Technical Report – September 2009

MedicInfoSys: An Architecture for an Evidence-Based Medical Information Research and Delivery System.

Pif Edwards Vlado Keselj

Faculty of Computer Science,
Dalhousie University
Halifax, Nova Scotia
Canada

{pedwards, vlado}@cs.dal.ca

Abstract

Medical information is growing at an exponential rate. The majority of physicians information needs are not being met. Present information systems are insufficient. Knowledge-based methods and resources, once brittle and unreliable, have matured. Resources such as the UMLS open promising new avenues for experimentation, new implementations and better relevance performance. This paper explores information systems in the context of Evidence-Based Medicine (EBM) and the information needs of physicians.

In this work we identify 3 primary problems specific to this domain, and propose a solution in the form of an architecture. The first problem is time; physicians spend on the average 2-8 minutes per question and it takes on the average 10-45 minutes to answer all but the most simple clinical query. The answer to this problem is delegation. Just as in the primary care context, physicians often delegate tasks to specialists, the same must be done in the information context: physicians must delegate the information finding tasks to 'informationists' (a.k.a. medical librarians). Studies show questions answered from a central location can be done at an average cost $27.50 per question with an average wait time of 6 hours (via FAX). This fast, inexpensive medical test is shown to increase the average quality of care 47%. The second problem is the average length of queries is 2-3 keywords, which is insufficient for medical question answering. The answer that we propose, an Evidence-Based Medicine (EBM) style "Well-made Question" approach to: a) **structure** the query for the *user*; and b) **contextualize** the query for the *system*. Furthermore, a structured query prompts the user to first form the question in their mind and thus form better queries. The third problem is the sheer volume of results: 1000s of results for even moderately specific well-formed queries is the norm. We propose a hierarchical categorization of search results. The maturity of knowledge-based resources in the medical domain allows a speedy, trustworthy and customizable categorization.

Our 3-layer architecture details a delivery system that is fashioned around the studied information needs of physicians, and it is not a simple adaptation of existing systems. The end-user layer structures the query, interacts with the Informationist layer and shows the results after passing through each layer. The informationist layer is where the medical information specialist uses information need to skillfully query, browse and form a solution with interaction with the system layer. The system layer filters, categorizes, extracts and presents the information from multiple sources to the informationist layer.

Table of Contents

List of Figures

1. Introduction

"Me dicine, in modern jargon, is a knowledge based business, and experienced doctors use about two million pieces of information to manage their patients. ...Clinical information can be defined as 'the commodity used to help make patient care decisions."[12]

The above quote is very helpful in framing the modern medical situation in a way computer scientists can appreciate. Given a patient's situation, a physician is either certain or uncertain on how to proceed. We must provide a system that increases the level of medical certainty in patient care, which benefits both patients health and doctors confidence in a present, and similar future situations.

1.1. Clinical Context

There are three important factors at the center of the clinical context that motivate and mold our efforts in health informatics: 1. the stakes are very high (for physicians as well as patients); 2. time is in short supply and; 3. doctors have sophisticated and context-specific information needs which must be satisfied by an equally sophisticated and comprehensive knowledge base.

Doctors are experts; they have fourteen years of post-secondary education and their level of diction reflects that education. This high level of diction makes the source material – medical documentation – often beyond the understanding of anyone outside the medical field, and its interpretation into medical practice requires years of experience. In order to plumb this highly sophisticated source material, equally as sophisticated methods of information retrieval are required.

Physicians bring a huge volume of medical knowledge to bear on any reading and interpretation of a medical article. This fact makes seemingly straight-forward tasks in this domain, like the identification of similarities/ differences in sentences, exceedingly difficult for computers to perform [17].

The conclusion that the systems which presently exist are insufficient is supported by the fact that the majority of medical questions go unanswered [8,12,71]. Doctors have much less time to pursue information needs (2–8 minutes) [8,71], then it takes to satisfy all but the simplest of them (10–43 minutes on average) [7,12,30]. Several studies have shown that the majority of unanswered questions were often times answerable – between 77%–92% [7,12] of the time – with present resources and changed patient management 40%–47% [7,12] of the time. The answers are there, but within the present clinical context, doctors cannot find them due to a lack of time and the inadequacy of search systems.

It is important to note there are two distinct user-groups, those in the research context and those in the clinical context. The research information tasks produce results which are meant to be generalized, the clinical information tasks are meant to be interpreted into the context of a specific patient.

In the terms of information retrieval (IR) evaluation, the time constraints present in the clinical context support the weighting of precision over recall as the prudent evaluator of IR systems. That is, finding a small number of good articles is sufficient, perhaps even one if it has the precise answer. Contrast this with users in the research context which require a system that performs strongly in terms of recall [7]. That is, finding articles which cover many-to-all different perspectives on a topic.

1.2. The Clinical Information Climate

> *"U S medical care is some 30% more expensive than that in Canada and Europe, where quality is comparable; and US medicine also has the most litigious malpractice climate in the world. Some have argued that this 30% surcharge on US medical care, about US$1000 per capita annually, is mostly medico-legal: either direct legal costs, or else the overhead of "d efensive medicine," i.e. unnecessary tests ordered by physicians to cover themselves in potential future lawsuits. In this tense climate, physicians and other medical data-producers are understandably reluctant to hand over their data to data miners."*[18]

With the stakes as high as they are in medicine, where daily decisions have life and death impacts, information must be accurate and timely, and sources must be reliable and trustworthy. If not, patients face death and injury and doctors face lawsuits and the loss of their livelihoods. With stakes this high questions of ethical responsibility must be addressed.

Data mining in the medical domain has three primary ethical issues: data ownership, fear of lawsuits and privacy [18]. First the question of **data ownership;** do *patients* own data about them, or do *physicians* own the data they collect, or do the *insurance providers* who paid for the tests own the data? Adding to the confusion are ethical questions surrounding the sale of human data and tissues. Second, there is the threat of **lawsuits.** Physicians and medical data-producers are wary that data provided could be used against them in a court of law, that accidental omissions and unrepresented context specific information may generate – or add leverage to – a case of malpractice. This is a situation where physicians have much to lose and little to gain. Finally the issue of **privacy.** Patient-physician confidentiality is a legal contract which patients and doctors both take very seriously. If there was any doubt in this confidence, patients may not be as forthcoming and the care of patients would suffer. These ethical question can be contextualized by the following four levels of identification:[18]

1. **Anonymous data:** No identification. (e.g. Tissue sample from a corpse.)
2. **Anonymized data:** Identification irrevocably removed.
3. **De-identified data:** Patient ID encoded and encrypted.
4. **Identified data:** Patient given written informed consent.

These ethical questions are moot for level 1 and level 2; has increasing impact on level 3; and these questions are vital and explicit for level 4 data. For many questions specific to patient diagnosis the context information available only in level 3 and 4 is critical and highly valued.

1.3. The Advantages of Medical literature.

For all experts, text is the primary channel for information exchange [1], the medical domain is no different; the medical literature is the predominant medium for researchers to make known their findings. Medical articles have well-structured conventions for the presentation of the material which provides multiple entry points into the information (Title, Abstract, Introduction, and specific sections headings to direct the readers' attention.) In 1987, The Ad Hoc Working Group for Critical Appraisal of the Medical Literature established guidelines for structuring headings within abstracts to reflect the content of publications in an effort to help people quickly assess content [17], increasing the usability of the literature for users. When present, this standardized structure of headings (Objective, Method, Results, Conclusion) can be used to the advantage of an IR system made sensitive to it.

Clinicians are not the only ones who have unmet medical information needs. Pharmaceutical companies in development of medications use the same resources and it is estimated that these companies derive 90% of drug targets from the literature [13]. Unfortunately, the amount of information is curbing their advancements, "su rveys suggest that about 50% of all potentially therapeutic compounds undergo attrition due to safety concerns and that about 50% of them had some indication in the literature already" [13].

Computerized medical records have been a suggested new source of research data. However, since medicine is primarily a patient-care activity and only secondarily acts as a research resource [18], it must be noted that there is a clear advantage of finding evidence in scientific literature since it is *intended* to be used as evidence, where data generated from medical records is not. When filling out patient medical records doctors are meant to focus on patient health not on the future needs of researchers. Though the use of these records is rife with pitfalls (privacy, legal-responsibility, their anecdotal and idiosyncratic nature, and habitual incompleteness [18]) they can be effectively used in a supporting role, for example in the assistance of automatic and interactive query formulation.

1.4. Information Glut.

The sheer volume of documents in this domain is its greatest blessing and ultimate obstacle. There is a vast array of information sources: commercial, governmental, academic, open-access, all of varying reliability. Over the last 20 years, primary sources (such as MEDLINE) are growing at a double-exponential pace [14]. MEDLINE specifically, has grown at a ~4.2% compounded annual growth rate [14], and as of September 2009, MEDLINE has 17,634,342 [54] citations indexed and was increasing at a rate of more than 2300/day [54]. Medical research produces medical findings at a reported rate of publishing 55 clinical trials a day [74]. Each medical specialty has its own tale, but the same story; Epidemiologists "would need over 600 hours a month to read every new article published in their field" [5] and "t he body of information on HIV doubles every 22 months, and, although half of that information is concentrated in 30 journals, the other half is spread through 593" [12]. In fact, according to research done in 1985, the biomedical knowledge-base doubles every 19 years, meaning that medical knowledge will quadruple during a professional lifetime [12]. If you consider the rate of growth of medical information is double-exponential the doubling time of medical knowledge, is shortening, thereby worsening this problem.

This problem not only has the medical impacts of lost opportunities for improved patient care but financial impacts as well. “Studies by International Data Corporation estimate that an enterprise employing 1,000 knowledge workers wastes nearly $2.5 million per year due to an inability to locate and retrieve information” [13].

Medical professionals stand at the foot of an exponentially growing “mountain” of information. For the proper functioning of the medical system patients must have confidence in their doctors level of knowledge and for doctors to provide a state-of-the-art level of care they must have the state-of-the-art tools to navigate this “mountain”.

In summary, the sheer volume of information and lack of an adequate way of searching it has the following consequences: (1) searching for answers to clinical questions is likely to fail; (2) keeping up to date in even one medical field requires an enormous effort and time – time and effort which doctors prefer to spend caring for patients; (3) advances in the field, medical breakthroughs, and all the “effort, creativity, and money that go into biomedical research is simply wasted” [11].

1.5. Information Needs of Clinicians

There are two types of information needs of clinicians: focused and general. The **focused** need is one where a specific question is formulated, specific situational factors are in play and the clinician requires an exact answer. The **general** information need is one where an overview is necessary and sufficient to satisfy the need, but from which a focused question may emerge [4].

Ely et al [20] divided the process of asking and answering clinical questions can into five steps: (1) recognizing an uncertainty, (2) formulating a question, (3) pursuing an answer, (4) finding an answer, and (5) applying the answer to patient care. Ely et al also compiled a complete taxonomy of obstacles to clinical question answering [53] and is available in my extractive summary. Both the question process steps and obstacle taxonomy are useful in identifying where things go wrong, and to generate ideas on how we can help.

The satisfaction of an information need begins with recognizing one. The lack of recognition of a need, that is, the problem of not knowing that you don't know, is aggravated by rate which new clinical information is being generated. We could provide tools to help in this capacity, for example, article recommendations and new finding notifications may help recognize uncertainties by increasing awareness.

The primary reasons not to pursue questions were lack of time and lack of confidence that an answer could be found [32]. Thus any new information system must demonstrate what sort of questions can be answered and how quickly, especially when systems are faster and when questions that a system is capable of handling are unlike those in past systems.

Several studies show that when they are answered patients benefit. It has been shown “that conducting a MEDLINE search early in the hospitalization of a patient could significantly lower costs, charges, and lengths of stay” [51] and that “...answers to these questions came from MEDLINE and the information from the articles changed patient management 47% of the time” [7] and another study “reported that the use of an online information retrieval system improved the quality of clinicians' answers to clinical questions by 21%” [52].

1.6. Obstacles.

Doctors have obstacles in question answering which have nothing to do with the technology or resources available. These **physician-related** obstacles are not the problems we should be trying to solve, they include:

- the failure to recognize information needs,
- the decision to pursue questions only when answers are thought to exist,
- the preference for the most convenient rather than the most appropriate resource,
- and the formulation of questions in a way that is difficult to answer with general resources. [20]

The best use of our time is to focus on the development of a system which overcomes the **resource-related** obstacles:

- the excessive time and effort required to find answers in existing resources,
- the difficulty navigating the overwhelming body of literature to find the information needed,
- the lack of access to information resources,
- search technology that is unable to directly answer clinical questions,
- and the lack of evidence that addresses questions arising in practice. [20]

However, the side-effect of faster, easier to use systems, which provide precise answers to specific questions will be the redefinition of expectations. In this way, better technology will also help overcome physicians-related obstacles.

If a fast and reliable system was prevalent, a lower level of uncertainty may cause doctors to initiate a search. Furthermore, a routine "better safe then sorry" search before prescription, diagnosis or treatment may become commonplace. Such searches would reveal new findings and updated recommendations, exposing information needs which would have gone otherwise undetected. This is a circumstance where the mentioned physician-related obstacles are essentially solved. Due to obstacles created by the sheer volume of clinical information this circumstance necessitates better information systems than are presently available.

1.7. How physicians search

In a 2008 study [31] of American emergency rooms where a follow-up visit is unlikely and need is immediate, the following table (Table 1.) was developed:

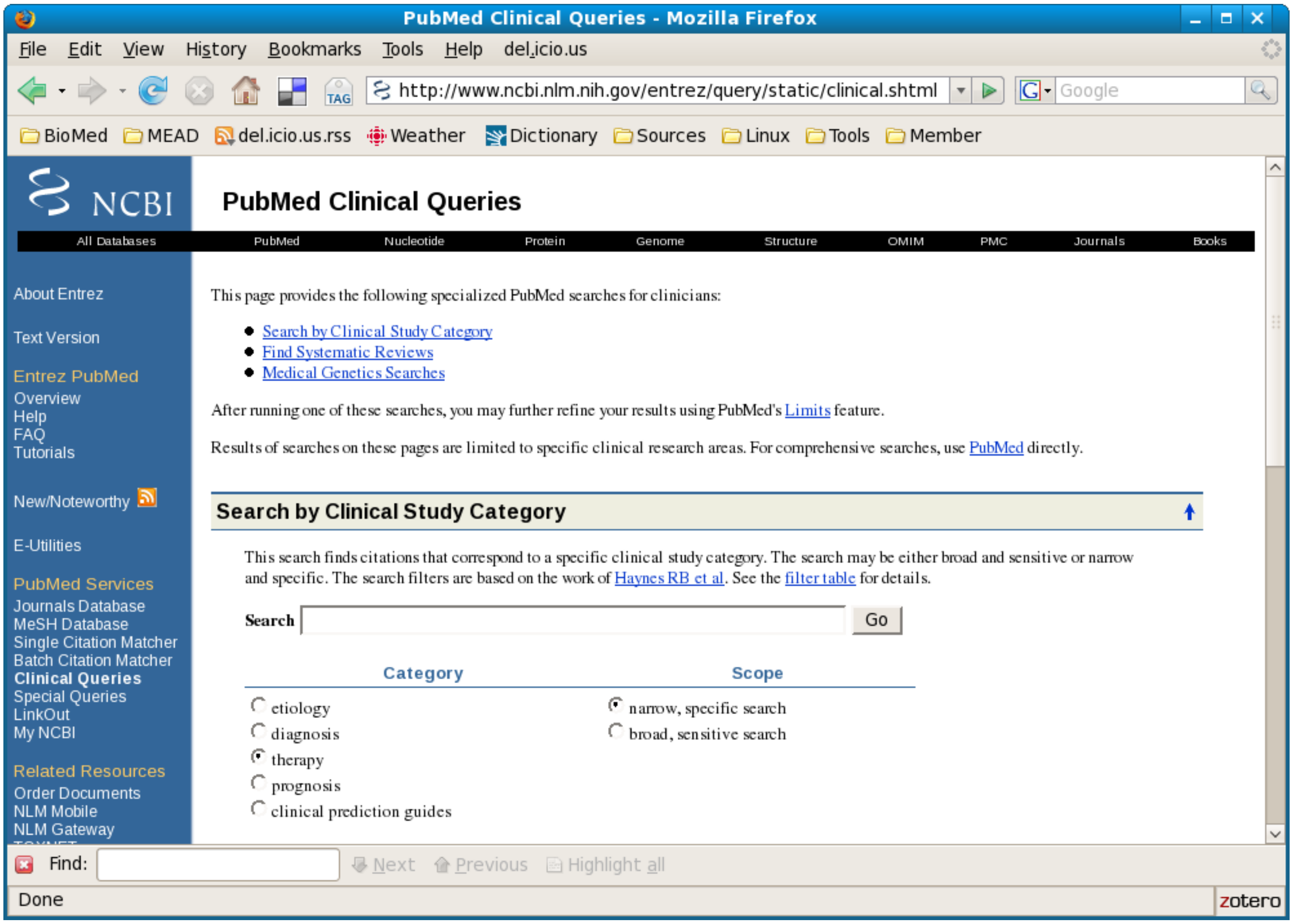

***Figure 1**: The PubMed clinical queries search UI.*

Source	Frequency	%
PDA-based drug information: Epocrates/Tarrascon/ Clinical Pharmacology	22	17.5
Micromedex	14	11
Pocket Pharmacopeia (print version)	11	8.5
Google	11	8.5
UpToDate.com	11	8.5
Consulted specialist	10	8
Miscellaneous texts	9	7
Consulted ED colleague	7	6
Tintinalli et al, “Emergency Medicine: a comprehensive study guide.” 6th edition. 2003.	7	6
PubMed	5	5
Red Book: 2006 Report of the Committee on Infectious Diseases 27th edition; American Academy of Pediatrics, 2006; Elk Grove Village, Ill	4	3
Harrison's Online	3	2

PDA-other (personal notes, 5-Minute Consult, PEPID)	3	2
eMedicine.com	3	2
Lange, EM On Call	2	1.5
Willis Eye Manual	2	1.5
Sanford Guide	2	1.5
Total	126	98.5

Table 1: Shows the sources Emergency Room doctors use to successfully satisfy information needs in an American 2008 study.

