Automatically Building a Repository to Support Evidence Based Practice

Dina Demner-Fushman, Joanna L. Karpinski, George R. Thoma
Lister Hill National Center for Biomedical Communications (LHNCBC)
U.S. National Library of Medicine, National Institutes of Health, Bethesda, MD 20894
{ddemner|karpinskij|gthoma}@mail.nih.gov

Abstract

Our long-term goal is to find, store, update, and provide access to key facts needed to support clinical decision making. Presently, the facts are extracted automatically from clinical narrative and biomedical literature sources, primarily MEDLINE, and stored in the Repository for Informed Decision Making. We envision expert community validation of the extracted facts and peer-reviewed direct deposit of key facts in the future. The key facts reported in publications do not change and can be extracted in advance and retrieved as needed. We chose an alternative approach to building the repository: extraction of key facts for specific clinical tasks and clinical scenarios. Combined with providing extracted facts (linked to the original publications) as service at the point of care, this approach allows capturing clinicians' relevance judgments and leads to gradual, weakly-supervised construction of a collection of facts and documents pertaining to specific clinical situations and expert judgments of relevance and quality of the documents. In this paper, we demonstrate our approach to corpus construction using the clinical task of patient care plan development. We provide an overview of the process and focus on the automatic construction of PubMed queries – an essential step in finding documents containing key facts.

1. Introduction

Collections containing biomedical documents, key facts extracted from the documents, and judgments on relevance and value of the documents to specific clinical tasks and questions are essential for many reasons, including clinical decision support and further development of biomedical natural language processing methods. Our research into methods for building collections of key clinical facts relatively fast and at low cost is motivated by the well-known desire of clinicians to have research evidence provided in the form of bottom-line advice (Ely et al., 2005) on the one hand, and significant manual efforts presently needed to find and summarize key facts in the biomedical domain, on the other hand. For example, creation of the 2007 Text REtrieval Conference (TREC) Genomics track collection involved extensive interviewing of biologists to obtain real-life questions of interest to biological domain, as well as recruiting judges with significant domain knowledge, typically in the form of a PhD in a life science (Roberts et al., 2009).

We propose inferring clinical questions using formal representation of a patient's case, rather than actively soliciting information needs. The second step of our repository building process is fairly typical for building collections and consists of literature retrieval. Our retrieval process is complicated by the fact that rather than having a short description of information need provided by a user or a relevant document (which a search engine could use to find similar documents) we are presented with the description of the patient's status and need to find documents relevant both to the patient's status and the clinical task to be performed. We describe our approach to automatic construction of expert queries in Section 3 and the evaluation of the method in Section 4. The rest of the paper is organized as follows: in Section 2 we introduce the clinical task of patient care plan development and the framework for formal representation of a patient's case. Section 5 presents the structure of our repository and the mechanism for obtaining relevance judgments as part of the healthcare workflow. We conclude with a discussion of the preliminary results of our approach to building the repository.

2. Patient Care Plan Development

Mobility Problem: limited mobility due to huge mass right arm.
Skin Problem: st 3 sacral decub
GI Problem: nausea/vomiting-new onset
Respiratory Problem: emphysema /smoking cessation
Psychosocial Problem: depression
Neurological/Cognition Problem: Declining cognitive function
Pain Problem: pain rigth arm
Pain Problem: R arm tumor pain
Pain Problem: patient c/o intermittent pain to surgical site
Pain Goals: pt able to do ADL with minimal pain
Pain Goals: Pt able to rate pain <3/10.
Pain Goals: pt with pca hydromorphone. cont. dose. also reciving bupivicaine 0.25% via epineural
Pain Interventions: cont with pca dosing, prn hydromorphone also to be given for break through pain
Respiratory Interventions: pt moved to ICU early AM for persistent dyspnea unknown etiology includ wheezing, tachypnea despite ok saturations and adequate pain control: plan for further diagnostics

Figure 1: Semi-structured interdisciplinary team note. The format combines problem types restricted to controlled vocabulary (shown in bold) and free text description of the problems entered by the team members

Care plan development starts with assessment of a patient's status[1]. The assessment results are documented

[1] The assessment part of the note entered into a patient's chart is sometimes preceded by patient's description of the patient's current condition (mostly in narrative form) and registration of objective conditions, such as vital signs, patient's status observed by the clinician during examination, results of laboratory tests, and other observations.

in the care plan as descriptions of the patient's problems. The form of the description ranges from a list of problems selected from a controlled vocabulary to a narrative summary of the problems. Once the problems are established, the clinician reviews each problem and establishes goals to be achieved while addressing the problems, and plans interventions to achieve the goals. An example of a de-identified note derived from the interdisciplinary notes used to build our collection is shown in Figure 1. Ideally, a clinician would seek evidence support for all three steps of care plan development.

