms found in documents to those in the original query. Based on these relations, the documents are categorized. So far, completely automated tools that use the Semantic Net are scarce because of its structure: Semantic relations exist between semantic types, not between the concepts that belong to those types. For example, the semantic type "Medical Devices" has a "treat" relation with the semantic type "Sign and Symptom". However, not every concept belonging to "Medical Device" will "treat" every concept belonging to "Sign and Symptom": Bone screws do not treat nausea.

There are different approaches to these problems. Cimino et al. [25] use predefined generic queries. User queries are mapped to their equivalent generic query. Once the queries are matched, the appropriate information sources are selected and the information is retrieved. The advantage is that all UMLS Knowledge Sources can be optimally used if a user query is correctly mapped to a generic query. The disadvantage is that this tool is based on a limited set of generic queries and a good match between user and generic query is necessary. In another approach, users build the structure of the query by selecting the concepts, their semantic types and the semantic relations between those concepts [26, 27]. The result

is a conceptual graph for a query. These graphs can be compared against graphs of the information source, e.g. patient records, to find valid matches. The advantage of this approach is that potential matches can be limited based on their underlying structure. The disadvantage is that end-users have to decide on the medical relevance of their graphs. Usage will be limited to experts with the necessary knowledge of medical concepts and the validity of relations between them.

2.3 *Concept Space*

Concept Space was generated to facilitate semantic retrieval of information. In several studies, Concept Space improved searching and browsing. In the bio-sciences, Concept Space was successfully applied to the Worm Community System [28, 29] and the FlyBase experiment [30]. There also have been successful results in the Digital Library Initiative studies conducted on the INSPEC collection for computer science and engineering [29,31] and on Internet searching [32].

There are five major steps involved in creating and using Concept Space. During *Document Analysis*, a collection of documents is analyzed to separate free text from fields, which contain the names of authors, publication information, and document identification numbers. During the *Concept Extraction* phase, concepts are extracted based on automatic indexing techniques [29, 31, 33], or noun phrasing [3, 34]. During *Phrase Analysis*, phrase and document frequencies are computed. *Phrase Co-occurrence Analysis* is used to generate the phrase thesaurus. It is based on the asymmetric "Cluster Function" developed by Chen and Lynch [35]. During *Associative Document Retrieval*, term suggestion techniques help users refine queries. They invoke spreading activation algorithms for multiple-term, multiple-link phrase suggestions. The user can use these terms to select documents for display.

3 **Research Question and System Design**

Our purpose is to integrate all the previously discussed components and present users with a system that can take a medical query in any form and provide synonyms and important related (but not synonymous) concepts for the terms in the queries. Our research question is the following: How well can the components discussed above be used to expand a user query with appropriate medical concepts. This has been tested in two user studies.

The Medical Concept Mapper involves three consecutive phases. In the first phase the concepts are extracted from the given natural language query. The second phase consists of adding synonyms to these terms. Here, the two ontologies are used. WordNet provides synonyms for each term if there is only one synset for that particular term. As such, its power is cut but precision should not suffer. The second ontology is the UMLS Metathesaurus. For each term, the concept to which it belongs is retrieved. All terms representing that particular concept are regarded as synonyms. The third phase consists of mapping semantically related concepts to the original query. For this phase, Concept Space is used to suggest related terms, employing the Semantic Net to limit these terms by means of Deep Semantic Parsing (DSP).

The DSP algorithm itself contains three different steps. In the first step, the context of the original query is established by means of semantic types. The Semantic Net provides these types. The second step limits all concepts suggested by Concept Space based on this query context. The algorithm allows only terms having certain semantic relations and types relevant to the query context. This results in a concept set in which the concepts are semantically related to the initial query. Terms not found in the UMLS are not excluded from the expansion process. That would limit the vocabulary to the Metathesaurus lexicon, which is not a substitute for a complete lexicon [36]. During the last step, all concepts that are retained are re-ordered based on their Concept Space weights, and the most relevant subset is selected.

The ambiguity inherent in the relation between concepts and semantic types is eliminated by using the Semantic Net for limitation, not for expansion. Terms' being related in Concept Space means that they co-occur frequently. In this case, the chances are very slim that the relation between their semantic types would not be valid. Since the relations can be assumed to be valid, they are used for term limitation. For example if "aspirin" and "headache" are strongly related in a Concept Space, this would be a strong

indication that “Aspirin” (Pharmacologic Substance) "treats" (Semantic Relation) “headaches” (Sign or Symptom).