You can see in this study doctors favor IR systems 54% of the time, then print sources 28.5% and finally colleagues 14% of the time. The preference toward IR systems makes itself clear.

2. Evidence Based Medicine

> *“Evidence-based medicine (EBM) is a widely accepted paradigm for medical practice that involves the explicit use of current best evidence, that is, high-quality patient-centered clinical research such as reports from randomized controlled trials, in making decisions about patient care.” [17]*

Within the field of EBM, the problem of question formulation “is the first and arguably the most important step in the EBM process. Without a well-focused question, it can be very difficult and time consuming to identify appropriate resources and search for relevant evidence” [7]. To solve this problem “well-built question” methods, such as the PICO model have been suggested and are taught in EBM curriculum. For this problem of question formulation we could look at generic question templates, structured queries (implementing the PICO Model), and interactive query iteration may help in formulating questions.

2.1. PICO

PICO is a mnemonic which stands for **P**atient/Population/Problem, **I**ntervention/Exposure, **C**omparison and **O**utcome. This mnemonic is meant to be used by clinicians to aid in the formulation of an evidence-based question. This method, first suggested in 1995 [44], now pervasive, has generated a number of variants including PICOTT [7], PECODR [72] and PESICO [73]. As a companion to the PICO method questions were also divided into 6 types: ***Clinical evidence***, concerning interpretation and gathering of evidence; ***Diagnosis***, concerning selection and interpretation of diagnostic tests; ***Prognosis***, concerning predicting complications and mapping a patient's progress; ***Therapy***, concerning treatments; ***Prevention***, reduce risk; and ***Education***, how to teach patients, families and oneself [44]. Many descriptions of the question types have simplified these six, to four types of questions: *Diagnosis*, *Therapy*, *Prognosis* and *Etiology*. In the literature, these four are regularly referred to as the 'Clinical Tasks.'

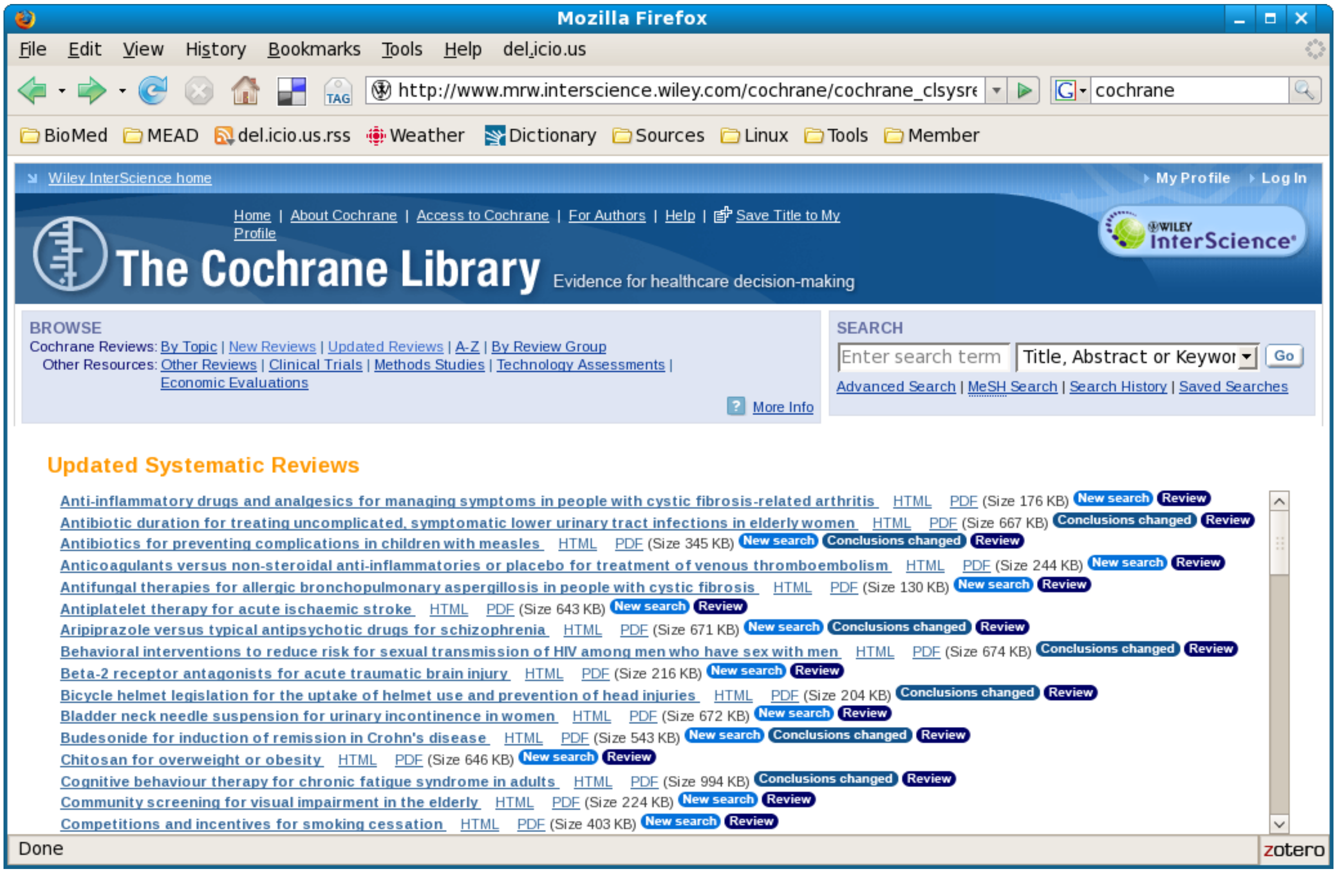

Figure 2: *A screenshot taken from The Cochrane Library of flagged search results.*

IR specialists are aware of the importance of query formulation, the users predisposition to 2-3 word queries, issues of lexical, syntactic and semantic ambiguity and the guess work commonly needed to predict: what the user means, is looking for, and their task. This PICO system implicitly gives keywords context and question type indicates search task. An IR expert can see the value of this and the improvement over standard 2-3 keyword queries.

PubMed has a clinical queries search mode which qualifies the search with the selection of radio button to indicate the type of clinical query as seen in **Figure 1**. This search tool specialized for clinical queries is based directly on the research of the Hedge filter group from McGill University [28].

A problem persists with the PICO method, not all questions can fit the PICO frame. Some drawbacks include inability to capture temporal information and anatomical qualifications [6]. It has also been noted that this model favors questions pertaining to treatment and interventions and is less conducive to well-built prognosis and etiology question formulation [6].

2.2. Strength of Evidence

One of the foundations of Evidence-Based Medicine is strong evidence, thus many models of evidence categorization have been created: **SORT** (Strength Of Recommendation Taxonomy) [45,46], **Oxford Centre Levels of Evidence** [47], and **GRADE** [48,49]. All systems are similar in that controlled randomized trials are the most highly rated form of evidence, followed by cohort studies in the middle and expert opinion ranking lowest.

2.3. EBM Informatics Infrastructure

If we look at this EBM informatics framework, the “five building blocks of an informatics infrastructure for evidence-based practice are proposed: 1) standardized terminologies and structures, 2) digital sources of evidence, 3) standards that facilitate health care data exchange among heterogeneous systems, 4) informatics processes that support the acquisition and application of evidence to a specific clinical situation, and 5) informatics competencies” [50]. We can see our project fits in the fourth category and is a key element of the overall EBM informatics task.

3. Knowledge Sources

3.1. Ambiguity in the medical domain

Ambiguity is a central obstacle in all levels of language processing and information retrieval. There is no difference in this domain, maybe even more of a problem then in the general domain. We listed five key examples of medical domain specific ambiguity which are particularly problematic: tokenization, acronyms, polysemy, synonymy and metonymy.

Tokenization. Identifying sentence boundaries is a problem, periods are used for sentences, abbreviations, decimals, and hierarchical delimiters, and it is not uncommon to have sentences that begin with lowercase letters [14].

Acronym/ abbreviation. With so many lengthy chemical compounds, anatomical terms and pathogen taxonomies it is easy to see the motivation to make regular use of shortened forms. (e.g. PDA = “patent ducus artteriosus”, “prosterior descending artery”, “phorbol 12, 13 diacetate”, “Parenteral Drug Association” [14], not to mention general uses like “personal digital assistant”.) New acronyms are being introduced to the domain at an alarming rate of one every introduced in every five to ten abstracts [42]. Making the problem bigger more difficult greater than 8% of acronyms are ambiguous [14] and there are, on average, more than 15 possible interpretations for a given acronym [14].

Polysemy. A single name can refer to more then one gene from a single species and from different organisms. (e.g. The Entrez Gene database contains more than 800 distinct gene that have been called P60) [14].

Synonymy. The problem of many words having the same meaning may be particularly acute in this domain where for example, many trademark names refer to the same compound (e.g. ibuprofen is sold as Advil, Bufren, Motrin, Nuprin and Nurofen) [1]. This and other factors create the situation where six or seven synonyms for a single concept is common, [1] resulting in a deeply problematic semantic ambiguity where “the probability of two experts using the same term to refer to the same concept is less than 20 per cent” [1].

Metonymy. The use of a word for a concept or object which is associated with the concept/object originally denoted by the word. For example, in the phrase: “*The White House* *phoned...*," the use of the word *White House* to mean *President* [35], is an example of the word "W hite House" used as a metonym. A string like *p53* could refer to the gene of that name, to the protein that it codes for, or to its mRNA [14].

3.2. PubMed and MEDLINE and ClinicalTrials.gov

Index Medicus, created in 1879, was a comprehensive index of medical journal articles which evolved into the US National Library of Medicine (NLM). This index was supplanted by the **PubMed** (also a NLM project) and ceased publication in 2004. **MEDLINE** is the biggest database of medical journal abstracts indexed and searched by PubMed [14].

MEDLINE is a collection of articles from 5,398 (as of July 2009) [55] medical journals with 17,634,342 (as of September 2009) [54] total records from 1966 to the present with articles added to MEDLINE at the average rate of over 2300/day [54]. Each of these articles have been manually indexed by one of 100 human indexers with MeSH terminology, 671,904 were indexed in 2008 [81]. MEDLINE is one of the resources searched by PubMed, both are maintained by the NLM. Since PubMed searches MEDLINE and other resources, it is a little larger: it has 19,174,957 (as of September 2009) [54] total records from 1948 to the present. Other sources it searches are, for example: (1) the 301,775 [54] articles not yet indexed with MeSH terminology, but in the process of being processed (i.e. indexed with MeSH) into the MEDLINE system, and (2) the 409,526 [54] records from OLDMEDLINE which contains records from the years 1948 to 1965. MEDLINE is free to search, 75% of the articles published in the last 25 years have abstracts in MEDLINE. As for usage, PubMed and MEDLINE were searched 671 million times in 2008 [81].

On April 11, 2003, in promotion of **open access** to scientific literature, a group drafted a statement known as *The Bethesda Statement*, this was followed by *The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, pushing for open access in reaction to rising subscription fees and decreasing library budgets [14]. In 2004, the NLM created PubMedCentral (PMC, http://www.pubmedcentral.gov/), an online digital library of open-access journal articles, containing some or all the articles from about 154 journals and individual article submissions from many others [14]. Since 2005, all NIH funded researchers (in part or in full) were requested to submit manuscripts to PubMedCentral, adding 430,000 manuscripts (5TB compressed) to PMC [14]. In late 2007 that changed from voluntary submission, to a legally binding one, with the *Consolidated Appropriations Act of 2007* (H.R. 2764) [82]. As of 2007, 18% of recent and 12% [57] overall PubMed articles are available as full-text through open-access sites such as PubMedCentral, BioMedCentral [65] and the Public Library of Science [66].

ClinicalTrials.gov [29], maintained by the NLM, currently contains 61,557 trials in their database from 157 countries and receives over 40 million page views per month [56]. It is by far the largest repository of controlled randomized trials and observational studies [56]. This is a major directory of primary sources for anyone interested in biomedical research and in Evidence-Based Medicine.

3.3. Cochrane Collaboration

> *"The Cochrane Collaboration is an international not-for-profit and independent organization, dedicated to making up-to-date, accurate information about the effects of health care readily available worldwide. It produces and disseminates systematic reviews of health care interventions and promotes the search for evidence in the form of clinical trials and other studies of interventions. The Cochrane Collaboration was founded in 1993 and named after the British epidemiologist, Archie Cochrane." [75]*

This collaboration, though originating in the UK, has branches in every continent for a total of 21 branches in 19 countries [69] (including Canada and the US). They produce a major EBM resource known as the Cochrane Database of Systematic Reviews. This collection is one of the sources (along with DARE, CENTRAL and others [67]) available as part of the Cochrane Library. Decisions regarding changes to its reviews are evaluated by committees of volunteers known as Cochrane Review Groups [68], which are made up of mostly medical professionals. Strictly organized and constantly updated, these reviews provide status flags which act as visual indicators of the content any changes in the Library. In **Figure 2**, you can see 3 (New Search, Conclusions changed and Review) of the 9 flags (Review, Protocol, Methodology, New, New Search, Conclusions changed, Major change, Withdrawn, and Comment) used. Though freely available in Canada, UK and much of Europe, limited public access in the United States has prevented its universal adoption [58].

3.4 Up-to-date, Dynamed, Google and Wikipedia

3.4.1. UPTODATE.COM

UptoDate.com [61] is a commercial, Internet-based service which provides medical information directed at primary care medical practitioners. As the name indicates, the published monographs from this source are regularly updated by its 3,800 author/editors, clinical experts all of whom are listed on the website. The service is available off-line for $1500 (with quarterly updates for a year), or on-line and on PDA for $500/ year. UptoDate.com practices many aspects of EBM including structured queries (like PICO) and uses the GRADE [48] system to indicate Strength of Recommendation.

