Semantic Retrieval of Medical Records Related to Patient Symptoms[1]

Hemant Jain, Huimin Zhao
Sheldon B. Lubar School of Business
University of Wisconsin–Milwaukee
jain@uwm.edu, hzhao@uwm.edu

David P. Klemer
College of Engineering and Applied Science
University of Wisconsin–Milwaukee
dklemer@uwm.edu

Carmelo Gaudioso
Medical Informatics Program
University of Wisconsin–Milwaukee
cgaudioso@sbcglobal.net

Abstract

There are currently many active movements towards computerizing patient healthcare information, which has been widely acknowledged as being both crucial and long overdue. As Electronic Medical Record (EMR) systems become widely adopted and deployed in the foreseeable future, the next big challenge will be the development of methods to effectively utilize this massive information source. Based on the observation that simple text-word based information retrieval fails to yield satisfactory performance—due to the complex semantic relationships among symptoms and diagnostic/therapeutic processes—we propose a framework for developing an intelligent information retrieval system which can serve as an adjunct to a unified EMR system. The proposed framework integrates various technologies, including information retrieval, domain ontologies (e.g., the Unified Medical Language System), automatic semantic relationship learning (e.g., term co-occurrence), as well as a body of domain knowledge elicited from healthcare experts. This will allow effective retrieval of healthcare records related to a patient's presenting signs and symptoms. Knowledge of semantic relationships among medical concepts, such as symptoms, exams and tests, diagnoses, and treatments, as well as knowledge of synonyms, hypernym/hyponyms, is used to expand and enhance initial queries posed by a user. Such a semantic retrieval system can liberate doctors from the daunting task of digging into voluminous medical records for a few relevant documents, with a consequent improvement in healthcare efficiency and quality. We will discuss the major challenges anticipated and research issues, and will outline our research plan in the paper.

Keywords: healthcare, electronic medical records, information retrieval, domain knowledge

1. Introduction

Having electronic medical records (EMR) in doctor's offices and hospitals is both crucial and long overdue (Kolata 2005). There are currently many active movements towards this end. The pressing need for effective healthcare information systems has recently resulted in "The Decade of Health Information Technology: Delivering Consumer-centric and Information-rich Health Care", initiative announced by President Bush in April 2004 (Thompson and Brailer 2004). Medicare is planning to provide its EMR software, which is being used by the military and the Department of Veterans Affairs, to all doctors free of charge (Kolata 2005). Many other related initiatives are currently underway. Some examples are the Master Patient Index (MPI) research funded by the National Institute of Standards (Bell and Sethi 2001) and the Integrating the Healthcare Enterprise (IHE) initiative jointly sponsored by the Radiological

---

[1] An earlier version of this paper has been presented at the WeB 2005 conference, Las Vegas, Dec. 10th 2005.

Society of North America (http://www.rsna.org) and the Healthcare Information and Management Systems Society (http://www.himss.org). Similar efforts are being made in many other parts of the globe, for example, the Integration Broker of Healthcare Information Sources (IBHIS) project in the UK (Kotsiopoulos et al. 2003; Zhu et al. 2004) and the Co-operative Health Information Networks (CHIN) developed in several European countries, including Finland, Germany, Greece, Spain, Sweden, and the UK (Tambouris et al. 2000).

As EMR system becomes a reality, an important challenge is to effectively utilize this massive information source. The difficulty first lies in the enormous size of the information source and is further amplified by the largely skewed distribution of healthcare information across patients; over 90% of the health records belong to 10% of senior patients who are frequent users of the healthcare system. It is still a daunting task for doctors to dig into the voluminous medical records and identify the few relevant documents based on a patient's symptoms, especially for patients with very long healthcare histories. An intelligent information retrieval system that can help doctors quickly sift through the medical records and identify the documents relevant to a patient's current symptoms will be extremely valuable.