4 User Studies

4.1 Design

Thirty real and cancer-related user queries were submitted to the system. The queries came from three sources. The first was a set of more than 1500 queries generated by medical doctors for usage with the UMLS [37]. Medical librarians submitted additional queries via e-mail. Two queries came from the literature [38]. For each query, a set of terms representing concepts of importance related to the query was put together by two expert groups: medical librarians and cancer researchers. As such, there were two Golden Standards for each query. None of the experts had seen the Medical Concept Mapper or any of its output. The results were evaluated by comparing the system’s output with the Golden Standards from the experts. Precision and Recall were calculated based on this comparison.

All queries were submitted in three different ways. First they were submitted in their original state. Second, they were submitted as cleansed queries that had had spelling corrected and unnecessary *conversational* information omitted. Third, the queries were submitted by means of search terms that represented each query. These terms were extracted directly from the query and were not altered in any way. Both the cleansing of the queries and the selection of relevant search terms were done by a medical expert.

In the first user study, three sets of synonyms were extracted. The first set consisted of WordNet synonyms added to the search terms. The second set consisted of UMLS synonyms added to the search terms, without any WordNet synonyms. The third set consisted of WordNet synonyms and UMLS synonyms. In this last case, the WordNet synonyms were used as input to find UMLS synonyms. In the second user study, queries were expanded by attaching semantically related concepts. The best synonym condition (determined in the first experiment) was used as the baseline. Two different expansion methods were tested. The first method employed Concept Space terms directly, without any limitation being imposed by the DSP algorithm. The second method used DSP.

4.2 Results

4.2.1 Golden Standard

There was an enormous difference between the two standards. The medical librarians gave long lists of terms, averaging 17.6 terms per query. They included many synonyms and spelling variations. The cancer researchers provided a smaller number of terms, averaging of 6.1 terms per query. They stipulated that any deviation (e.g., plural for a singular term) was not acceptable.

4.2.2 User study 1: Synonyms

All 30 queries were submitted to the system, and the results were compared to the two golden standards. With medical librarians, the conceptual relevance of terms was checked by a medical expert, which meant that small deviations (spelling, punctuation and some more specific synonyms) were acceptable. Using the standard provided by the cancer researchers, the strings had to be an exact match.

A. Medical Librarians’ Standard

Recall improved with expansion. For the Original Queries, recall was 14% with and without additional WordNet synonyms. Recall increased to 25% when the Metathesaurus was used and to 26% when WordNet was used to leverage the Metathesaurus. Similar results were found for the three input methods. Recall also benefited from cleaner input. Term Input resulted in higher recall than Original and Cleansed Queries.

Precision was affected by the input method, but not by expansion. For Original Queries, it was 54% when there was no expansion, 53% when WordNet synonyms were added, 59% when Metathesaurus synonyms were added, and 58% when WordNet was used to leverage the Metathesaurus. Similar results

were found for Cleansed Queries and Term Input. Precision was higher for Term Input compared to both Original and Cleansed Queries. For example, when no expansion was done, the precision was 54% for Original Queries, 57% for Cleansed Queries and 92% for Term Input.

B. Cancer Researchers' Standard

Recall improved with cleaner input. For example when no expansion was done, recall was 22% for Original Queries, 23% for Cleansed Queries, and 31% for Term Input. Expansion had no effect on recall.

Precision dropped when synonyms were added. For Original Queries, precision was 29% when no expansion was done. It was 6% with the Metathesaurus and 5% with both the Metathesaurus and WordNet. Precision was high for Term Input when there was no expansion (59%). It dropped to the same levels as the other input methods when synonyms were added.

4.2.3 User study 2: Semantically Related Concepts

A. Medical Librarians' Standard

Recall improved with expansion and was not affected by Input Method. For Original Queries, recall was 25% for the baseline (Syns), 30% for Concept Space without DSP, and 30% with DSP.

Precision dropped when no DSP was used in the expansion. For example, for the Original Queries, precision was 59% for the baseline (Syns). When Concept Space terms were added, precision was 46% without and 52% with DSP. Precision was also higher for Term Input. For the baseline (Syns), precision was 59% for Original Queries, 60% for Cleansed Queries and 79% for Term Input.