In a 2008 observational study of 424 hospitals increased usage of UpToDate.com (measured in hits per week) were "significantly associated with a shortened severity-adjusted length of stay and lower risk-adjusted patient safety adverse outcome rates" [59]. This study also showed that the 424 hospitals with UpToDate compared against the 3091 hospitals without UpToDate "were associated with significantly lower risk-adjusted complication rates and patient safety adverse outcome rates" [59]. This second point loses its potency when you notice that this is an observational study. That is to say, confounding factors must be considered. For example, there is the possibility that hospitals with UpToDate.com subscriptions were better hospitals on the whole and therefore provided better care and lower complication rates and a generally shorter length of stay.

3.4.2. DYNAMED

DynaMed is a similar regularly updated subscription-based service available on-line and on PDAs (Palms, Pocket PC, Windows Smartphone, BlackBerry, and iPhone). This site EBM based service uses the Strength of Recommendation Taxonomy (SORT) [45] to delineate the strength of evidence. The source material for the reviews within DynaMed are searched using PubMed Clinical Queries, the Cochrane Database of Systematic Reviews, and the National Guideline Clearinghouse is the source for medical guidelines. A complete list of primary and secondary sources is available on the DynaMed website [63].

3.4.3. GOOGLE AND WIKIPEDIA

Though patients may get nervous of the idea of doctors googling their symptoms on the Internet, there is mention of its use in the literature. Google scholar is used by doctors [76], sometimes preferred [33], and there is some evidence that it does provide decent results [33].

Wikipedia's quality is steadily increasing as is its reputation. Though still frowned upon in a court of law [64] and the medical office, some improvements and recent developments must be noted. The combination of concretely referenced articles which hyperlink to reputable sources, the integration of clinical taxonomies such as MeSH and the familiarity of its interface make it a useful starting point for many clinical queries and in some cases adequate for general information needs. Recent studies [76, 77] find that while only 10% of doctors edit or contribute to Wikipedia's content [77], nearly 50% use it for clinical queries [76] and that it is nearly error-free on the topic of drugs [77].

In both these cases, the use of these web services are inevitable due to the off-duty habits of clinicians and the reality that these resources are both easy to use and familiar, two qualities that are potent and desperately needed elements of a clinical information service. These popular services set the expectations, and for better or worse any other service must contend with them as competitors.

4. Knowledge-Based Sub-systems

4.1. Ontologies

A multitude of definitions and theory surround the concept of an ontology. Here, we will define what it is and what it is not in the context of this paper, based on the research in the subject. An ontology primarily serves as a tool to solve the problem of semantic ambiguity.

If I were to say "speaking in the language of a statistician... the result was 'statistically significant'" you would have a sense of the word 'significant' according to the domain of mathematical statistics. That is, according to the distribution and the experimental design we have a result which could lead to causation. A very specific meaning in which "statistics study" is the context for interpretation. Where the common use of "...the result was significant" would not carry those specific mathematical connotations. The ontology is the conceptualization of that domain knowledge, a context.

The ontology is a framework of communication. For two agents to agree on a subset of meanings (or senses) of a body of terminology is to agree to 'commit' to a specific ontology. So, an ontology needs only to define the terms of communication; two agents that 'commit' to an ontology are agreeing on a shared vocabulary. The deeper needs of answering arbitrary queries and solving problems is the concern of the knowledge base [2]. The ontology makes as few constraints about the world it is modeling as possible to maintain a consistent terminology and maximize the freedom of the ontology committal agents to instantiate as needed [2].

4.2. WordNet

The creators of WordNet would not consider it an ontology, but rather, an on-line lexical database, a dictionary designed for a computer to read. Where human dictionaries define words with a list of descriptive sentences, WordNet is more like a thesaurus, defining words in relation to other words which share its meaning – more specifically – share the same sense. Each of these "w ord-senses" are a collection of synonyms called "syn sets" and are meant to represent one distinct concept. Presently, WordNet contains 155,327 words, 117,597 synsets and 207,249 word-sense pairs [78].

Defining words in a way computers can use to interpret human communication means addressing the ambiguous ways humans use language: **polysemy,** the same word form may belong to more than one set; **synonymy**, different word-forms belong to the same synset; **hyper/hyponymy**, noun synsets must be organized hierarchically to represent "I SA" relations; **meronymy**, synsets which are conceptually related components of each other must indicate "H ASA" relations; to name a few. Thus, the demarcation of these synsets and the definition of their relationships was not a trivial task. A task that was performed painstakingly by George Miller and his team of linguists at Princeton from 1985–1995, and is on-going, with the most recent version released in 2006. With this difficult groundwork laid, WordNet (free to download and use) has become a central resource for computational linguists, so much so that 434 papers have been published on WordNet [79] and the conference dedicated to its study and use is now in its 5th year [80].

4.3. UMLS

> *"T he objective of this program . . . is to solve what is the most fundamental barrier to the application of computers in medicine; namely, the lack of a standard language in medicine. We will attempt to build that vocabulary, a language that will cross between the biomedical literature and the observations on the patient, as well as the educational applications in the school, a language which allows those areas to be interrelated. – Donald A. B. Lindberg, M.D., March 19, 1985["3]*

The UMLS is not strictly a formal ontology as described in the first sub-section of this section. The UMLS is more similar to WordNet, that is, organized like a very precise thesaurus with several distinct frameworks of hierarchical and semantic relations added to its structure [8].

Before I begin describing the details of the UMLS, I would like to make some clarifications. The UMLS is a project, an acronym, which stands for (U)nified (M)edical (L)anguage (S)ystem. Under the umbrella of this project are several components. First, there are 3 knowledge sources the UMLS Metathesaurus; UMLS Semantic Network and the SPECIALIST Lexicon and Lexical tools [36]. Second, there is the UMLS Knowledge Server. Thirdly, the MetaMap program and finally the RRF Browser. In general, when people refer to the UMLS they are referring to the UMLS Metathesaurus, Semantic Network and SPECIALIST Lexicon in combination.

The UMLS Metathesaurus attempts to integrate all of the disparate and specialized medical terminologies, categorizations and thesauri into one unified superset hence the name 'Metathesaurus'. It includes more than 100 source vocabularies from the entire domain of medicine, including such varied sources as:

- Diagnostic and Statistical Manual of Mental Disorders (DSM-IV),
- HCPCS Version of Current Dental Terminology,
- WHO Adverse Drug Reaction Terminology (WHOART),
- Standard Product Nomenclature (USFDA)...

A complete list is available from http://www.nlm.nih.gov/research/umls/metaa1.html.

The metathesaurus attempts to tackle the problem of synonymy – different lexical forms (words) with the same meaning – by linking synonymous words to distinct, unique (and numbered) concepts it has defined. This way all synonymous concepts from all the source materials can be equated, allowing a framework for the exchange of knowledge between these vocabularies. The 2007 release of the UMLS Metathesaurus contains information about 1,436,586 biomedical concepts [36], has over 5 million concept names [17] from 123 controlled vocabularies, and is available (at least in part) in 17 different languages [36].

The Semantic Network provides categorization for all the concepts represented in the UMLS Metathesaurus and adds a hierarchical semantic structure to the Metathesaurus through a set of semantic types and relations between these types [22]. This is done in the attempt to tackle the problem of hyponymy, for example, *ibuprofen* is a subclass of *anti-inflammatory,* and both are is a subclasses of *drug.* All terms from every source vocabulary is linked to at least one concept in the Metathesaurus. All concepts in the Metathesaurus are linked to at least one of the 135 semantic types in the current Semantic Network. These Semantic types are related to each other by at least one of the 54 relationships currently in use by the Semantic Network [34].

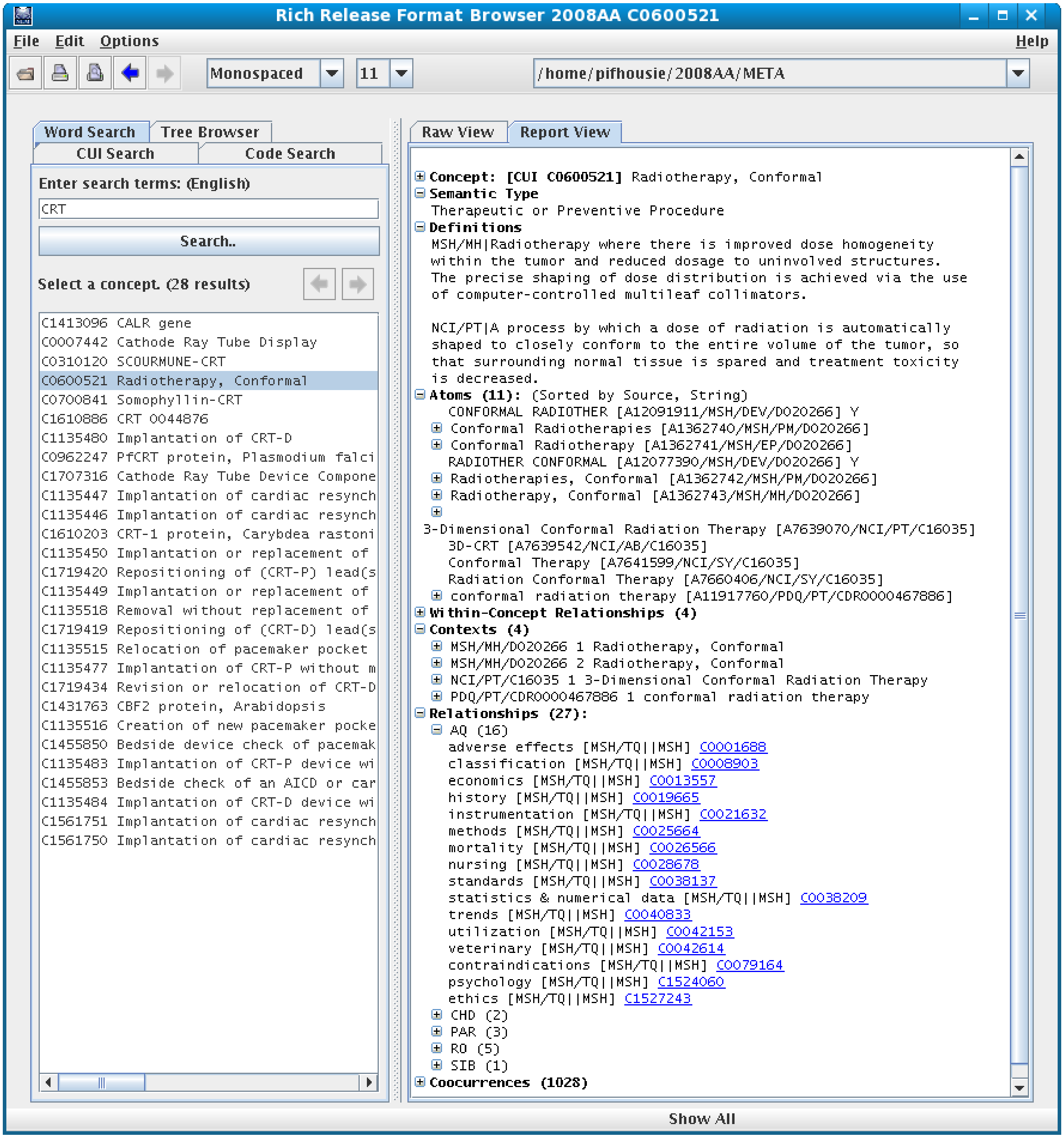

***Figure 6**: UMLS Metathesaurus search results for "CRT" as shown in the RRF browser. On the left side we see the search term and results. "Conformal Radiotherapy" is selected. In the Report View on the right hand side we see: the unique concept ID (CUI); the semantic type (taken from the Semantic Network); a short definition; variants; contexts (showing in which taxonomies the term is represented); and relationships shows connections to other concepts.*

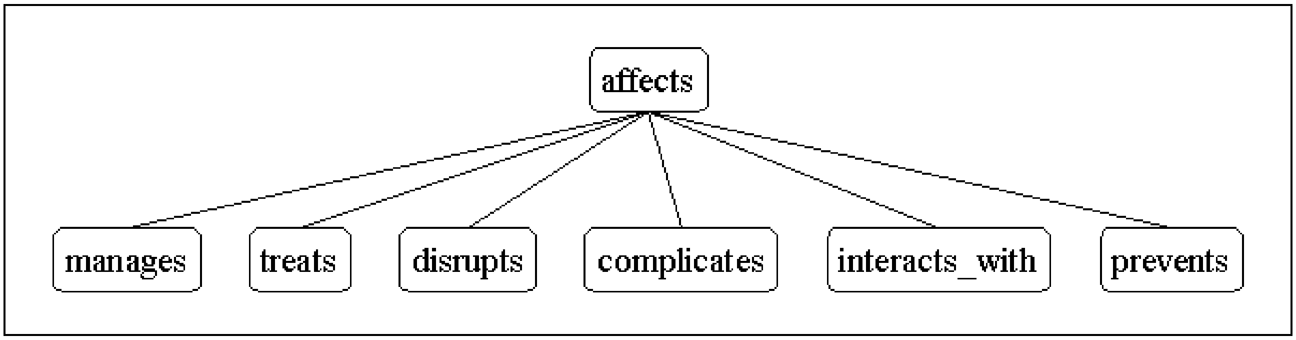

Figure 3*: An example hierarchy for network relationships; the relationships used in the Semantic Network are they themselves hierarchically related with ISA relationships [60].*