Two commonly-encountered scenarios will illustrate this need. (1) A physician in the emergency department may be encountering a patient for the first time, presenting with shortness-of-breath due to fluid overload. The patient's relevant medical history will have a significant impact on the approach to diagnosis and treatment -- is this the first manifestation of these symptoms for this patient, caused by acute heart failure of unclear origin? Or, is this the patient's fifteenth visit over the past 6 months, resulting from suboptimal dosing of his cardiac medications and/or his unwillingness to follow his doctor's low-sodium-diet instructions? Of course, the patient may be able to supply some history, but oftentimes a patient's recollection of his past medical history and treatment may not be as reliable as the medical record itself. (2) A family practitioner may be seeing a new patient, recently transferred from a different health plan, who needs immediate refills for a long list of heart medications of which she has incomplete details. In this case, the physician has immediate need of a very specific subset of the patient's medical record, i.e., the medications and doses which the patient is currently taking.

This paper presents a framework for developing an intelligent information retrieval system on top of a unified EMR system. Assuming that patient history data have been integrated into a unified EMR system, such an information retrieval system helps care providers (e.g., family practitioners, emergency department staff, and others) effectively and efficiently retrieve the relevant documents (test reports, diagnoses, etc.) given a query based on a patient's symptoms. This system may require several types of techniques, including word based information retrieval (Baeza-Yates and Ribeiro-Neto 1999), semantic relationships based on domain ontologies—such as the International Classification of Diseases (ICD), the Medical Subject Headings (MeSH), and the Unified Medical Language System (UMLS) (Cimino 1993; Bodenreider 2001; Burgun and Bodenreider 2001; Leroy and Chen 2001)—or learned from healthcare document corpus (e.g., term co-occurrence) (Baeza-Yates and Ribeiro-Neto 1999), template assisted semantic indexing of narrative EMR notes (Mikkelsen and Aasly 2002), and adaptive learning based on relevance feedback from users (Baeza-Yates and Ribeiro-Neto 1999).

However, these techniques are not sufficient to completely capture the medical domain knowledge that links symptoms to diagnosis and therapeutics. Symptoms, exams and tests, diagnoses, and treatments may not appear in the same document simultaneously, preventing the retrieval of relevant documents based just on description of symptoms. Additionally, symptoms may have relationships with other related diseases whose records need to be retrieved. For example, a patient presenting with chest pain and shortness-of-breath will require that the treating physician has knowledge not only of the patient's cardiac history, but also of any history of other relevant diseases, such as diabetes mellitus, hypercholesterolemia, hypertension, and renal failure. Such relationships between particular symptoms, exams and tests, diagnoses, treatments and related diseases need to be acquired from domain experts. The proposed framework integrates information retrieval techniques with medical domain knowledge available in domain ontologies and/or elicited from healthcare experts, and allows effective retrieval of relevant healthcare records based on a patient's symptom. It can liberate doctors from the arduous manual identification of medical records and consequently improve healthcare efficiency and quality.

The paper is organized as follows. In the next section, we review some related work. We then present the proposed framework in section 3 and discuss major challenges and research issues in section 4. Finally we conclude the paper and outline our research plan in section 5.

2. Related Work

Horsch and Balbach (1999) classified telemedical information systems (TIS) into three types: patient related TIS, knowledge related TIS, and meta TIS. As an important component of patient related TIS, EMR keeps track of all the medical data about a patient, including administrative data, diagnostic data, and therapeutic data.

Retrieval of relevant medical records from a large EMR is a challenging task. Simply text word based information retrieval may not yield satisfactory performance, due to the complex semantic relationships among symptoms and diagnostic/therapeutic processes. An evaluation study on the retrieval performance by physicians when searching literature databases showed that only one fourth to one half of the relevant articles on a given topic were retrieved for most searches, thus calling for further research and development to improve retrieval performance (Hersh and Hickam 1998).

Several approaches have been proposed for enhancing free-text queries for better retrieval performance. Baud et al. (2001) proposed using medical lexicons to enhance user queries with more semantic information. Brown and Sonksen (2000) developed and empirically evaluated a semantic terminological model build within Clinical Terms Version 3 (CTV3) for retrieving clinical findings from a patient database and found that this semantically enriched method performed better than free-text search. Bayegan (2002) proposed a framework for a problem-oriented, knowledge-based medical record system. This system ranks the relevance of patient information in a given context using knowledge about physicians' work processes.