B. Cancer Researchers' Standard

Recall benefited from cleaner input but not from expansion. *Precision* was affected by concept expansion, and by the input method. Precision dropped with expansion. It was somewhat higher for Term Input than for the other two input methods. For example, looking at the baseline (Syns), precision for Term Input was 11%, and 6% for both the Original and Cleansed Queries.

5 Discussion of Results

In general, we can say that we successfully combined human-created ontologies and our computer-generated tools. The Medical Concept Mapper can automatically double the number of useful search terms extracted from queries. It can also suggest related terms with high precision. This is helpful for users interested in using alternative terms in their search.

It is clear from our results that not all users use many terms when they are searching for information. The two expert groups differed in how they performed searches. The medical librarians started out with an extensive list of terms and their possible synonyms and spelling variations. The cancer researchers used a limited number of terms. If we think of search strategies (local/global) as a continuum, it seems that our two groups were at the extremes. A search with the medical librarians' terms could result in an all-encompassing list of documents that probably would have to be narrowed down. A search with the cancer researchers' terms would not always result in documents being found, in which case other synonyms would have to be used. These differences had a clear impact on the evaluation of the Medical Concept Mapper. When no synonyms were included in the Golden Standard, a system explicitly built to provide these and other terms, did not perform as well on precision as when the medical librarians' standard was used for evaluation. In addition, terms provided by the Medical Concept Mapper that do not appear in the Golden Standard may not necessarily have been incorrect terms. This needs further evaluation.

Another interesting observation is that we found no differences between Original and Cleansed Queries. The system is robust enough to deliver the same recall and precision regardless of the format of the query. It had been expected that the unnecessary information in the Original Queries would lead to a major number of irrelevant terms and concepts. This did not happen.

6 Future Directions

Having developed a number of medical information access tools, the AI Lab has processed increasingly larger document collections ranging from specialized document sets of 7500 documents for AIDS and epilepsy to the entire MEDLINE collection of nearly 10 million documents [39]. We started by generating concept space applications for data access to the collections and enabling query refinement to reduce the document space required. The Concept Space for CANCERLIT used by the Medical Concept Mapper also provides online access to more than 750K documents in the National Cancer Institutes bibliographic collection. The URL to access this system is <http://ai4.bpa.arizona.edu/cgi-bin/tng/npj2>. Concept Space provides the user with a document index and term thesaurus. The document index, built on the collection, allows for fast real-time access to the documents in the collection. Users can choose to display the text of the documents, or look at the bibliographic information. They can cluster and visually display the document collection using the Dynamic SOM (DSOM).

In the DSOM's display, similar categories of documents are clustered together in regions on the map. Each region represents an important topic or theme in the underlying collection. Similar topics appear in the same neighborhood, thus creating a cognitively intuitive display and summary for users. Colors represent distinct topics and the height of a cylinder represents the number of documents grouped under a topic. The label of each category is generated automatically. It is the most representative term for documents appearing in a specific region.

We are currently developing a medical MetaSpider. This tool will be used to gather information interactively during user sessions or to gather updates periodically when changes occur in the monitored collections. At the time of this writing we have integrated National Library of Medicine's PDQ and MEDLINE sites along with National Cancer Institute's CANCERLIT. For clinicians we will also include in future versions evidence-based medicine (EBM) meta-information sites such as Database of Abstracts of Reviews of Effectiveness (DARE), Cochran Abstracts, Bandolier and NLM's HSTAT. Users will enter a natural language query or multi-word search terms and spiders will search for relevant documents from the various data sources and on the web. The user can choose to immediately view the documents or can categorize the results visually by invoking the Dynamic SOM (DSOM), which has been successfully used for categorizing and visually displaying retrieved document sets [40, 41]. The goal of the AI Lab is to combine proven medical information retrieval tools in a complete customizable search interface. This evaluation of the Concept Mapper shows that it is a useful tool for integration with Concept Space and the Medical MetaSpider described above. It will be used for high-precision term suggestion or for automated query expansion, depending on the preferences of the user.