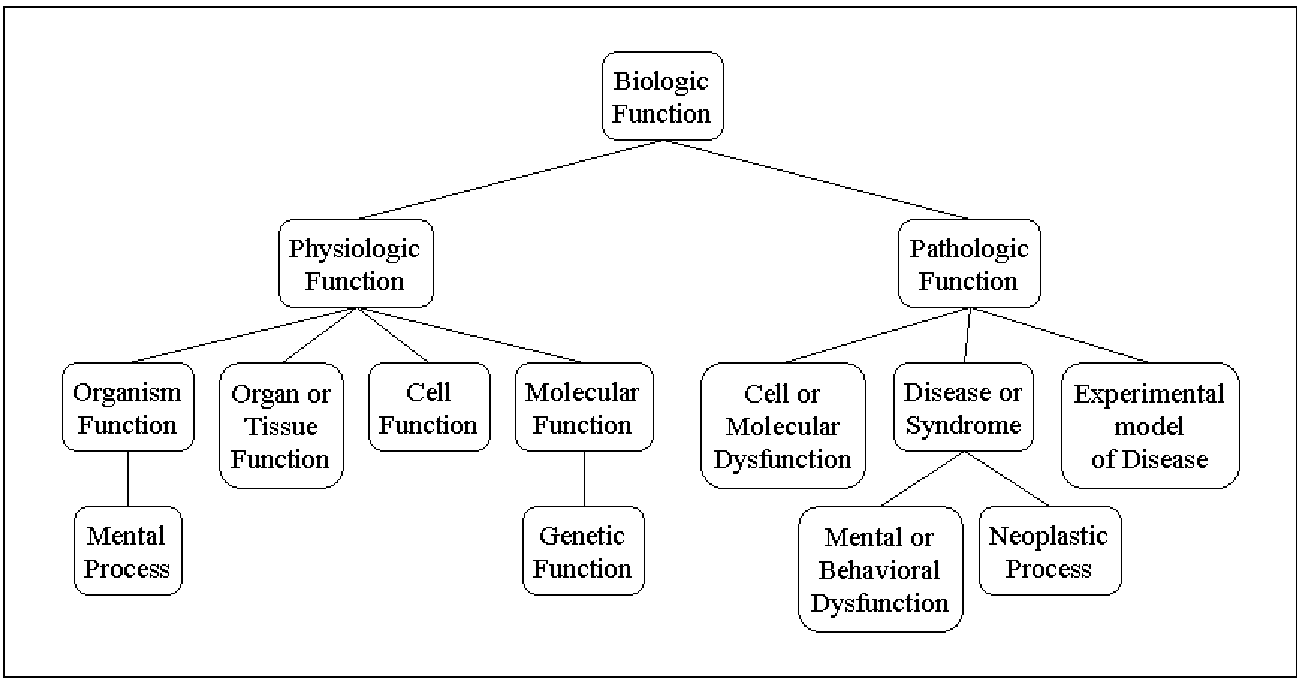

Figure 4*: This sample from the Semantic Network shows ISA relationships.*

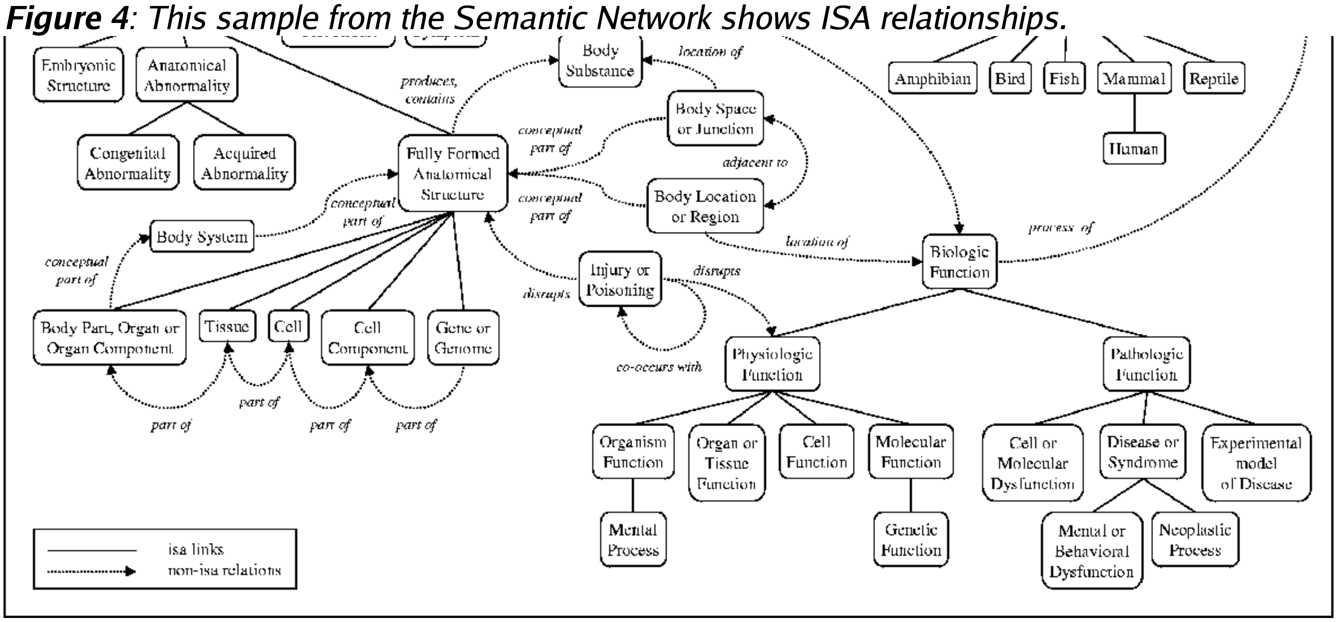

Figure 5*: Shown here are example hierarchical and associative relationships between semantic types in the Semantic Network [60].*

The primary relationship in the Semantic Network is the "isa" relationship (**Figure 3**), this is used to create the hierarchy of concepts necessary to solve semantic issues arising from hyponymy. In addition, five major categories of **associative relationships** are defined which are themselves relations: "physically related to", "spatially related to", "temporally related to", "functionally related to", and "conceptually related to" [34]. **Figure 4** shows an example hierarchy of an associative relationship and **Figure 5** shows examples of associative relationships and hierarchical relationships in a single graph representation.

The SPECIALIST Lexicon and NLP Tools

The need was recognized for a bridge between the UMLS Metathesaurus and free text applications. These components of the UMLS were developed to foster development of – and for use in – natural language processing and information retrieval systems. The SPECIALIST lexicon "is a syntactic lexicon of biomedical and general English words, providing orthographic, morphological and syntactic information," [22] has 297K records (over 482K inflectional forms) [36]. There are 6 tools in the 2008 SPECIALIST NLP Toolkit, they are open source, freely available and each is developed specifically for a standard NLP task.

Tokenization - **Wordind** - Wordind is a tokenizer and word index generator.

Normalization - Norm – Normalizes strings and words into the a form preferred by the UMLS Metathesaurus that is ignoring alphabetic case, inflection, spelling variants, punctuation, genitive markers, stop words, diacritics, symbols, ligatures, and word order. [39]

Part-of-speech tagging - **dTagger** is a Part of Speech (POS) tagger specifically built for use in the medical domain. It includes a trained model, one trained on a set of annotated MEDLINE abstracts from MedPost corpus (genomics) [39].

Spell Checking - **GSpell** is a spell checker but it treats a space as a letter allowing the correction of errors in word compounding. [39]

LexAccess2008/2009

To allow easy access to the SPECIALIST Lexicon, **LexAccess2008** is provided. It is written in Java and provides Java API's for use as a component in other applications or can be used as an end-user tool. Here is example output from this tool:

```
$> CRT
{base=CRT
entry=E0420176
   cat=noun
   variants=uncount
   variants=groupuncount
   variants=plur
   variants=metareg
   acronym_of=Certified Record Techniques
   acronym_of=cardiac resuscitation team|E0420190
   acronym_of=cathode-ray tube|E0420189
   acronym_of=choice reaction time|E0420188
   acronym_of=chromium release test|E0420187
   acronym_of=complex reaction time|E0420186
   acronym_of=computerized renal tomography|E0420185
   acronym_of=copper reduction test|E0420184
   acronym_of=corrected retention time|E0420183
   acronym_of=cortisone resistant thymocyte|E0420182
```

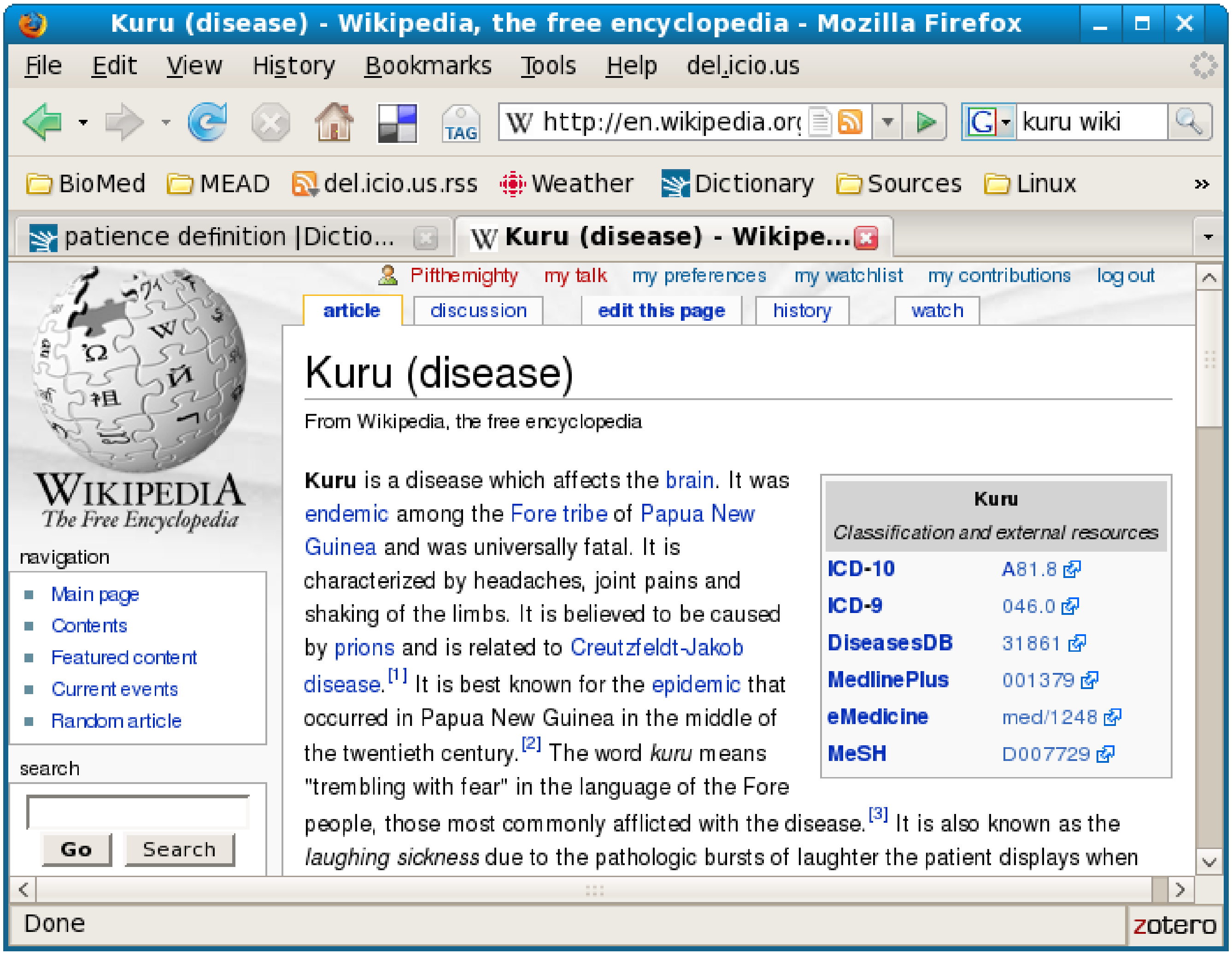

***Figure 7**: shows Wikipedia's search results for the term “Kuru,” a disease. You can see on the right side to box containing links to several resources including MedlinePlus and MeSH.*

```
   acronym_of=cranial radiation therapy|E0420181
   acronym_of=capillary refilling time|E0420180
   acronym_of=chemoradiation therapy|E0420179
   acronym_of=conformal radiation therapy|E0420178
}
{base=CRT
entry=E0420177
   cat=adj
   variants=inv
   position=attrib(3)
   position=pred
   stative
   abbreviation_of=certified|E0220630
   abbreviation_of=corrected
}
{base=Crt
spelling_variant=CRT
entry=E0420191
   cat=noun
   variants=uncount
   acronym_of=calreticulin|E0304049
}
```

```
{base=cRT
entry=E0420192
   cat=noun
   variants=uncount
   acronym_of=competitive reverse transcriptase|E0420193
}
{base=CrT
entry=E0420194
   cat=noun
   variants=metareg
   acronym_of=crista terminalis input site|E0420195
}
```

And another:

```
$> be
{base=be
entry=E0012152
        cat=aux
        variant=be;infinitive
        variant=is;pres(thr_sing)
        variant='s;pres(thr_sing)
        variant=isn't;pres(thr_sing):negative
        variant=are;pres(fst_plur,second,thr_plur)
        variant='re;pres(fst_plur,second,thr_plur)
        variant=aren't;pres(fst_plur,second,thr_plur):negative
        variant=am;pres(fst_sing)
        variant='m;pres(fst_sing)
        variant=was;past(fst_sing,thr_sing)
        variant=wasn't;past(fst_sing,thr_sing):negative
        variant=were;past(fst_plur,second,thr_plur)
        variant=weren't;past(fst_plur,second,thr_plur):negative
        variant=been;past_part
        variant=being;pres_part
}
{base=Be
entry=E0581585
        cat=noun
        variants=uncount
        abbreviation_of=beryllium|E0012491
}
{base=BE
entry=E0581586
        cat=noun
        variants=metareg
        variants=uncount
        acronym_of=base excess|E0581581
        acronym_of=Barrett's esophagus|E0217024
        acronym_of=Barrett esophagus|E0581582
        acronym_of=beta-endorphin|E0012611
        acronym_of=benzoylecgonine|E0303750
        acronym_of=bacterial endocarditis|E0011702
        acronym_of=barium enema|E0011948
        acronym_of=butoxyethanol|E0424081
        acronym_of=bronchiectasis|E0014147
        acronym_of=brain edema|E0013941
        acronym_of=bioequivalence|E0013004
        acronym_of=branching enzyme|E0013962
        acronym_of=bicycle ergometry|E0581583
        acronym_of=backscattered electron|E0581584
        abbreviation_of=bestatin|E0303779
        abbreviation_of=benzene|E0012391
}
```

In the first example you can see sensitivity to spelling variants such as 'CrT', Crt' and 'cRT' demonstrating potential pitfalls due to the ambiguous nature of the domain. Not only are there 14 possible acronyms of 'CRT' could be referring to, but 1 abbreviation (an adjective) and 3 spelling variants. The second example reinforces this point by demonstrating an expected list of verb tenses for the word "be" and an unexpected list (at least to me) of 17 domain specific acronyms and abbreviations are also returned. This demonstrates not only how simple it is to find ambiguity in this domain, but also points to the use of this tool as a piece of the solution.

MetamorphoSys

The third component, MetamorphoSys, 'the UMLS installation wizard and Metathesaurus customization tool, installs one or more of the UMLS Knowledge Sources and enables us to create customized Metathesaurus subsets' [40]. Part of the MetamorphoSys package is the RRF browser, which allows us to browse your customized installation of the Metathesaurus, as shown in **Figure 6**.

MetaMap

The last tool from the NLM to be discussed is MetaMap. MetaMap. Also known as MMTx, it was developed to map biomedical text to the Metathesaurus. MetaMap is also used to semi-automatically relate MeSH terminology to MEDLINE papers [39]. It is semi-automatic in that human indexers approve MetaMap's choices, by selecting the specific MeSH terms on which both MetaMap and the indexers agree and removing the others [40]. Two of the Q&A systems discussed later, CQA-1.0 and Essie, make use of MetaMap.

As a word of warning, installation the UMLS Knowledge sources takes about 2-12 hours and requires about 20GB to 45GB of storage (our installation took 5 hours and totals 42.5GB). MetaMap must be installed separately.

4.4. MeSH

MeSH, is an acronym which stands for Medical Subject Headings. This terminology was first created by the NLM as a terminology for organizing medical literature. In the 2008 version of MeSH there are 24,767 descriptors. The top level of the MeSH Hierarchy:

1. Anatomy [A]
2. Organisms [B]
3. Diseases [C]
4. Chemicals and Drugs [D]
5. Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]
6. Psychiatry and Psychology [F]
7. Biological Sciences [G]
8. Natural Sciences [H]
9. Anthropology, Education, Sociology and Social Phenomena [I]
10. Technology, Industry, Agriculture [J]
11. Humanities [K]
12. Information Science [L]
13. Named Groups [M]
14. Health Care [N]

15. Publication Characteristics [V]
16. Geographicals [Z]

For example, if you were to look at the children of "Anatomy":

Anatomy [A]

- Body Regions [A01] +
- Musculoskeletal System [A02] +
- Digestive System [A03] +
- Respiratory System [A04] +
- Urogenital System [A05] +
- Endocrine System [A06] +
- Cardiovascular System [A07] +
- Nervous System [A08] +
- Sense Organs [A09] +
- Tissues [A10] +
- Cells [A11] +
- Fluids and Secretions [A12] +
- Animal Structures [A13] +
- Stomatognathic System [A14] +
- Hemic and Immune Systems [A15] +
- Embryonic Structures [A16] +
- Integumentary System [A17] +

(http://www.nlm.nih.gov/mesh/MBrowser.html)

Besides these descriptors there are 97,000 entry terms, synonymous with descriptors, to assist entry to the MeSH system, for example "Heart Attack" is an entry term for "Myocardial Infarction." As well as being the key indexing and categorization paradigm for the NLM, this MeSH terminology is one of the source vocabularies in the UMLS Metathesaurus.

As point of interest, the MeSH terminology has begun to infiltrate the main stream. If you were to type in a disease name into Wikipedia most times a call-out box in the upper right-hand corner displays links to the MeSH terminology (see F**igure 7** below). Clicking on word "MeSH" will take you to the Wikipedia entry for "Medical Subject Headings" and clicking on the number beside it will take you to the entry in the MeSH browser on the NLM website.

5. Current Solutions to Health Information Needs

5.1. Informationist

> *We believe it's time to face up to the fact that physicians can't, and shouldn't, try to do all or even most medical information retrieval themselves. ...Better they should focus their scarce discretionary professional time on reading, discussing, and reflecting in ways that truly deepen their conceptual and practical understanding of*

medicine than on the mechanics of finding, extracting, and synthesizing information from the published literature [11].