The elements of the care plan are in essence the same elements that are used in the evidence based practice (EBP) framework for finding information to ensure the best possible care in a given clinical situation. The elements of the framework, called PICO, are: the description of the patient and problem, intended interventions and comparisons, and desirable outcomes (Richardson et al., 1995). Clearly, these are the problems, interventions, and goals sections of the care plan and we, therefore, can use the framework developed within EBP for construction of well-formed clinical questions to formally represent the clinical situation. We use an existing EBP-based question-answering system, CQA, (Demner-Fushman and Lin, 2007) to find and extract key facts from publications relevant to the patient's case. The CQA system showed good performance answering clinical questions when queries and clinical scenario frames were developed manually. Manual formulation of the query is not ideally suited for use in clinical setting because it interrupts the workflow and is often perceived as less useful than spending the time with the patient (Bond, 2007).

To replace manual initiation of the search for key facts, we developed an automatic process for extracting information from a patient's record and properly formulating a query to identify appropriate evidence using the National Library of Medicine (NLM) resource, PubMed®. Our solution to automatic construction of clinical questions and initiation of the search process is described next.

3. Query Formulation Algorithm

To identify search strategies that would yield relevant results, we manually developed a set of reference PubMed search strings to analyze for text elements, query forms, and search processes that are most likely to yield successful search results. The set was developed by a medical librarian (the second author) using 254 records of patient encounters from 52 patients selected from a dataset of more than 4500 patient encounters. The 254 records were selected because the formal representations of the encounters using the PICO framework and simple searches (that combined all identified PICO elements) retrieved at least one MEDLINE® citation or other evidence (for example, a MedlinePlus® article). The reference queries retrieved the greatest proportion of relevant results, presented the most relevant results at the top of the results display, and retrieved a total number of hits that could be easily perused by a busy clinician in two minutes or less. These strings were constructed using the patient's primary diagnosis (*Chief Complaint*) and problems found in the interdisciplinary notes (*IDP Problem*).

3.1 PICO Representation of Patient Records

The *Problem* and *Intervention* extraction modules of the CQA system were used to represent patients' cases. Given a clinical note, the system automatically generates a question frame using one of the Named Entity Recognition (NER) tools (MetaMap (Aronson, 2001), NER modules of the NLM experimental search engine Essie (Ide et al., 2007) or CQA internal dictionary-based NER module) and a set of rules for extraction of the elements of a clinical scenario.

IDP Pain Problem: patient c/o intermittent pain to surgical site
IDP Pain Interventions: cont with pca dosing, prn hydromorphone also to be given for break through pain
NER (MetaMap Mappings): Intermittent pain [Sign or Symptom]; Surgical Site (Operative site) [Spatial Concept]; Hydromorphone [Organic Chemical,Pharmacologic Substance]; …; Pain [Sign or Symptom]
CQA Problem(s): Intermittent pain
CQA Patient:
CQA Intervention(s): Hydromorphone

Figure 2: Clinical question frame is used to formally represent a patient's note using NER

To generate question frames (see Figure 2 for an example), the CQA system extracts from the NER output concepts that belong to the following semantic groups: *Problems/findings* (meant to represent a patient's problem list), *Interventions*, and *Anatomy* (which provides details about the patient). The semantic groups are based on the Unified Medical Language System® (UMLS®) (Lindberg et al., 1993) Metathesaurus semantic types. The *Problems/findings* semantic group is based on the UMLS semantic group *Disorders* (McCray et al., 2001). The *Interventions* group includes therapeutic and diagnostic procedures, drugs, and drug delivery devices. The *Anatomy* group includes semantic types in the anatomy and physiology groups excluding those on the cell and molecular level (for example, *Cell* or *Molecular Function*).

3.2 Analysis of Reference Queries

We evaluated the manually constructed search string using SAS®[2] hypothesis testing (we used the SAS 9.0 SURVEYLOGISTIC Procedure with the Cumulative Logit logistic regression model and Fisher's Scoring optimization) and SPSS[3] linear and cubic regression analysis. The set of citations retrieved by each search

[2] http://www.sas.com/
[3] http://www.spss.com/

string was evaluated on the following criteria:

- Overall success of the query (evaluated by the second author on a scale of zero to four, 0=no hits, 4=ideal results set) as a function of the number of relevant results in the top 10, number of hits retrieved, and the positions of all relevant results
- Total number of citations retrieved
- Number of relevant citations in top 10
- Position of the first relevant citation
- Number of queries executed prior to success or termination
- Number of review article citations retrieved
- Total number of relevant review article citations in top 10
- Position of first relevant review article citation.