More research has been devoted to information retrieval from general medical information sources rather than EMR systems. Fontelo et al. (2005) developed a free-text, natural language search tool, named askMEDLINE, for the medical literature system MEDLINE/PubMed without using any domain-specific vocabularies. Suarez et al. (1997) developed a Web search engine, named Medical World Search, for medical information using UMLS. Malet et al. (1999) proposed a model for enhancing the retrieval of online medical documents using MeSH. Göbel et al (2001) used the controlled vocabulary of MeSH to enhance the queries posed by normal users who are not familiar with standard medical terms and expressions. Leroy and Chen (2001) used UMLS to provide users with appropriate medical terms when posing queries. Plovnick and Zeng (2004) found in an evaluation study that query reformulation using standard terminology from UMLS improved retrieval precision. Liu and Chu (2005) expanded free-text medical queries using the semantic relationships among terms in UMLS, as well as knowledge specific to the scenario (e.g., diagnosis, treatment). While these studies have targeted at different purposes, the developed techniques may be adapted to the retrieval of medical records.

Past research on medical information retrieval has utilized two major sources of semantic relationships among medical terms: standard ontologies and domain knowledge. However, the models and systems have not adequately integrated these two sources together and the use of domain knowledge has been especially preliminary. In our proposed framework, we integrate standard ontologies and domain knowledge acquired from experts, as well as semantic relationship heuristics automatically learned from

EMR databases. As we posit that domain knowledge is crucial for capturing the complex semantic relationships among healthcare concepts that are not available in ontologies and cannot be easily learned, we plan to construct a more comprehensive domain knowledge base that relates symptoms to medical records generated during various diagnostic and therapeutic activities.

3. Proposed Framework

We propose a framework for symptom based retrieval of medical records. Figure 1 illustrates the major components of the framework. It consists of an information retrieval system, which searches a unified EMR database, with facilitation of a global knowledge base. The information retrieval system communicates with the EMR system and the global knowledge base through the Internet using the Web services technology and healthcare industry standards, such as HL7. In an earlier paper, we have presented a service oriented architecture for integrating various distributed, heterogeneous local systems into a unified EMR database (Jain and Zhao 2004). In this paper, we focus on the information retrieval system.

The information retrieval system allows the user to pose a simple query based on a patient's symptom through an interactive user interface. The query enhancing engine will then expand the query using several sources of knowledge, including semantic relationships (e.g., synonyms, hypernyms, and hyponyms) based on standard ontologies, other heuristics (e.g., term co-occurrence) automatically learned from the EMR database, and domain knowledge of healthcare experts. The search engine will then run the enhanced query against the EMR database and present the returned relevant medical records (possibly along with their relevance scores) to the user. The system also allows the user to provide feedback on the relevance of the retrieved documents through the interactive user interface and further enhance the query according to the user's relevance feedback using query expansion and term re-weighting (Baeza-Yates and Ribeiro-Neto 1999). This interaction may be repeated for a few iterations under the user's control.