The idea behind this solution is to create a position akin to a medical librarian. A person whose primary responsibility is to answer doctors clinical questions, present during rounds, available after out-patient visits, to be seen as an important member of the medical team. A person trained in equal parts clinical work and information science [11]. It has been shown that they often help clinicians formulate their questions [11], which is one of the major obstacles in clinical question answering [20,53].

Unfortunately, on-site medical librarians and 'informationists' are uncommon outside academic centers [11]. As an alternative, off-site clinical question answer services have be suggested, established and studied.

A system in the UK called ATTRACT, started in 1997, would deliver (via Fax) an Evidence-Based medicine summary created by an information manager within 6 hours. This service was rated 'useful' by 31% and 'very useful' by 69% of the 40 doctors participating in the study. Over half said the summaries changed their practice [9]. The average cost per question was $27.30. In an similar study in Australia, questions were answered for a fee of 27.50 per question and questions were answered within 1 to 12 days [9]. Following the study all 9 doctors said they were willing to pay for the service and 50% said they would use it at least twice per month. The time to respond to questions was seen as an important factor is the perceived usefulness, by the participating doctors [9].

These services go a long way to help solve obstacles in answering clinical questions, such as “lack of time,” “difficulty formulating questions,” “selection of resources,” and “difficulty finding optimal search strategies” [9,20].

If this service is seen as an alternative (in some cases) to blood tests or scans, one could quickly alleviate a portion of the workload on expensive equipment and services with a price tag of $27.50 per question which is well below the price of even the most inexpensive test.

5.2. Essie

...O, be some other name!
What's in a name? that which we call <concept-name='a rose'>
By any other name would smell as [valid];

-William Shakespeare, Romeo and Juliet, 1594.

Essie (formerly referred to as SE) [26] is a concept-based search engine developed in 2000 at the NLM for it's site ClinicalTrials.gov [29,30]. Essie's ranking algorithm can be best described as “all the right pieces in all the right places” [26]. Since the search engine is phrase based, (as opposed to single word tokenized), 'the right pieces' are these phrases from the query. Since the engine heuristically ranks locations in the structure of the document differently, such as the title which is ranked high and footnotes which are ranked low, 'the right places' are where in the document these pieces are found [26].

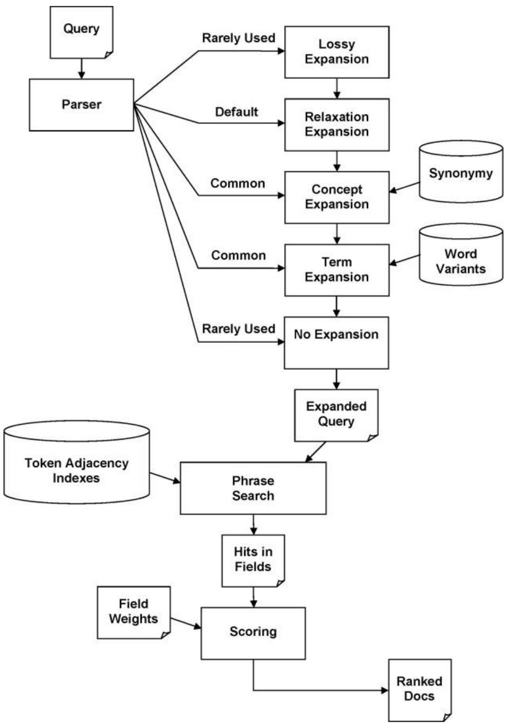

Figure 8*: The Essie query architecture.*

Essie uses the UMLS to identify concepts and as the basis for synonymous phrase-based query expansion. Once a phrase token is identified the UMLS is used to identify the concept it references, Essie then 'expands' the query by adding other synonymous phrases, phrases which also reference the same concept in the UMLS, thus searching for matches of all synonymous phrases, and thus searching for the concept 'by any other name' not just the queried phrase.

This concept based query expansion necessitates unusual documents scoring, not on the usual *how many* occurrences (frequency) but on *where* in the document (location) a concept is placed. This is due to the fact that phrase proliferation generated by query expansion equates many very different phrases [26].

Essie competed in the 2003 and 2006 TREC Genomic track. In 2003 it was the best performing search-engine and in 2006 it "a chieved results comparable to those of the highest-ranking systems" [26].

One advantage of the concept-based system is the utilization of the 97000 entry terms from MeSH (as part of the UMLS). For example the common term for 'ascorbic acid' is 'vitamin C', Essie would relate the query term 'vitamin C' to the concept 'ascorbic acid' (via the UMLS) and search both, performing innate translation of many common usage words into clinical terminology.

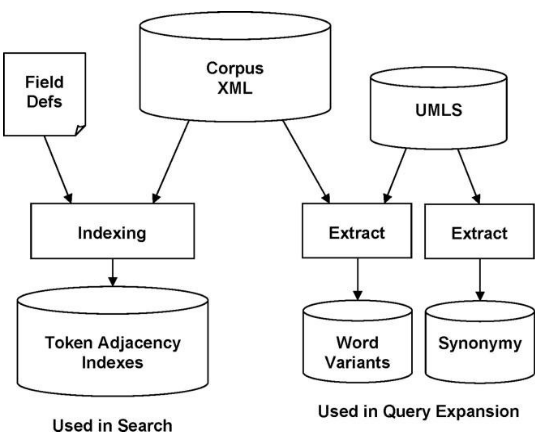

Figure 9*: The Essie index architechure.*

Essie has two main phases. First it indexes the search corpus, by tokenizing and recording the position of every token occurrence in the corpus. Position information is important for ranking results, and determining token adjacency, which is key information for phrase matching (i.e. words in a phrase are adjacent). This indexing results in a look-up table shown in **Figure 8**.

During this phase two other tables are generated, the **synonymy dataset** and the **variant dataset**. The former using concept expansion, the latter using term expansion. Concept expansion uses the UMLS in the manner described above and term expansion, uses the UMLS SPECIALIST Lexicon to include term variants such as plurals, possessives, hyphenation, compound words and alternate spellings but not non-noun inflectional variants.

The second phase is the search phase. In this phase the query is entered, parsed, broken into fragments (called relaxation expansion in **Figure 8**) and then expanded through concept expansion then term expansion. All of the phrases generated are searched, scored and returned according to rank, as shown in **Figure 9**.

There are a few limitations to this system. A relatively static corpus is a must to allow time required for the extensive indexing [26]. Query expansion has its dangers as "the explosive nature of the expansions makes the implementation vulnerable to failure when given a very long query" [26].

It is important to note the hardware requirements of this type of a architecture: "The strategies used by Essie are computationally expensive and resource intensive. The token adjacency index can be up to ten times the size of the document set. The Essie Medline implementation makes use of servers that have a considerable amount of random-access memory (64 GB) to reduce the use of slower disk access" [26]. These are very heavy requirements when compared to the MedQA system – discussed next – which runs on a personal computer.

5.3. MedQA

Since 40% of medical questions are "What is..." questions [71], the MedQA question answering system focused on these definitional questions. It uses both the World Web Web via Google and MEDLINE (indexed with Lucene) as resources for answering questions. MedQA is available for use on-line at: http://askhermes.org/MedQA/. The MedQA team plan on adding other question types as they development of the system.

In MEDLINE, normally articles that report original research use a document structure known by the mnemonic: IMRAD (Introduction, Methods, Results, and Discussion) [33]. The authors took advantage of this structure in two distinct ways. First, to determine relevancy of an given article to the query, the "Results" section was the focus since a recent user study had shown that physicians prefer this section when determining the relevancy [33]. Second, in identifying definitional sentences they found these sentences were more likely to be found in the Introduction and Background sections [33]. The authors made a training set of sentences classified by section, then used machine learning techniques (naïve Bayes) to classify unknown sentences from MEDLINE into the classes Introduction, Background, Methods, Results, Conclusion and Other with 78.6% accuracy.

To retrieve candidate sentences, all non-factual statements had to be filtered out. To do this "MedQA applied cue phrases (e.g., suggest, potential, likely, may, and at least)... which was reported to outperform machine-learning approaches, to separate facts from speculations" [33] in combination with the writing convention "...to use the past tense when reporting original work and present tense when describing established knowledge" [33].

The retrieval of definitions from these candidate sentences required a clever use of Google, "...we applied all of the terms that are included in the Unified Medical Language System (UMLS 2005AA) as candidate definitional terms, and crawled the Web to search for definitions. We built our crawler on the Google:Definition service. ...With this set of definitions, we then automatically identified lexico-syntactic patterns that comprise the definitions" [33]. In a nutshell, they used Google to build a corpus of definitional phrases, then used probabilistic machine learning techniques on the corpus to build a system to identify definitional phrases. A beneficial side effect of this process was a pre-processed static collection of 36,535 definitions of the 1 million UMLS terms.

To initiates a search, the user first types in a definitional question. Next, MedQA identifies noun phrases, and forms a query with only these terms. The query is then used to retrieve relevant documents which are tokenized into sentences and clustered. To generate an answer, centroid-based summarization is applied twice. First, to remove redundancies, MedQA selects one sentence based on TF*IDF weighted cosine similarity to be the most representative of its cluster. Then again to the collection of selected representative sentences to generate a final coherent summary. The user receives a result separated in two sections; Web and MEDLINE, (see **Figure 10**).

Web search has its pitfalls. On-line definitions can often be irrelevant to the medical domain. "For example, "heart" was defined as both "a... ne of the most successful female fronted bands in the annals of hard rock" and "a hollow, muscular organ that pumps blood through the blood vessels by repeated, rhythmic contractions;" [33] To deal with this problem on-line medical dictionaries are also queried, the TF*IDF scores are then compared, if this similarity measure fall below a given threshold the web result is discarded.

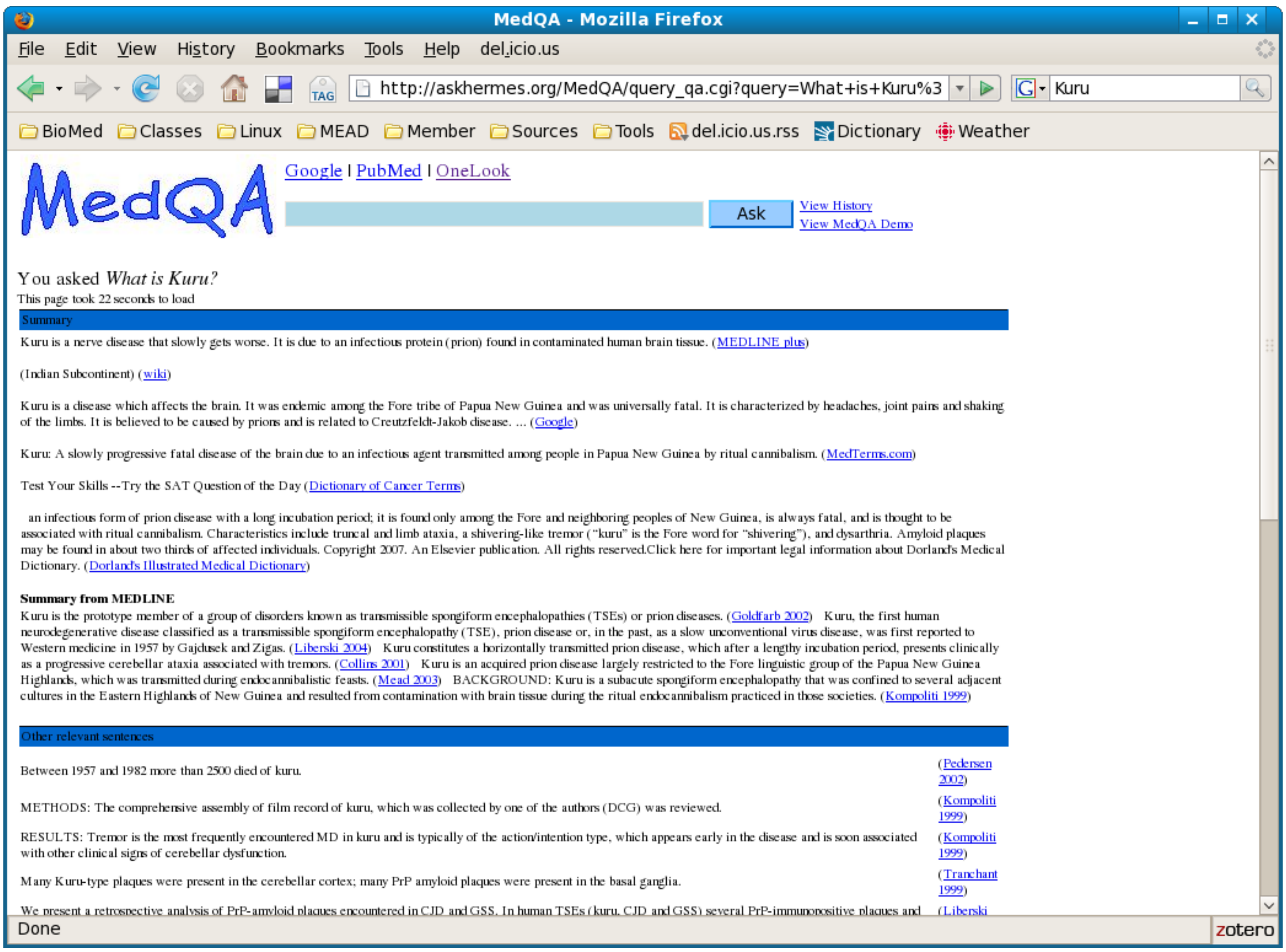

Figure 10. *MedQA search results for the term "Kuru." In summary section, each extracted sentence is followed by a link to the source and each source in the "Summary from*

MEDLINE" subsection is hyperlinked. The second 'Other relevant sentences' section provides highly ranked non-definitional extractions all of which are liked to primary sources through MEDLINE.

MedQA was evaluated by four physicians in comparison to three other on-line systems Google, One-Look and PubMed. We assume reader's familiarity with Google, and PubMed was described in detail previously. OneLook, however, requires a brief description. OneLook is a federated search engine which has indexed over 900 other dictionaries [70]. A search on OneLook returns relevant results from any dictionary it has indexed. Results appear in the form of a list of hyperlinks to the source site, broken into categories such as General, Art, Science and most importantly Medicine.

Evaluation results indicated PubMed and OneLook were bettered in most evaluation criteria, quality of answer, ease of use, time taken, and actions taken, by Google and MedQA. Qualitative test scores, gathered by questionnaire showed Google was the preferred system in terms of ease of use and quality of answer overall. Quantitatively, Google provided ranked results in less then a second and MedQA generated its summary in an average of 16 seconds. However, since information is spread among sites, the evaluation of Google results (identifying definitions) was more time consuming. MedQA was the highest rated system in terms of time spent and number of actions.

An interesting and encouraging point is one of hardware requirements, MedQA was written in Perl and runs on a Macintosh PowerPC with dual 2 GHz CPUs and 2GB of memory [33].