7 Acknowledgements

This project was supported in part by the following grants:

NSF/ARPA/NASA Digital Library Initiative, IRI-9411318, 1994-1998 (B. Schatz, H. Chen, et al., Building the Interspace: Digital Library Infrastructure for a University Engineering Community'), NSF CISE, IRI-9525790, 1995-1998 (H. Chen, Concept-based Categorization and Search on the Internet: A Machine Learning, Parallel Computing Approach),

National Library of Medicine (NLM) Toxicology and Environmental Health Research Participation Program through the Oak Ridge Institute for Science and Education (ORISE), 1996-1997.

Additional funding and support were provided by the National Cancer Institute and the National Institutes of Health.

From the University Medical Center, we thank Margaret Briehl and Steven Stratton for devoting their time to making the cancer researchers' standard. From the Health Sciences Library we thank Gerald Perry and Jeffrey Middleton for making the medical librarians' standard, Rachael Anderson for her continued support and Robin Sewell for evaluating queries and terms. We also want to thank Alexa McCray at the National Library of Medicine for her continued support.

We want to thank the National Library of Medicine and Princeton University for making the UMLS and WordNet ontologies freely available to researchers.

8 References

1. Anick PG, Vaithyanathan S.
Exploiting Clustering and Phrases for Context-based Information Retrieval. In: 20th Annual International ACM SIGIR Conference on Research and Development; 1997: 314-23.
2. Girardi MR, Ibrahim B.
An Approach to Improve the Effectiveness of Software Retrieval. In: Richardson DJ, Taylor RN, eds. 3rd Annual Irvine Software Symposium; 1993: 89-100.
3. Brill E.
A Corpus-Based Approach to Language Learning [Ph.D. Dissertation]. Philadelphia: University of Pennsylvania; 1993.
4. Tolle KM, Chen H.
Comparing noun phrasing techniques for use with medical digital library tools. Journal of the American Society of Information Systems 1999;"in press".
5. Farquhar A, Fikes R, Rice J.
The Ontolingua Server: a tool for collaborative ontology construction. Int. J. Human-Computer Studies 1997;46:707-27.
6. Gruber T.
A translation approach to portable ontology specifications. Knowledge Acquisition 1993;5 (2):199-220.
7. Grüninger M, Uschold M.
Tutorial SA1, Ontologies: Principles, Applications and Opportunities. In: Thirteenth National Conference on Artificial Intelligence; 1996;166-137.
8. Fellbaum C.
WordNet : An Electronic Lexical Database. Cambridge, Mass: MIT Press; 1998.
9. Rodríguez H, Climent S, Vossen P, Bloksma L, Peters W, Alonge A, et al.
The top-down strategy for building Eurowordnet: Vocabulary coverage, base concepts and top ontology. Computers and the Humanities 1998;32:117-52.
10. Cycorp.
Cyc. [Http://www.cyc.com](http://www.cyc.com)
11. Uschold M.
Converting an Informal Ontology into Ontolingua: Some Experiences. In: ECAI '96 Workshop on Ontological Engineering; 1996; 1-17.
12. Mock KJ, Vemuri VR.
Information filtering via Hill Climbing, WordNet, and Index Patterns. Information Processing & Management 1997;33(5):633-44.
13. Stairmand MA.
Textual Context Analysis for Information Retrieval. In: 20th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval; 1997: 140-47.
14. Voorhees EM.
Using WordNet for Text Retrieval. In: Fellbaum C, ed. *WordNet: An Electronic Lexical Database*. Cambridge, MA: The MIT Press; 1998:285-303.
15. Lindberg DA, Humphreys BL, McCray AT.
The Unified Medical Language System. Methods Inf Med 1993;32(4):281-91.
16. McCray AT, Nelson SJ.
The representation of meaning in the UMLS. Methods Inf Med 1995;34(1-2):193-201.
17. Humphreys BL, Lindberg DA.
The UMLS project: making the conceptual connection between users and the information they need. Bull Med Libr Assoc 1993;81(2):170-77.
18. McCray AT, Aronson AR, Browne AC, Rindfleisch TC, Razi A, Srinivasan S.
UMLS knowledge for biomedical language processing. Bull. Med. Libr. Assoc. 1993;81(2):184-94.