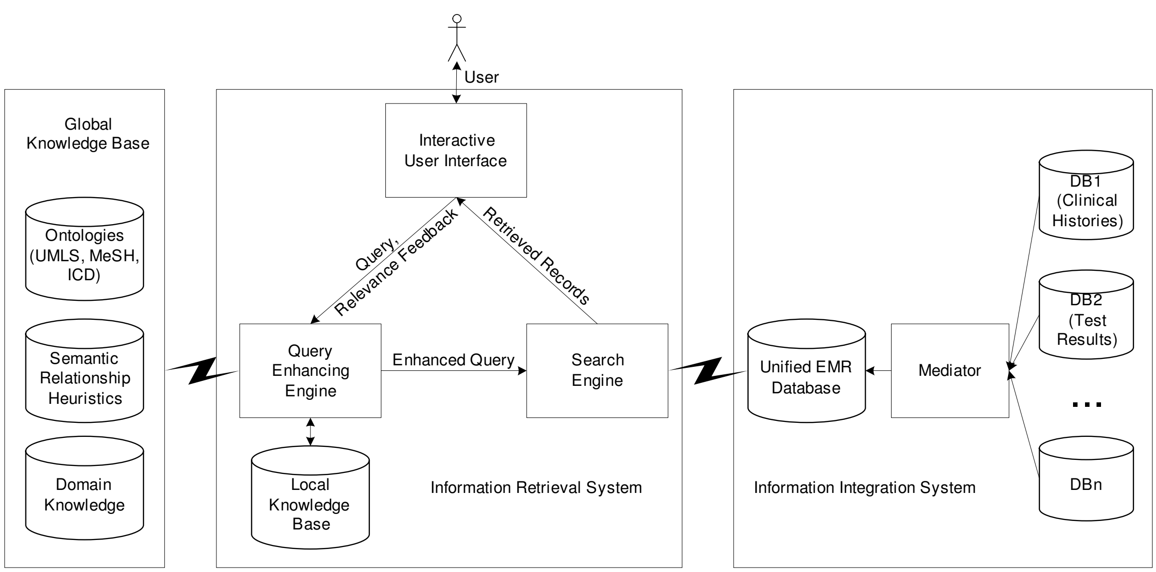

Figure 1. A Framework for Symptom Based Retrieval of EMR

The most critical components in the framework are the sources of semantic relationships among healthcare terms, which are used to expand and enhance user queries. The basic idea of query expansion is to expand a query by adding additional terms that are closely related to the original query terms in some manner, potentially better matching relevant documents. The sources of term relationships include:

- *Domain ontologies*. In this paper, we informally use the term "domain ontology" to refer to any dictionary, lexicon, thesaurus, terminology, classification, or semantic network that is widely shared within the healthcare industry. Such well developed and widely adopted ontologies will be an important cornerstone of the knowledge base, providing some basic term relationships (e.g., synonyms, hypernyms, and hyponyms) and guiding the further development of the domain knowledge base.

- *Domain knowledge*. Some of the medical domain knowledge that links symptoms to diagnosis and therapeutics may not be available in the domain ontologies and need to be acquired from domain experts. For example, the relationship between "symptoms," "disease," and "therapy" is highly specific for various types of cancer. Leukemia, for example, is divided into four types based on the onset ("acute" vs "chronic") and cell line ("lymphocytic" or "myelogenous"); one type alone has seven subtypes. All aspects–and thus the domain knowledge–of leukemia, such as treatment, course, prognosis, and population at risk, are type- and sub-type specific, a fact that holds true for all types of cancers. For a cancer which is treated by surgical intervention, the healthcare practitioner must know whether a patient's medical status will permit an invasive procedure, while a cancer that is treated with chemotherapy or radiation will require consideration of other aspects of a patient's medical history. These considerations must all be incorporated into the collective set of domain knowledge.

  Clearly, the body of domain knowledge is constantly changing, with an evolving understanding of human pathophysiology and development of new therapies. For example, the intimate relationships among obesity, hypercholesterolemia, hypertension and dysregulation of blood sugar control in diabetes have recently led to the definition of a new disease entity, "Metabolic Syndrome," which ties together these ostensibly disparate medical conditions. Fortunately, these new definitions and relationships are incorporated into the medical Standard of Care through readily-accessible documents, such as reports of expert panels convened by medical specialty groups (e.g., the American College of Surgeons, American College of Physicians, American College of Obstetricians and Gynecologists and others), documents issued by the U.S. Centers for Disease Control, such as their *Morbidity and Mortality Weekly Report,* position statements by the United States Preventative Services Task Force, and reports from other governmental and professional healthcare organizations. Incorporation of this new information will form an essential aspect of the maintenance and updating of the domain knowledge, and is not unlike the process which routinely occurs on an ongoing basis in present physician "knowledge databases," such as the physician information database *UpToDate,* and the widely-used *ePocrates Drug and Formulary Reference.*