5.4. CQA-1.0

This prototype Q&A system was developed by the NLM around the fundamentals of EBM: PICO built questions, strength of evidence and clinical task type. This system views the clinical answering task "...as 'semantic unification' between information needs expressed in a PICO-based frame and corresponding structures automatically extracted from MEDLINE citations" [17]. The idea is that "EBM offers three orthogonal facets that, when taken together, provide a framework for codifying the knowledge involved in answering clinical questions." These facets are:

(1) the four main clinical tasks:

> **Therapy:** Selecting treatments to offer a patient, taking into account effectiveness, risk, cost, and other relevant factors (includes Prevention—selecting actions to reduce the chance of a disease by identifying and modifying risk factors).
>
> **Diagnosis:** This encompasses two primary types: Differential diagnosis: Identifying and ranking by likelihood potential diseases based on findings observed in a patient. Diagnostic test: Selecting and interpreting diagnostic tests for a patient, considering their precision, accuracy, acceptability, cost, and safety.
>
> **Etiology/Harm:** Identifying factors that cause a disease or condition in a patient.
>
> **Prognosis:** Estimating a patient's likely course over time and anticipating likely complications.[17]

(2) a well-built clinical question (Patient/Problem, Intervention, Comparison, and Outcome. (PICO)):

- What is the primary *problem* or disease? What are the characteristics of the patient (e.g., age, gender, or co-existing conditions)?
- What is the main *intervention* (e.g., a diagnostic test, medication, or therapeutic procedure)?
- What is the main intervention *compared to* (e.g., no intervention, another drug, another therapeutic procedure, or a placebo)?
- What is the *desired effect* of the intervention (e.g., cure a disease, relieve or eliminate symptoms, reduce side effects, or lower cost)? [17]

(3) strength of evidence, they use the *Strength of Recommendations Taxonomy* (SORT):

- ***A-level evidence*** is based on consistent, good-quality patient outcome-oriented evidence presented in systematic reviews, randomized controlled clinical trials, cohort studies, and meta-analyses.
- ***B-level evidence*** is inconsistent, limited-quality, patient-oriented evidence in the same types of studies.
- ***C-level evidence*** is based on disease-oriented evidence or studies less rigorous than randomized controlled clinical trials, cohort studies, systematic reviews, and meta-analyses. [17]

The authors do not believe that free-form natural language queries are well-suited to question-answering systems. Instead, their system structures the query according to the familiar PICO framework, a standard of the EBM curriculum. The benefit is the physician – instead of the system – translates their information need into a frame-based representation [17], a problematic interpretation for a computer system. This interface also “...force[s] physicians to “think through” their questions” [17] which leads to better “thought out” queries.

Here are some example questions and their PICO query frames:

> ***Does quinine reduce leg cramps for young athletes?*** (Therapy)
> *search task:* therapy selection
> *primary problem:* leg cramps
> *co-occurring problems:* muscle cramps, cramps
> *population:* young adult
> *intervention:* quinine
>
> ***How often is coughing the presenting complaint in patients with gastroesophageal reflux disease?*** (Diagnosis)
> *search task:* differential diagnosis
> *primary problem:* gastroesophageal reflux disease
> *co-occurring problems:* cough
>
> ***What's the prognosis of lupoid sclerosis?*** (Prognosis)
> *search task:* patient outcome prediction

primary problem: lupus erythematosus
co-occurring problems: multiple sclerosis

What are the causes of hypomagnesemia? (Etiology)
search task: cause determination
primary problem: hypomagnesemia [17]

Though the system looks exclusively at abstracts, structured abstracts are common, though varied in naming structure. This system takes advantage this structure when present.

5.4.1 SAMPLE OUTPUT FROM PICO EXTRACTORS

The PICO extractors parse the abstract, tagging phrases and sentences as 'Problem', 'Population', 'Intervention', or 'Outcome'. It was noted that outcomes are usually complete sentences and tagged as such, while interventions, population, and problems are noun phrases [17]. In this sample output from the PICO extractors the *<u>italic underlined text</u>* is the extracted text and the subscript immediately following is the tag. This sample is in response to the question "*In children with an acute febrile illness, what is the efficacy of single-medication therapy with acetaminophen or ibuprofen in reducing fever?*" [17]

Antipyretic efficacy of ibuprofen vs acetaminophen
Kauffman RE, Sawyer LA, Scheinbaum ML
Am J Dis Child. 1992 May;146(5):622-5

OBJECTIVE–To compare the antipyretic efficacy of ibuprofen, placebo, and acetaminophen. DESIGN–Double-dummy, double-blind, randomized, placebo-controlled trial. SETTING–Emergency department and inpatient units of a large, metropolitan, university-based, children's hospital in Michigan. PARTICIPANTS– *<u>37 otherwise healthy children aged 2 to 12 years</u>*$_{\text{Population}}$ *<u>with acute, intercurrent, febrile illness</u>*$_{\text{Problem}}$. INTERVENTIONS–Each child was randomly assigned to receive a single dose of *<u>acetaminophen</u>*$_{\text{Intervention}}$ (10 mg/kg), *<u>ibuprofen</u>*$_{\text{Intervention}}$ (10 mg/kg) (7.5 or 10 mg/kg), or <u>placebo</u>$_{\text{Intervention}}$ (10 mg/kg). MEASUREMENTS/MAIN RESULTS– Oral temperature was measured before dosing, 30 minutes after dosing, and hourly thereafter for 8 hours after the dose. Patients were monitored for adverse effects during the study and 24 hours after administration of the assigned drug. <u>All three active treatments produced significant antipyresis compared with placebo</u>.$_{\text{Outcome}}$ *<u>Ibuprofen provided greater temperature decrement and longer duration of antipyresis than acetaminophen when the two drugs were administered in approximately equal doses</u>*.$_{\text{Outcome}}$ No adverse effects were observed in any treatment group. CONCLUSION–*<u>Ibuprofen is a potent antipyretic agent and is a safe alternative for the selected febrile child who may benefit from antipyretic medication but who either cannot take or does not achieve satisfactory antipyresis with acetaminophen.</u>*$_{\text{Outcome}}$

Publication Type: Clinical Trial, Randomized Controlled Trial
PMID: 1621668
Strength of Evidence: grade A [17]

The user could be shown any of the extracted they prefer, the default is to only return the outcome sentences (with the title, and bibliographic information) as studies show they are key sentences in abstracts for determining relevancy [17]. This system uses MetaMap to map noun phrases into the concepts in the UMLS. This mapping must be applied differently to each of the elements of the PICO frame, "...problems and interventions can be directly mapped to UMLS concepts, and populations easily mapped to patterns that include UMLS concepts, outcome statements follow no predictable pattern" [17].

Population and problem were separated due to there conceptual differences and the fact that often the are not presented together in abstracts. Intervention and comparison were merged as they are conceptually similar and difficult for the system to distinguish. The identification of these elements is dependent on the semantic groups MetaMap links to:

concept	UMLS Semantic Group (and its children)
population	GROUP
problem	DISORDER
intervention	DIAGNOSTIC PROCEDURE, CLINICAL DRUG, and HEALTH CARE ACTIVITY

Outcomes are full sentences unlike the other elements. Each sentence in the abstract is given a score to determine the likelihood that it is an outcome sentence, then the system returns all those rated above a certain threshold. The score is a combination of elements captured in the following formula: $S_{outcome} = \lambda 1S_{cues} + \lambda 2S_{unigram} + \lambda 3S_{n\text{-}gram} + \lambda 4S_{position} + \lambda 5S_{length} + \lambda 6S_{semantic\ type}$. Here is a brief description of the components of the above formula:

S_{cues}	uses cue phrases heuristically developed by the team
$S_{unigram}$	uses a 'bag-of-words" classifier from the MALLET toolkit
$S_{n\text{-}gram}$	developed on corpus of positive outcome predictors using odds ratio
$S_{position}$	closer to the end of the abstract is better
S_{length}	a probability based on the length of the abstract that it contains an outcome statement
$S_{semantic\ type}$	contains UMLS concepts related to outcome statements.

Strength of evidence evaluation is based on three components. First, the most recent articles are given greater weight. Second, highly trusted sources are given greater weight. The third weight depends on the where the type of study sits in the SORT taxonomy. The type of determined by is determined by the metadata associated with the article. Specifically the MeSH tags which accompany most MEDLINE citations and the publication type:

Evidence	MeSH/Publication type
Level A	Meta-analysis, randomized controlled trials, cohort study, follow-up study
Level B	Case-control study, case series
Level C	Case report, in vitro, animal and animal testing, alternatives studies [17]

To determine the clinical task, MeSH terms (and their children) are categorized into indicators of the four task types:

Clinical Task	Positive indicators	Negative indicators
Therapy	MeSH Terms: CLINICAL TRIALS, RANDOM ALLOCATION and THERAPEUTIC USE	
Diagnosis	MeSH Terms: DIAGNOSIS	Positive Therapy

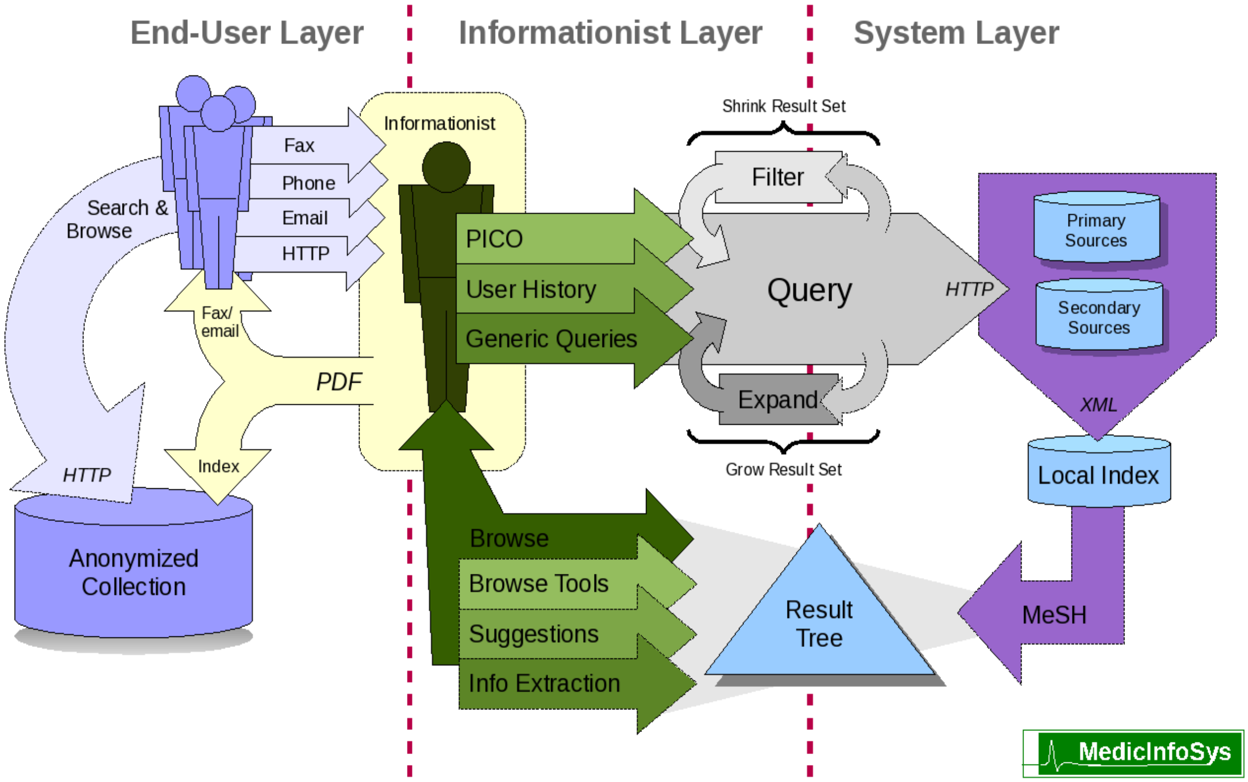

Figure 11: *A proposed architecture for this MedicInfoSys (pronounced: "medicine-fo-sys") medical information system is divided into 3 layers: 1) the End-user layer; 2) the Informationist layer; and 3) the System layer. The dotted lines represent these boundaries and indicate the interface of each user to the rest of the system to right.*

		indicators
Prognosis	MeSH Terms: SURVIVAL ANALYSIS, DISEASE-FREE SURVIVAL, TREATMENT OUTCOME, HEALTH STATUS, PREVALENCE, RISK FACTORS, DISABILITY EVALUATION, QUALITY OF LIFE, and RECOVERY OF FUNCTION.	
Etiology	MeSH Terms: POPULATION AT RISK, RISK FACTORS, ETIOLOGY, CAUSALITY, and PHYSIOPATHOLOGY.	Positive Therapy indicators

If any of these terms are marked as a major theme (indicated by a * next to the MeSH term) that terms weight is increased.

5.4.2. RESULTS

The baseline for comparison are expertly generated boolean PubMed queries. Each of these queries took an average of 40 minutes for the first author (a Medical Librarian & Medical Doctor) to generate. These go far beyond the ability of your average PubMed user, but definitely demonstrate the system against the most expert of PubMed users. For example, the question, '*What is the best treatment for analgesic rebound headaches?*' resulted in the

PubMed query:

> ((("analgesics"[TIAB] NOT Medline[SB]) OR "analgesics"[MeSH Terms] OR "analgesics"[Pharmacological Action] OR analgesic[Text Word]) AND (("headache"[TIAB] NOT Medline[SB]) OR "headache"[MeSH Terms] OR headaches[Text Word]) AND ("adverse effects"[Subheading] OR side effects[Text Word])) AND hasabstract[text] AND English[Lang] AND "humans"[MeSH Terms][17]

The relevancy was evaluated based on following criteria:

P10 - Precision at ten retrieved documents (P10) measures the fraction of relevant documents in the top ten results.
MAP - Mean Average Precision (MAP) is the average of precision values after each relevant document is retrieved.
MRR - Mean Reciprocal Rank (MRR) is a measure of how far down a hit list the user must browse before encountering the first relevant result.
TDRR - Total Document Reciprocal Rank (TDRR) is the sum of the reciprocal ranks of all relevant documents.

The results from the lenient test, meaning all results that are helpful or contain the answer count as a successful result. All results with '*' following the score are results which are statistically significant at the 1% level.

	P10	*MAP*	*MRR*	*TDRR*
PubMed (baseline)	0.281	0.356	0.526	1.353
Term	0.481 (+29%)*	0.481 (+71%)*	0.513 (+44%)	1.945 (+44%)*
EBM	0.677 (+141%)*	0.718 (+102%)*	0.936 (+78%)*	2.671 (+98%) *
Combo	0.688 (+145%)*	0.718 (+102%)*	0.962 (+83%)*	2.680 (+98%) *

The results from the strict test, meaning only results that contain the answer to the question counts as a successful result. All results with '*' following the score are results which are statistically significant at the 1% level.

	P10	*MAP*	*MRR*	*TDRR*
PubMed (baseline)	0.069	0.045	0.190	0.328
Term	0.150 (+117%)*	0.092 (+105%)*	0.346 (+82%)*	0.632 (+93%)*
EBM	0.196 (+183%)*	0.129 (+187%)*	0.433 (+127%)*	0.765 (+133%)*
Combo	0.219 (+217%)*	0.138 (+207%)*	0.494 (+160%)*	0.851 (+159%) *

CQA-1.0 and Essie were used in combination with the idea that a fusion of output of the systems might provide better results. "For 15 questions, the CQA-1.0 improvement over PubMed is statistically significant ($p < 0.01$), and so is the improvement of the fused results of

SemRep, Essie, and CQA-1.0 over the individual systems and the baseline." [4]

5.5. Semantic Clustering

Semantic clustering is an attempt to improve the way we return results. Instead of returning a ranked list of results, semantic clustering returns the results grouped by concept. To indicate the content of each cluster, UMLS concepts are used to label clusters, this was the users gets an overview of the results inside.

The idea is "that grouping retrieved MEDLINE® citations into semantically-coherent clusters, based on automatically-extracted interventions from the abstract text, represents an effective strategy for presenting results, compared to a traditional ranked list. Experiments with our implemented system appear to support this claim" [27].

"The system starts by assigning each intervention (and the associated abstract) to its own cluster, and then iteratively merges clusters whose interventions share a common UMLS hypernym, ascending the UMLS hierarchy in the process....To avoid forming clusters under labels that are too general to be of interest, we truncated the tops of the UMLS hierarchies" [27].

Semantically-related results are organized into clusters. Cluster names are presented to the user a cluster can be selected and the contents of the cluster are displayed. Inside the cluster, each article is displayed as a short extractive summary of three parts: the title, the main intervention, and the top-scoring outcome sentence [23]. The extracted outcome sentence is the automatically identified using CQA-1.0 [17]. The outcome sentence serves as an entry point into the article, which the reader can use to judge relevance. The clusters are ordered by size (number of articles), the articles inside each cluster are sorted chronologically (newest first).