19. Carenini G, Moore JD.
Using the UMLS Semantic Network as a Basis for Constructing a Terminological Knowledge Base: A Preliminary Report. In: Seventeenth Annual Symposium on Computer Applications in Medical Care; 1994.
20. Suarez HH, Hao X, Chang IF.
Searching for information on the internet using the UMLS and Medical World Search. In: AMIA Annual Fall Symposium; 1997: 824-28.
21. NLM.
The Unified Medical Language System. [Http://umlsks.nlm.nih.gov](http://umlsks.nlm.nih.gov)
22. Rindflesch TC, Aronson AR.
Ambiguity resolution while mapping free text to the UMLS Metathesaurus. In: The American Medical Informatics Society Annual Symposium on Computer Applications in Medical Care; 1994: 240-44.
23. Aronson AR, Rindflesch TC.
Query Expansion Using the UMLS Metathesaurus. In: AMIA Annual Fall Symposium; 1997; 1997. p. 485-89.
24. Pratt W.
Dynamic organization of search results using the UMLS. In: AMIA Annual Fall Symposium; 1997: 480-84.
25. Cimino JJ, Aguirre A, Johnson SB, Peng P.
Generic queries for meeting clinical information needs. *Bull Med Libr Assoc* 1993;81(2):195-206.
26. Robert JJ, Joubert M, Nal L, Fieschi M.
A computational model of information retrieval with UMLS. *Proc Annu Symp Comput Appl Med Care* 1994:167-71.
27. Joubert M, Fieschi M, Robert JJ.
A conceptual model for information retrieval with UMLS. In: Seventeenth Annual Symposium on Computer Applications in Medical Care; 1994: 715-19.
28. Chen H, Schatz BR, Yim T, Fye D.
Automatic thesaurus generation for an electronic community system. *Journal of the American Society for Information Science* 1995;46(3):175-93.
29. Chen H, Martinez J, Ng DT, Schatz BR.
A concept space approach to addressing the vocabulary problem in scientific information retrieval: An Experiment on the Worm Community System. *Journal of the American Society for Information Science* 1997;48(1):17-31.
30. Chen H, Schatz BR.
Semantic Retrieval for the NCSA Mosaic. *Proceedings of the Second International World Wide Web Conference* 1994.
31. Chen H, Martinez J, Kirchhoff A, Ng TD, Schatz BR.
Alleviating search uncertainty through concept associations: automatic indexing, co-occurrence analysis, and parallel computing. *Journal of the American Society for Information Science* 1998;49(3):206-16.
32. Chen H, Houston A, Yen J, Nunamaker JF.
Toward intelligent meeting agents. *IEEE Computer* 1996;29(8):62-70.
33. Chen H, Ng DT.
an algorithmic approach to concept exploration in a large knowledge network (automatic thesaurus consultation): symbolic branch-and-bound vs. connectionist Hopfield net activation. *Journal of the American Society for Information Science* 1995;46(4):348-69.
34. Houston AL, Chen H, Schatz BR, Hubbard SM, Sewell RR, Ng TD.
Exploring the use of concept space to improve medical information retrieval. *International Journal of Decision Support Systems*, "in press".

35. Chen H, Lynch KJ.
automatic construction of networks of concepts characterizing document databases. IEEE Transactions on Systems, Man and Cybernetics 1992;22(5):885-902.
36. Johnson SB, Aguirre A, Peng P, Cimino J.
Interpreting natural language queries using the UMLS. Seventeenth Annual Symposium on Computer Applications in Medical Care. 1993:294-98.
37. Stavri Z.
Queries generated by medical doctors for usage with the UMLS. Personal Communication.
38. Hersh WR, Hickam DH.
An evaluation of interactive Boolean and natural language searching with an online medical textbook. Journal of the American Society for Information Science 1995;46(7):478-89.
39. Chung Y, He Q, Powell K, Schatz B.
Semantic Indexing for a complete Subject Discipline. In: The fourth ACM Conference on Digital Libraries; 1999: 39-48.
40. Chen H, Schuffels C, Orwig R.
Internet categorization and search: A machine learning approach. Journal of Visual Communications and Image Representation 1996;7(1):88-102.
41. Orwig R, Chen H, Nunamaker JF.
A graphical, self-organizing approach to classifying electronic meeting output. Journal of the American Society for Information Science 1997;48(2):157-70.

Address of Author:

Gondy Leroy
University of Arizona, Management Information Systems
McClelland Hall, Room 430
1130 E. Helen Str.
Tucson, AZ 85721
USA
Phone: (520) 626-9239
Fax: (520) 621-2433
Email: gleroy@bpa.arizona.edu