- *Other heuristics*. There are also techniques for automatically learning term relationship heuristics, which can then be used to expand user queries and potentially improve retrieval performance. Two particular examples are term co-occurrence thesaurus (global analysis) and pseudo relevance feedback (local analysis) (Baeza-Yates and Ribeiro-Neto 1999; Chu et al. 2005). These techniques automatically identify frequently co-occurring terms and use this information to expand user queries. Global analysis builds a global term co-occurrence thesaurus offline from the EMR database and other large healthcare related document bases. This thesaurus can be used not only during query expansion but also in building the domain knowledge base by providing hypotheses on term relationships. The learning of this thesaurus should be incremental to adapt to changes in the underlying EMR database and document bases. Local analysis evaluates the retrieved documents for each query and can better exploit the context of the particular query.

The knowledge is stored in both a centralized global knowledge base and local knowledge bases. The global knowledge base captures common knowledge that is shared by the entire industry and can be accessed by multiple local systems. The use of the global knowledge base reduces duplicated effort in knowledge engineering and maintenance and prevents potential conflicts across local systems. The local knowledge base captures the specific knowledge, preferences, and specialties of the end users of a particular system and allows further customization.

4. Developing the Knowledge Base

Domain ontologies will be an important foundation of the knowledge base. They can not only provide some basic term relationships (e.g., synonyms, hypernyms, and hyponyms), but also guide the further development of the domain knowledge base. One of the leading domain ontologies is UMLS (http://www.nlm.nih.gov/research/umls/), developed by the National Library of Medicine (NLM). NLM hosts the UMLS Knowledge Source Server, which supports both an online query interface and Java APIs. We are considering UMLS as the first candidate for an ontology database in our system.

UMLS consists of three knowledge sources: the Metathesaurus, the Semantic Network, and the SPECIALIST lexicon. The Metathesaurus is built from over 60 source vocabularies, including MeSH, SNOMED, and ICD. It contains several types of information about biomedical and healthcare related concepts. These include the synonyms, narrower concepts (hyponyms), broader concepts (hypernyms), and other related concepts for each term. In certain circumstances, for example, "colon cancer" may be used interchangeably with numerous other synonyms, such as "colorectal cancer", "colon carcinoma", "colon neoplasm or neoplasia", "colonic neoplasms malignant", and "malignant tumor of colon". There are also many narrower concepts for "colon cancer", such as "malignant neoplasm of hepatic flexure of colon", "malignant neoplasm of splenic flexure of colon", "malignant neoplasm of transverse colon", "malignant neoplasm of sigmoid colon", "malignant neoplasm of descending colon", and "malignant neoplasm of ascending colon", broader concepts, such as "malignant neoplasms", "gastrointestinal diseases", "gastrointestinal neoplasms", and "cancer of intestines", and other related concepts, such as "large intestine", "colon", and "episodicities". More generally, UMLS provides the ancestors, parents, siblings, and children of each concept. These can be directly used to expand a user query.

The Metathesaurus also contains term co-occurrence statistics for some concepts based on contributing sources, including MeSH and AI/RHEUM. For example, the concepts that are most frequently co-occurring with the term "colon" in MeSH include "intestinal mucosa", "rectum", "colonic neoplasms", "smooth muscle (tissue)" "gastrointestinal motility", "ulcerative colitis", "colitis", "ileum", "colonic diseases", and "colorectal neoplasms". Such statistics can be used as initial heuristics of term co-occurrence information, which is then refined via learning in real EMR databases. The frequently co-occurring concepts can also be used as hypothetical related concepts during the development of the domain knowledge base. Such potential semantic relationships between concepts can be presented to domain experts for their evaluation and confirmation. This may significantly reduce the time and effort required to elicit the knowledge of semantic relationships from experts, compared with knowledge acquisition from scratch.