The idea is to "drill-down" into the information, to view your results at deepening levels of granularity, unlike Google which presents pages of results which hyperlink out of the Google interface. This would, through series of turn-down switches, show more information about an article at each deeper level. For example, "Top-level answers to 'What is the best drug treatment for X?' consist of categories of drugs that may be of interest to the physician. Each category is associated with a cluster of abstracts from MEDLINE about that particular treatment option. Drilling down into a cluster, the physician is presented with extractive summaries of abstracts that outline the clinical findings. To obtain more detail, the physician can pull up the complete abstract text, and finally the electronic version of the entire article (if available)" [23].

This clustering method was compared to lexical clustering, clustering based solely on keyword, not only did it not improve the PubMed baseline but the cluster names were incoherent and therefore unhelpful in organizing results [27,23].

There are 4 advantages we would like to mention. First is redundancy management; that is, redundant information is gathered together, since all interventions in a particular cluster are conceptually related [23]. Secondly the concepts labels give an information overview; in this form of presentation the categorized results provide a feel for information landscape that is something difficult to get a sense of from browsing a ranked list. Third, obviously irrelevant articles are bundled together and can be categorically ignored, saving the precious resources of time, patience and screen "real-estate". And finally, it provides an opportunity for easy semantic-based relevance feedback. Clusters can be selected and deselected to indicate preference and focus iterative searches.

6. Proposed Architecture

In this section we present an architecture to improve delivery of information to those working in the medical domain, **Figure 11** is a diagram of this architecture and Appendix A is a larger annotated version of this diagram. The following sub-sections describe its aspects.

6.1. End-User Layer

The highly valued time of medical professionals is limited and overworked, this coupled with their varying degrees of comfort, competency and fluency with information technology points to the need for delegation to a specialist. In this layer, the end user's view of the system is represented. Once they identify an information need they have two avenues of to find answers: search the collection of previous adequately answered medical queries for their query/answer pair, or communicate their information need to an Informationist. In some cases the request will be a one-way communication of of a well-defined or generic sort, other times the user will benefit from interaction with the Informationist to help the user clarify their need and focus it into a precise query.

6.1.2. END USERS

The intended core users of this system would be those directly involved with patient care (such as general practitioners, surgeons, specialists, nurses, psychiatrists, rehabilitators, therapists). However, the information needs of health administrators, social scientists, health advisors and public servants may be well served by this system. The primary sources in collections such as PubMed, contain the newest research and statistics on all health related fields including economics, ethics, social trends, prevention, law, and technology; which are valuable sources of information for making informed budgetary decisions and strengthening policy positions.

6.1.3. PDF REPORT

The keys to this system are brevity, accessibility and transparency (of its evidence-based sources). The time constraints placed on medical professionals are severe, therefore customized reports must be timely and brief. In addition, with the stakes as high as they are in the medical field, all reports must be aggregate in nature, gathered from trusted EBM sources, and explicitly referenced from sources accessible to the end-user preferable immediately via the Internet. Experiments with a similar phone/fax system in the UK reports [9] the mean time to answer clinical questions at 45 minutes, a maximum length of two pages (one sided), a maximum turn-around time of 8 hours and an average cost of $27.50 per answer.

The PDF is a format of choice for medical articles, compact and secure. Many freely available readers exist for any platform, it prevents manipulation, capable embedding high-resolution images for medical diagrams, hyperlinks to web-resources, and has security layers to password protect printing and viewing which is a priority if handling private patient data.

This PDF is faxed/emailed to the user (users preference) and sent to searchable collection for later reference. If the user approves the content the report is anonymized, indexed and made available to all authorized users of the system.

6.1.4. INTERFACE WITH INFORMATIONIST LAYER

Bad assumptions and ambiguity create extra-work and wasted effort in every workplace in any domain where one person gives a task to another. In a domain as sophisticated and time-sensitive as this is, it is imperative to make an extra effort to ensure that all parties are "on the same page." The structured query is a way to mitigate time wasted because of the delegation of search tasks; to promote a mutual understanding and limit time-wasting irrelevant results.

We need to structure the query to tease out the users' information need. The user usually does not instantly know the best words or how to phrase an uncertainty. A structured model like PICO helps the user through the uncertain process of developing a query. Along the way a structured query contextualizes the components (i.e. query keywords) for the system implicitly, thus reducing the ambiguity inherent in common unqualified keyword searches. The qualifications (such as Population) help the system/informationist to narrow query expansion, prioritize the "where" a word of phrase was found according to rhetorical structure (e.g. introduction, methods...) for the purposes of ranking, summarization and information extraction. Other qualifications clarify what task (such as diagnosis) the doctor is engaged in, which will help the system direct the search, better rank what is relevant, and best frame the answer. To put it a different way, help them give us the pieces we need to solve their question by helping them formulate the query through a process we define, and define that process with terminology they are familiar with and we can compartmentalize.

6.2. Informationist Layer

This layer concerns the use of the system from the Informationists point of view. This position as defined in the literature [11] historically goes by many names: medical librarian, medical researcher or medical knowledge worker. The qualified user is one with a medical background strong enough to research the most detailed technical medical questions health professionals have to offer and to provide answer or direct the end-user to the source of the answer to their information need. This service could be provided by a single government agency, a single private company, a selection of regional or specialist providers, or by accredited public, private and individual "freelance" service providers. Renumeration could be based on an hourly fee, on a per query basis, a monthly retainer, or yearly contracts. Quality Assurance could be guaranteed by the particular end-user using the service in combination with a regulatory body. Since Canada's public healthcare system would pay for this service the resultant research can be collected free of copyright restrictions, indexed and made available as web-based public medical information resource.

6.2.1. INPUT

The ways the user has to input queries into the system have been categorized into three avenues: PICO, User Profile and Generic Questions.

6.2.1.1. PICO

In the PICO query input category the user enters the appropriate information into specialized fields (Population, Intervention, Comparison, Outcome). This information is used to formulate and query the system. PICO is the most common EBM 'well-built question' method, but several others exist including PICOTT and PEDCOR the user may select the method which is most appropriate for their information need.

6.2.1.2. User Profile

All queries in the User Profile input category depend on some user specific information. For example, user search history is needed to revisiting past query result sets; and private information access is necessary for automatic query generation based on specific electronic patient records. A set of articles selected from the results of previous queries can be selected and used to automatically generate queries-by-example, a user profile is needed to collect, store and retrieve this set.

6.2.1.3. Generic Queries

Certain questions are common enough that effective heuristics have been developed to answer them specifically. This input category helps users make use of these customized search methods. The user selects from a short list of generic questions, fills in the appropriate fields, then the system employs the optimized search method developed for that generic query type. Examples of these generic questions include: “What is the cause of symptom X?”, “What is the recommended dosage of X?”, and “How should I manage disease or finding X?”. These three generic questions taken together make up the majority of all generic clinical queries [71].

6.2.1.4. Filter

This filtering module is meant to narrow results sets and increase precision. The three input categories above influence the filtering of the query, but the user may customize these filters arbitrarily. These filters include: the Clinical Query Filters developed at McGill University, specifying the Strength of Evidence by various models, the use of standard Boolean Operators, and a set of filters like those used in PubMed's Advanced Search to limit publication type, specify database, journal, author, etc. Also, an automatic disambiguation function will detect ambiguous word candidates and interact with the user to disambiguate candidate terms (e.g. “Did you mean... X or Y?”)

6.2.1.5. Expand

This Query Expansion module is meant to broaden results sets and increase recall. Like the filter module this module is partly in the domain of the system layer, but the user has influence and can make customizations. This module uses the UMLS as its primary means of query expansion. The UMLS Metathesaurus can be used for the identification medical concepts in the query, allowing for the addition of synonyms for those terms into the query and for queries based on concept ids for databases equipped to accept them. The UMLS Semantic Network can be utilized to add related concepts appropriate to the query. Abbreviations, metonymy, hypo/hypernymy can also be addressed and/or exploited by this module, increasing recall and affecting better re-ranking.

6.2.2. OUTPUT

The output of the system is displayed in a browsable hierarchical tree structure based-on the MeSH controlled vocabulary. Results are organized into categories based on the conceptual attributes they represent, an they are ordered by their degree of lexical similarity to the query. This structure retains its state during and after browsing, providing a predictably behaving structure for interaction, easing a users re-visitation of previous query result sets and maintaining progress from complete or interrupted browsing sessions. Nodes, branches and subtrees may be deleted at the users discretion, for the purpose of trimming dead ends and focusing the result set. These result set maybe saved for future reference and used as the basis for a query-by-example type automation.

6.2.2.1. Browse Tools

This category represents tools meant to aid the user in navigating the result set and include: keyword highlighting, "Find in MeSH" and secondary search. Keyword highlighting is fairly straight-forward, the user enters terms they would like to be highlighted in the results to draw attention to them as they scan the results. "Find in MeSH" is a tool that locates MeSH concepts by name in the result tree hierarchy and centers the view over them for the users convenience. Secondary search is meant to focus a given result set by searching within it using a keyword search.

6.2.2.2. Suggestions

These navigation aids are meant to use the information available on the user and the query to direct the user to likely starting points (MeSH categories which match query terms) for browsing, to rank results within each category in the tree and to recommend articles based on the results so-far selected, based on the user profile and based on information task.

6.2.2.3. Information Extraction

As needed and appropriate automatic extractive summarization will be utilized. Some queries may be sufficiently specific, adequately detailed or heuristically predisposed to precise answers or customized extractions. For example, well-structured clinical queries and certain structural features of many medical articles can enable outcome extraction which is very useful for determination of relevance for many clinical users. Also identification of salient sentences is likely for some query types and all properly defined queries will produce candidates sentences for extraction and those ranked highly enough to surpass a heuristically based threshold would be presented to the user using this tool.

6.2.3 UPDATE PDF FILES

Quotes sources could be checked against Cochrane Library by any user viewing the PDF after the creation date for updates and corrections. If an update to an existing PDF is requested from the user layer, the references in the existing PDF would be used to search for new papers which use them as references, as a starting point for the Informationist.

6.3. System Layer

6.3.1. SOURCES

Since all sources recommended are available on-line they will be queried via HTTP. Many sources make their results available via XML, further research is need to investigate XML availability for each on-line system. The results from all sources will be collected and indexed into one local database for every individual query.

6.3.1.1. Primary Sources

Primary Sources are the unfiltered root sources of medical information, medical journals and Clinical trials. The primary source in this category in MEDLINE. MEDLINE, PubMed and all NLM collections can be searched via HTTP with a publicly available tool developed by the NLM known as *efetch*. Results can be returned as XML formatted data. NLM sources are freely available and well-respected.

6.3.1.2. Secondary Sources

Secondary sources are collected, edited aggregates of the primary sources. These include clinical guidelines, Cochrane reviews, UpToDate.com, Dynamed, Micromedex, Harrison's on-line and Wikipedia. All these sources listed are known to be used by physicians and have varying licensing agreement, levels of quality and utility. Some heuristics for generic queries use specific collections in this set.

6.3.2. LOCAL INDEX

The local index is the collection point for all these disparate sources of on-line information. Since indexing is done during run-time and results are to be displayed to the user as categorized a boolean indexing system is recommended. Stop-words would be removed and a Porter stemmer would be applied to the set. This local index would be used primarily building the MeSH tree, and searched for secondary searching and keyword highlighting.

6.3.3. MESH-BASED BROWSE TREE

Each article is categorized into a MeSH conceptual heading these articles are then placed in a tree where each leaf is an article and the parent nodes between the leaf and the root are the concepts names in the MeSH hierarchy. These categories are very general near the root and increase in specificity the further from the root in the hierarchy you venture. The tree starts with an upper bound of 16 main MeSH categories and is at maximum 11 levels deep. The user navigates this hierarchy in the same way as a file hierarchy where folders are concepts and files are results.

With a hierarchical model, uninteresting results are often categorized together, these subtrees or leaves can be minimized, ignored or deleted. If the user wants to get a feel for the information landscape – for what is 'out there' – the hierarchal structure acts as an informative guide for exploration, each node a signpost, rather than a long uninformative ranked list where each result is independent of the one before and the one after.

7. Future Work

To build this system we must first build the components and test them, layer by layer (layers shown in **Figure 11**), starting with the Systems layer, then the Informationist layer and finally the End-User layer.

Work has begun on the IR system: the local indexer, MeSH categorizes and browsable result tree. A prototype has been built, pilot study has been completed, a second prototype is in development and a user study is planned. Following this user study, the structured query input, query expansion/filtering and information extraction components would be added to the IR prototype. Following systems testing, a user-study with an informationist demographic user-base would be completed using questions picked randomly from a large collection of actual clinical questions that is available on the Web [83]. Once these systems perform acceptably together, the tools and protocols needed to interact with the Anonymized Collection will be developed. Once these layers behave and perform satisfactorily, a user study with Canadian physicians would take place and be evaluated. Iterations of user-study, development and deployment would be necessary until stake-holders are satisfied.

Once stake-holders are satisfied, the system would be put on-line and the service promoted.

8. Conclusion

Our architecture puts physicians first. We have researched the information needs physicians, the complex linguistic nature of the medical information, and the mature knowledge-based resources available in the medical domain and suggest a solution based on that research.

A system to make the best use of the human knowledge worker in support of physicians is at the heart of this work. One could postulate, that some day these knowledge workers could be replaced by an AI agent, presently and for the foreseeable future we still need trained and knowledgeable people to perform these tasks. Still this system is needed to meet the needs of medical professionals which are short on time.

There are two outside perspectives on this service that accurately position it in the mind of the end-user and give a good idea of how it is intended to be used. First, the paradigm in which this service should fit, to be used and properly thought of by physicians is of the same category as a blood test or a biopsy, one which aids medical understanding and assists in the tasks of diagnosis, prognosis, therapy or etiology. The second perspective is to view this as the consultation of a specialist, an information specialist. This is a colleague – an expert in the ways of information – to listen and help answer questions for physicians, to allow physicians to focus on what they do best: diagnose, treat, and heal patients.

To succeed it is vital that this system be centralized to give equal access to those in remote regions, to share in the benefits provided by academic centers which generally only exist in the larger centers. Note that a central database can be anywhere as the URL is of course consistently located in cyberspace. It is the informationists which could be anywhere and need not be centralized; in fact, several main agencies such as academic centers, hospitals, NGOs and government bodies could work in tandem.

In short, we have provided a survey of medical knowledge-based resources, discussed linguistic difficulties in the medical context, detailed the information needs and obstacles facing physicians and provided a domain- and user-specific solution to these obstacles in the form of an architecture. We have also outlined future work to implement this architecture and test the resulting system.

9. Acknowledgments

This research was supported by NSERC USRA program, as well as the NSERC CRD grant with support from AlphaGlobal IT Inc.

10. References

1. Spasic, Irena, Sophia Ananiadou, John McNaught, and Anand Kumar. “Text mining and ontologies in biomedicine: Making sense of raw text.” *Brief Bioinform* 6, no. 3 (January 1, 2005): 239-251.
2. Gruber, T. R. “Towards Principles for the Design of Ontologies Used for Knowledge Sharing.” In *Formal Ontology in Conceptual Analysis and Knowledge Representation*, edited by N. Guarino and R. Poli. Deventer, The Netherlands: Kluwer Academic Publishers, 1993.
3. Humphreys, B L, D A Lindberg, H M Schoolman, and G O Barnett. (1998) “The Unifed Medical Language System: an informatics research collaboration.” *Journal of the American Medical Informatics Association: JAMIA* 5, no. 1: 1-11.
4. Sneiderman, Charles A., Dina Demner-Fushman, Marcelo Fiszman, Nicholas C. Ide, and Thomas C. Rindflesch. “Knowledge-based Methods to Help Clinicians Find Answers in MEDLINE.” *J Am Med Inform Assoc* 14, no. 6 (November 1, 2007): 772-780.
5. Demner-Fushman, Dina, Susan Hauser, and George Thoma. “The role of title, metadata and abstract in identifying clinically relevant journal articles.” *AMIA Annual Symposium Proceedings* (2005): 191-5.
6. Huang, Xiaoli, Jimmy Lin, and Dina Demner-Fushman. “Evaluation of PICO as a knowledge representation for clinical questions.”*AMIA Annual Symposium Proceedings* (2006): 359-63.
7. Schardt, Connie, Martha B Adams, Thomas Owens, Sheri Keitz, and Paul Fontelo. “Utilization of the PICO framework to improve searching PubMed for clinical questions.” *BMC Medical Informatics and Decision Making* 7 (2007).
8. Zweigenbaum, P. “Question answering in biomedicine.” In EACL 2003 *Workshop on Natural Language Processing for Question Answering.*, 2003.
9. Coumou, Herma C H, and Frans J Meijman. “How do primary care physicians seek answers to clinical questions? A literature review.” *Journal of the Medical Library Association: JMLA* 94, no. 1 (January 2006): 55-60.