The Semantic Network provides a categorization of the concepts (155 semantic types) in the Metathesaurus and a set of 54 relationships among these semantic types. For example, "Bacterium" (semantic type) "causes" (relates to) "Disease or Syndrome" (another semantic type) and "Antibiotic" "treats" "Disease or Syndrome". While the Semantic Network only contains such a schema about semantic relationships, it is very useful as a blueprint to develop the domain knowledge base. The types of relationships defined by UMLS can be evaluated and the ones most useful for the symptom based EMR retrieval can be selected and then populated with detailed instances. For example, "colon cancer," a subset of "neoplasms of the gastrointestinal tract," is a "disease" which has a "treatment" which may involve "medical intervention" (chemotherapy) and/or "surgical intervention" (colectomy) and followed with certain “laboratory tests” (tumor markers or histopathology). Such instances of semantic relationships need to be elicited from domain experts.

The SPECIALIST lexicon contains the syntactic, morphological, and orthographic information about both commonly occurring English words and biomedical terms. For example, “pain” can be used as a regular noun with a plural form of “pains”, as an uncountable noun, and as a verb with different forms, such as “pains”, “pained”, and “paining”. This information allows different forms of the same term to be standardized into a uniform basic form, such that mismatches due to only form discrepancies between user queries and EMR records can be reduced.

5. Major Challenges and Research Issues

We foresee several major research challenges in implementing and deploying the proposed framework:

- *Heterogeneity across local systems.* Healthcare related information is generated and managed by numerous information systems owned by hospitals, clinics, independent physicians, testing laboratories, and other healthcare providers. These local systems are typically heterogeneous on various levels, including hardware, operating system, database management system, data format, terminology, and semantics. Integrating these heterogeneous systems into a unified EMR system is a prerequisite for the effective functioning of the retrieval system. The recent development in the Internet, the semantic Web (Berners-Lee et al. 2001), and Web services has provided a technological foundation for bridging the hardware and software gaps across local systems (Kotsiopoulos et al. 2003; Zhu et al. 2004). Healthcare industry standards, such as HL7, have also been developed to facilitate the interoperability and data exchange across systems within the industry (Lucy-Bouler and Morgenstern 2003). We have proposed an approach for information integration using these technologies and industry standards (Jain and Zhao 2004) and plan to implement a prototype system as a foundation for the retrieval system, based on a set of common signs, symptoms and disease entities.

- *Diversity of media types*. EMR is typified by diversity of media types, including structured data, free text, and images (such as X-ray, cardiology, and ultrasound images). In addition, previous legacy manual reports may be simply scanned into the computerized EMR database and are hard to interpret. Semantic annotation of these records is a challenging task. Tools, such as templates assisting the semantic indexing of narrative EMR notes (Mikkelsen and Aasly 2002), need to be developed. There are also techniques for converting scanned images into text [Auto OCR 2005] and techniques for automatically indexing text documents [Baeza-Yates and Ribeiro-Neto 1999]. These techniques need to be applied to facilitate the extraction of information from scanned manual reports.

- *The knowledge acquisition bottleneck.* Eliciting the domain knowledge form healthcare experts is a difficult task, known as the knowledge acquisition bottleneck. Several knowledge acquisition techniques—such as interviews, protocol analysis, observation, and focus groups—may be used to facilitate knowledge transfer from experts. The global knowledge base is developed and maintained by a centralized team of knowledge engineers and domain experts. The local knowledge base is maintained locally by the end users.

- *Evolution of domain knowledge*. With the fast development in medicine and pharmaceutics, the healthcare domain knowledge changes constantly and frequently. Classifications for diagnosis and treatment recommendations are constantly being refined and re-defined; consider, for example, the new guidelines for classification of “prediabetes” offered in late 2003, and the thresholds for diagnosis of “pre-hypertension” announced in 2003 as well. Similarly, management of tuberculosis infections is the subject of CDC guidelines which have become increasingly complex over the last decade, and which now take into account a patient's co-morbidities and other risk factors. The knowledge base must be updated accordingly. The system must have such adaptation ability. At the same time, some learning mechanisms that can automatically detect new events, findings, and patterns from medical and healthcare information sources and alert knowledge engineers are preferred. In addition, the learning of the semantic relationship heuristics should also be incremental and adaptive.