10. Brassey, J, G Elwyn, C Price, and P Kinnersley. “Just in time information for clinicians: a questionnaire evaluation of the ATTRACT project.” *BMJ (Clinical Research Ed.) 322*, no. 7285 (March 3, 2001): 529-30.

11. Davidoff, F, and V Florance. “The informationist: a new health profession?” *Annals of Internal Medicine* 132, no. 12 (June 20, 2000): 996-8.

12. Smith, R. “What clinical information do doctors need?” BMJ (Clinical Research Ed.) 313, no. 7064 (October 26, 1996): 1062-8.

13. Hale, Roger. “Text mining: getting more value from literature resources.” *Drug Discovery Today* 10, no. 6 (March 15, 2005): 377-379.

14. Hunter, Lawrence, and K. Bretonnel Cohen. “Biomedical Language Processing: What's Beyond PubMed?” *Molecular Cell* 21, no. 5 (March 3, 2006): 589-594.

15. Zweigenbaum, Pierre, Dina Demner-Fushman, Hong Yu, and Kevin B Cohen. “Frontiers of biomedical text mining: current progress.” *Briefings in Bioinformatics* 8, no. 5 (September 2007): 358-75.

16. Cohen, K. Bretonnel, and Lawrence Hunter. “Getting Started in Text Mining.” *PLoS Computational Biology 4*, no. 1 (January 2008).

17. Demner-Fushman, Dina, and Jimmy Lin. “Answering Clinical Questions with Knowledge-Based and Statistical Techniques.” *Computational Linguistics* 33, no. 1 (2007): 63-10

18. Cios, Krzysztof J., and G. William Moore, “Uniqueness of medical data mining,” *Artificial Intelligence in Medicine* 26, no. 1-2 (2002): 1-24.

19. Krallinger, Martin, Ramon Alonso-Allende Erhardt, and Alfonso Valencia, “Text-mining approaches in molecular biology and biomedicine.” *Drug Discovery Today* 10, no. 6 (March 15, 2005): 439-445.

20. Ely, John W, Jerome A Osheroff, M Lee Chambliss, Mark H Ebell, and Marcy E Rosenbaum. “Answering physicians' clinical questions: obstacles and potential solutions.” *Journal of the American Medical Informatics Association*: JAMIA 12, no. 2 (2005): 217-24.

21. Aronson, A R, and T C Rindflesch. “Query expansion using the UMLS Metathesaurus.” *Proceedings: A Conference of the American Medical Informatics Association / AMIA Fall Symposium* (1997): 485-9.

22. McCray, A T, A R Aronson, A C Browne, T C Rindflesch, A Razi, and S Srinivasan. “UMLS knowledge for biomedical language processing.” *Bulletin of the Medical Library Association* 81, no. 2 (April 1993): 184-94

23. Demner-Fushman, Dina, and Jimmy Lin. “Answer extraction, semantic clustering, and extractive summarization for clinical question answering.” In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL*, 841-848. Sydney, Australia: Association for Computational Linguistics, 2006.

24. McCray, A T, A Burgun, and O Bodenreider. “Aggregating UMLS semantic types for

reducing conceptual complexity.” *Studies in Health Technology and Informatics* 84, no. Pt 1 (2001): 216-20.

25. Demner-Fushman, D., S. M. Humphrey, N. C. Ide, R. F. Loane, P. Ruch, M. E. Ruiz, et al. “Finding relevant passages in scientific articles: fusion of automatic approaches vs. an interactive team effort.” *In Proceedings of the Fifteenth Text REtrieval Conference*. Gaithersburg, MD, 2006.
26. Ide, Nicholas C, Russell F Loane, and Dina Demner-Fushman. “Essie: a concept-based search engine for structured biomedical text.” *Journal of the American Medical Informatics Association*: JAMIA 14, no. 3: 253-63.
27. Lin, J., and D. Demner-Fushman. “Semantic Clustering of Answers to Clinical Questions.” *AMIA Annual Symposium Proceedings.* Chicago, 2007.
28. Wilczynski, N L, K A McKibbon, and R B Haynes. “Enhancing retrieval of best evidence for health care from bibliographic databases: calibration of the hand search of the literature.” *MEDINFO* 10, no. Pt 1 (2001): 390-3.
29. ClinicalTrials.gov. 2008. Available at: http://clinicaltrials.gov/ Accessed August 26, 2008.
30. McCray, A T, and N C Ide. “Design and implementation of a national clinical trials registry.” *Journal of the American Medical Informatics Association*: JAMIA 7, no. 3: 313-23.
31. Graber, Mark A., Bradley D. Randles, John W. Ely, and Jay Monnahan. “Answering clinical questions in the ED.” *The American Journal of Emergency Medicine* 26, no. 2 (February 2008): 144-147.
32. Ely, J W, J A Osheroff, P N Gorman, M H Ebell, M L Chambliss, E A Pifer, et al. “A taxonomy of generic clinical questions: classification study.” *BMJ (Clinical Research Ed.) 321*, no. 7258 (August 12, 2000): 429-32.
33. Yu, Hong, Minsuk Lee, David Kaufman, John Ely, Jerome A. Osheroff, George Hripcsak, et al. “Development, implementation, and a cognitive evaluation of a definitional question answering system for physicians.” *Journal of Biomedical Informatics* 40, no. 3 (June 2007): 236-251.
34. U.S. National Library of Medicine. “2008AA Section 3 UMLS SEMANTIC NETWORK.” Technical Documentation. http://www.nlm.nih.gov/research/umls/meta3.html. Accessed September 9, 2009.
35. Jurafsky, Daniel, and James H. Martin. *Speech and Language Processing. 1st ed.* New Jersey, NJ, USA: Prentice Hall Inc., 2000.
36. Alan Aronson, Allen Browne, Anantha Bangalore, Betsy Humphreys, Brian Carlsen, Chris Lu, David Sheretz, Guy Divita, John Kilbourne, Karen Thorn, Kin Wah Fung, Laura Roth, Mark Tuttle, Olivier Bodenreider, Stephanie Lipow, Steve Emrick, Stuart Nelson, Suresh Srinivasan, Tammy Powell, Tom Rindflesch, Vivian Auld, and William Hole. “UMLS_Basics.pdf.” http://www.nlm.nih.gov/research/umls/pdf/UMLS_Basics.pdf. Accessed August 26, 2008.

37. NLM. "Appendix A.1 2008AA UMLS Appendix to the License Agreement." Technical Documentation. http://www.nlm.nih.gov/research/umls/metaa1.html. Accessed August 26, 2008.

38. NLM. "SPECIALIST Lexicon." Fact Sheets. http://www.nlm.nih.gov/pubs/factsheets/umlslex.html. Accessed August 26, 2008.

39. Aronson, A R. "Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program." *AMIA Annual Symposium Proceedings. AMIA Symposium* (2001): 17-21.

40. Aronson, Alan R, James G Mork, Clifford W Gay, Susanne M Humphrey, and Willie J Rogers. "The NLM Indexing Initiative's Medical Text Indexer." *Studies in Health Technology and Informatics* 107, no. Pt 1 (2004): 268-72.

41. "PubMed Clinical Queries." http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.shtml. Accessed August 26, 2008.

42. Chang, Jeffrey T, Hinrich Schütze, and Russ B Altman. "Creating an online dictionary of abbreviations from MEDLINE." *Journal of the American Medical Informatics Association: JAMIA* 9, no. 6: 612-20.

43. Liu, Hongfang, Stephen B Johnson, and Carol Friedman. "Automatic resolution of ambiguous terms based on machine learning and conceptual relations in the UMLS." *Journal of the American Medical Informatics Association: JAMIA* 9, no. 6: 621-36.

44. Richardson, W S, M C Wilson, J Nishikawa, and R S Hayward. "The well-built clinical question: a key to evidence-based decisions." *ACP Journal Club* 123, no. 3 (November 1995): A12-3.

45. Ebell, Mark H, Jay Siwek, Barry D Weiss, Steven H Woolf, Jeffrey Susman, Bernard Ewigman, et al. "Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature." *American Family Physician* 69, no. 3 (February 1, 2004): 548-56.

46. "SORT: the strength-of-recommendation taxonomy." *American Family Physician* 71, no. 1 (January 1, 2005): 19-20.

47. Phillips, Bob, and Chris Ball. "Levels of Evidence." *Centre for Evidence-Based Medicine*, May 11, 2007. *http://www.cebm.net/index.aspx?o=1025*. Accessed August 26, 2008.

48. Atkins, David, Dana Best, Peter A Briss, Martin Eccles, Yngve Falck-Ytter, Signe Flottorp, et al. "Grading quality of evidence and strength of recommendations." *BMJ (Clinical Research Ed.) 328*, no. 7454 (June 19, 2004): 1490.

49. "GRADE working group." *http://www.gradeworkinggroup.org/intro.htm*. Accessed August 26, 2008.

50. Bakken, S. "An informatics infrastructure is essential for evidence-based practice."

Journal of the American Medical Informatics Association : JAMIA 8, no. 3 (January 22, 2001): 199-201.

51. Klein, M S, F V Ross, D L Adams, and C M Gilbert. "Effect of online literature searching on length of stay and patient care costs." *Academic Medicine: Journal of the Association of American Medical Colleges* 69, no. 6 (June 1994): 489-95.
52. Westbrook, Johanna I, Enrico W Coiera, and A Sophie Gosling. "Do online information retrieval systems help experienced clinicians answer clinical questions?" *Journal of the American Medical Informatics Association: JAMIA* 12, no. 3: 315-21.
53. Ely, John W, Jerome A Osheroff, Mark H Ebell, M Lee Chambliss, Daniel C Vinson, James J Stevermer, et al. "Obstacles to answering doctors' questions about patient care with evidence: qualitative study." *BMJ (Clinical Research Ed.)* 324, no. 7339 (March 23, 2002): 710.
54. *NLM. "Data, News and Update Information." Announcements, News. http://www.nlm.nih.gov/ bsd/revup/revup_pub.html#med_update. Last Accessed September 11, 2008.*
55. NLM. "Number of Titles Currently Indexed for Index Medicus® and MEDLINE® on PubMed®." Procedures. *http://www.nlm.nih.gov/bsd/num_titles.html*. Last Accessed September 11, 2008.
56. "Board of Regents Minutes – February 2008." NLM Board of Regents. *http://www.nlm.nih.gov/ od/bor/2-08bor.pdf*. Last Accessed September 11, 2008.
57. "Board of Regents Minutes – September 2007." NLM Board of Regents. *http:// www.nlm.nih.gov/od/bor/9-07bor.pdf*. Last Accessed September 11, 2008.
58. Grimes, David A., Melody Y. Hou, Laureen M. Lopez, and Kavita Nanda. "Do Clinical Experts Rely on the Cochrane Library?." *Obstet Gynecol* 111, no. 2 (February 1, 2008): 420-422.
59. Bonis, Peter A., Gary T. Pickens, David M. Rind, and David A. Foster. "Association of a clinical knowledge support system with improved patient safety, reduced complications and shorter length of stay among Medicare beneficiaries in acute care hospitals in the United States." *International Journal of Medical Informatics* In Press (2008), Corrected Proof.
60. NLM. "2008AA Section 3 UMLS SEMANTIC NETWORK." Technical Documentation. *http://www.nlm.nih.gov/research/umls/meta3.html*. Last Accessed September 11, 2008.
61. "UpToDate Inc." UpToDate. *http://www.uptodate.com/home/about/index.html*. Last Accessed September 11, 2008.
62. "Determining Validity & Level of Evidence." DynaMed. *http://www.ebscohost.com/dynamed/ levels.php*. Last Accessed September 11, 2008.
63. "Content Sources." DynaMed. *http://www.ebscohost.com/dynamed/sources.php*. Last

Accessed September 11, 2008.

64. "Asylum-Seeker Rejected Based On Wikipedia, Appeals Court Reverts." Wired. http://blog.wired.com/27bstroke6/2008/09/asylum-seeker-r.html. Accessed September 11, 2008.

65. "BioMed Central | The Open Access Publisher." BioMed Central. http://www.biomedcentral.com/. Accessed September 11, 2008.

66. "Public Library of Science." PLoS. http://www.plos.org/. Accessed September 11, 2008.

67. "Wiley InterScience: Reference Work: The Cochrane Library 2008, Issue 3." The Cochrane Library. http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/ProductDescriptions.html?CRETRY=1&SRETRY=0. Accessed September 11, 2008.

68. "The Cochrane Collaboration - Cochrane entities." Cochrane Collaboration. http://www.cochrane.org/contact/entities.htm#CRGLIST. Accessed September 11, 2008.

69. "Cochrane centres and branches." Cochrane Collaboration. http://www.cochrane.org/contact/centresmap_new0507.htm. Accessed September 11, 2008.

70. "OneLook Dictionary Search." OneLook. http://www.onelook.com/about.shtml. Accessed September 11, 2008.

71. Ely, J W, J A Osheroff, M H Ebell, G R Bergus, B T Levy, M L Chambliss, et al. "Analysis of questions asked by family doctors regarding patient care." *BMJ (Clinical Research Ed.) 319*, no. 7206 (August 7, 1999): 358-61.

72. Dawes, Martin, Pierre Pluye, Laura Shea, Roland Grad, Arlene Greenberg, and Jian-Yun Nie. "The identification of clinically important elements within medical journal abstracts: Patient-Population-Problem, Exposure-Intervention, Comparison, Outcome, Duration and Results (PECODR)." *Informatics in Primary Care 15* (January 2007): 9-16.

73. Schlosser, Ralf W., Rajinder Koul, and John Costello. "Asking well-built questions for evidence-based practice in augmentative and alternative communication." *Journal of Communication Disorders 40*, no. 3 (2007): 225-238.

74. Meats, Emma, Jon Brassey, Carl Heneghan, and Paul Glasziou. "Using the Turning Research Into Practice (TRIP) database: how do clinicians really search?" *Journal of the Medical Library Association 95*, no. 2 (April 2007). http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1852632.

75. "About The Cochrane Collaboration." The Cochrane Collaboration. http://www.cochrane.org/ docs/descrip.htm. Accessed May 18, 2009.

76. "Docs look to Wikipedia for condition info: Manhattan Research." Medical Marketing & Media, http://www.mmm-online.com/Docs-look-to-Wikipedia-for-condition-info-Manhattan-Research/article/131038/. Accessed September 9, 2009.

77. Grossman, Lisa. “Should you trust health advice from the web?” New Scientist 2718, July 9009: 20-21. http://www.newscientist.com/article/mg20327185.500-should-you-trust-health-advice-from-the-web.html. Accessed September 9, 2009.

78. Princeton University. “WordNet - About WordNet”, http://wordnet.princeton.edu/wordnet/. Accessed September 9, 2009.

79. Csomai, Andras. “WordNet Bibliography” http://lit.csci.unt.edu/%7Ewordnet/. Accessed September 9, 2009.

80. The Global WordNet Association. “The Global WordNet Association.” http://www.globalwordnet.org/. Accessed September 9, 2009.

81. NLM. “MEDLINE Key Indicators”, http://www.nlm.nih.gov/bsd/bsd_key.html. Accessed September 9, 2009.

82. Wikipedia. “PubMed” http://en.wikipedia.org/wiki/PubMed. Accessed September 9, 2009.

83. NLM. “The *Clinical Questions* Collection”, http://clinques.nlm.nih.gov/JitSearch.html. Last Accessed September 9, 2009.

10. Appendix A -- The MedicinfoSys Architecture Diagram.