- *Privacy, confidentiality, and security*. Apparently, EMR information is of the utmost sensitivity to patients (Rindfleisch 1997). Privacy, confidentiality, and security are the most critical issues that must be satisfactorily addressed to encourage patients and care providers to actually use the EMR systems. Although the proposed framework does not include components that directly address these issues, adequate policies, procedures, and technologies must be in place to guarantee the compliance with security and privacy regulations (Lutes 2000).

6. Conclusion and Research Plan

We have presented a framework for symptom based retrieval of medical records. It strives to improve retrieval performance by integrating information retrieval techniques with medical domain knowledge available in domain ontologies or elicited from healthcare experts, as well as automatically learned heuristics.

This research is still in an early stage. We are forming a research project team, consisting of IS researchers, IT engineers, and healthcare domain experts, and planning to realize the proposed approach using several EMR systems maintained by Veteran Affairs (VA) facilities and Intensive Care Units (ICU) in a Midwest county. We plan to develop and evaluate a prototype system based on the proposed architecture and validate the technical/clinical feasibility of the proposed approach. The development of the knowledge base will be initially focused on one healthcare sector (e.g., heart disease, diabetes, or asthma) and later generalized to other areas. Our long term goal is to develop a generic retrieval system that works with a variety of EMR systems, including, for example, the general EMR system Medicare will offer doctors free of charge (Kolata 2005) and the MPI funded by the National Institute of Standards (Bell and Sethi 2001).

References

Auto OCR, http://www.southwestdata.com/textocr.html, 2005

Baeza-Yates, R. and Ribeiro-Neto, B. *Modern Information Retrieval*. Addison-Wesley, Reading, MA, 1999.

Bayegan, E. *Knowledge Representation for Relevance Ranking of Patient-Record Contents in Primary-Care Situations*. Doctoral Thesis. Norwegian University of Science and Technology, 2002.

Baud, R. H., Lovis, C., Ruch, P., and Rassinoux, A. M. "A Light Knowledge Model for Linguistic Applications," *Proceedings of AMIA Symposium*, 2001, pp. 37–41.

Bell, G. and Sethi, A. "Matching Records in a National Medical Patient Index," *Communications of the ACM* (44:8), 2001, pp. 83-88.

Berners-Lee, T., Hendler, J., and Lassila, O. "The Sematic Web," *Scientific American* (284:5), 2001, pp. 34-43.

Bodenreider, O. *Medical Ontology Research - A Report to the Board of Scientific Counselors of the Lister Hill National Center for Biomedical Communications*, National Library of Medicine. 2001.

Brown, P. J. B. and Sonksen, P. "Evaluation of the Quality of Information Retrieval of Clinical Findings from a Computerized Patient Database Using a Semantic Terminological Model," *Journal of American Medical Informatics Association* (7:4), 2000, pp. 392–403.

Burgun, A. and Bodenreider, O. "Mapping the UMLS Semantic Network into General Ontologies," *Proceedings of AMIA Symposium*, 2001, pp. 81-85.

Cimino, J. J., Johnson, S. B., Peng, P., and Aguirre, A. "From ICD-9-CM to MESH using the UMLS: a how-to-guide." *Proceedings of the Annual Symposium on Comput Appl Med Care,* 1993, pp. 730-734.

Chu, W. W, Liu, Z., Mao, W., and Zou, Q. "A Knowledge-based Approach for Retrieving Scenario-specific Medical

Text Documents," *Control Engineering Practice (CEP), Special Section on Biomedical Control*. 2005.
Fontelo, P., Liu, F., and Ackerman M. J. "askMEDLINE: A Free-Text, Natural Language Query Tool for MEDLINE/PubMed," *BMC Medical Informatics and Decision Making* (5:5), 2005, pp. 1-14.
Göbel, G., Andreatta, S., Masser, J., and Pfeiffer, K. P. "A MeSH Based Intelligent Search Intermediary for Consumer Health Information Systems," *International Journal of Medical Informatics* (64), 2001, pp. 241-251.
Hersh, W. R. and Hickam, D. H. "How Well do Physicians Use Electronic Information Retrieval Systems? A Framework for Investigation and Systematic Review," *Journal of the American Medical Association* (280:15), 1998, pp.1347-1352.
Horsch, A. and Balbach, T. "Telemedical Information Systems," *IEEE Trans. Inform. Technol. Biomed.* (3), 1999, pp. 166–175.
Jain, H. and Zhao, H. "Integrating Healthcare Information Systems Using Web Services and Vertical Industry Languages." *Proceedings of the Third Workshop on e-Business (WeB 2004),* Washington DC, 2004, pp. 460-468.
Kolata, G. "U.S. Will Offer Doctors Free Electronic Records System," *US News*, July 21, 2005.
Kotsiopoulos, I., Keane, J., Turner, M., Layzell, P., and Zhu, F. "IBHIS: Integration Broker for Heterogeneous Information Sources," *Proceedings of 27th Annual International Computer Software and Applications Conference (COMPSAC 2003)*, Dallas, 2003.
Leroy, G., and Chen, H. "Meeting Medical Terminology Needs—the Ontology-Enhanced Medical Concept Mapper," *IEEE Transactions on Information Technology in Biomedicine* (5:4), 2001, pp. 261-270.
Liu, Z. and Chu, W. W. "Knowledge-based Query Expansion to Support Scenario-Specific Retrieval of Medical Free Text," *Proceedings of the 20th ACM Symposium on Applied Computing (SAC 2005), Special Track on Information Access and Retrieval*, Santa Fe, NM, 2005.
Lucy-Bouler, T., and Morgenstern, D. "Is Digital Medicine a Standards Nightmare?" *Proceedings of the Workshop on Standard Making: A Critical Research Frontier for Information Systems*, John L. King and Kalle Lyytinen, (eds.), Seattle, Washington, 2003, pp. 354-360.
Lutes, M. "Privacy and Security Compliance in the E-healthcare Marketplace," *Healthcare Financial Management* (54:3), 2000, pp. 48-50.
Malet, G., Munoz, F., Appleyard, R., and Hersh, W. "A Model for Enhancing Internet Medical Document Retrieval with "Medical Core Metadata,"" *Journal of the American Medical Informatics Association* (6:2), 1999, pp. 163-72.
Mikkelsen, G. and Aasly, J. "Manual Semantic Tagging to Improve Access to Information in Narrative Electronic Medical Records," *International Journal of Medical Informatics* (65), 2002, pp. 17-29.
Plovnick, R. M. and Zeng, Q. T. "Reformulation of Consumer Health Queries with Professional Terminology: A Pilot Study," *Journal of Medical Internet Research* (6:3), 2004.
Rindfleisch, T. C. "Privacy, Information Technology, and Health Care," *Communications of the ACM* (40:8), 1997, pp. 92-100.
Suarez, H., Hao, X., and Chang, I. "Searching for Information on the Internet Using the UMLS and Medical World Search," *Proceedings of AMIA Annual Fall Symposium,* 1997, pp. 824-8.
Tambouris, E., Willimas, M. H., and Makropoulos, C. "Co-operative Health Information Networks in Europe: Experience from Greece and Scotland," *Journal of Medical Internet Research* (2:2), 2000.
Thompson, T. G. and Brailer, D. J. *The Decade of health Information Technology: Delivering Consumer-centric and Information-rich Health Care*, Office of the National Coordinator for Health Information Technology (ONCHIT), Department of Health & Human Services, July 2004.
Zhu, F., Turner, M., Kotsiopoulos, I., Bennett, K., Russell, M., Budgen, D., Brereton, P., Keane, J., Layzell, P., Rigby, M., and Xu, J. "Dynamic Data Integration Using Web Services," *Proceedings of 2nd International Conference on Web Services (ICWS 2004)*, San Diego, California, 2004, pp. 262-269.