Five variables of query construction were evaluated for their effects on quality of search results:

1. Increased use of Medical subject headings (MeSH® terms) -- controlled vocabulary terms assigned to MEDLINE citations during NLM manual indexing process.
2. Varied use of the Chief Complaint
3. Increased use of advanced search strategies (use of subheadings when appropriate; identification of terms to search as major topic headings)
4. Use of complex query forms (use of Boolean AND/OR/NOT; addition of nested search strings)
5. Application of search limits to retrieve only review articles to reduce total number of results retrieved when the retrieved set is too large.

SAS hypothesis testing identified variable 1, increased use of MeSH terms, variable 3, increased use of advanced search strategies, and variable 4, use of complex query forms, as statistically significant ($p < 0.001$) to a successful search outcome. SPSS regression analysis was then performed on the number of MeSH terms in the query, revealing that searches that used between two and five MeSH terms were far more likely to be successful than searching using fewer than two or more than five MeSH terms. These factors were used to guide the development of the query formulation process.

3.3 Query Formulation Rules

Based on the experience gained during creation of the reference queries and the SAS and SPSS analysis results, the second author derived the following rules for the automatic query formulation:

1. Identify MeSH terms
2. Construct Chief Complaint String: If multiple MeSH terms are identified, combine terms with Boolean "AND".
3. Prior to constructing IDP Problem string
   - 3.1. Extract any subheading terms (Identify terms of an identical semantic type to PubMed subheadings and apply to Chief Complaint string (i.e. A drug name in the IDP Problem field would give the subheading "drug therapy" to the Chief Complaint string)
   - 3.2. Identify terms in the IDP Problem field that are explicitly subheadings. Apply those to the Chief Complaint string. (eg. Text phrase "surgery on Tuesday" would give the subheading "surgery" to the Chief Complaint terms.)
4. Construct IDP Problem string:
   - 4.1. Use dependency parser to identify relationships between MeSH terms
   - 4.2. Remove second-child terms
   - 4.3. Combine parent terms with Boolean "AND"
   - 4.4. Combine first-child terms with Boolean "OR" and nest this string.
5. Combine Chief Complaint and IDP Problem strings with Boolean "AND."

Chief Complaint: Diffused B Cell Lymphoma
IDP Problem Text: Maintain perfusion of right hand and fingers, preserve neural function right hand and wrist, track hematoma for status
Baseline Query: Diffused B Cell Lymphoma AND perfusion AND right hand AND fingers AND neural function AND wrist AND hematoma
Advanced Query: (Diffused B Cell Lymphoma) AND (Perfusion AND (Hand OR Fingers) AND (Nerves OR Wrist) AND hematoma)

Figure 3: Automatic query formulation strategies

6. Run iterative searches as necessary:
   - 6.1. If, when Chief Complaint and IDP Problem strings are combined and hits retrieved are less than or equal to 5, then combine any multiple Chief Complaint terms with Boolean "OR." Re-execute search.
   - 6.2. If retrieved set is less than or equal to three, then re-execute the search using only the IDP Problem string.
   - 6.3. If the results set retrieved from step (a) is between 20 and 100 hits, then search for all Chief Complaint terms as major topic headings.
   - 6.4. If retrieved set is still greater than or equal to 75, limit results display to review articles.

We evaluated the developed algorithm on 30 additional randomly-selected patient encounter records using a naïve ANDing of all identified PICO elements as the baseline. Figure 3 demonstrates the differences in the advanced and baseline query formulation.

4. Experimental Evaluation of the Query Formulation Algorithm

We evaluated citation sets retrieved by the baseline and advanced search strategies using the evaluation criteria for the reference searches (scale of zero to four; zero being the lowest, with no results retrieved, and four being the highest, with the greatest overall relevancy and usability of the results). Table 1 presents the results of this evaluation.

Relevance Ranking	Baseline	Advanced search
0 (no results retrieved)	4	0
1,2 (not relevant)	26	20
3,4 (relevant)	0	10

Table 1: Comparison of results retrieved using the baseline and advanced search strategies.

Assuming that a search query that retrieved results receiving a 0, 1, or 2 relevance score were not likely to be useful to a clinician, and that only those receiving a 3 or 4 should be considered successful retrievals that would be useful for patient care plan development, none of the results retrieved by the baseline strategy would thus be considered useful for care plan development, compared with 33.3% (10) of the results retrieved with the updated algorithm.

Result sets retrieved for all 30 queries were also evaluated for the placement of the first relevant result. Of the ten sets derived with the baseline algorithm that contained relevant citations in the top ten results, the mean location was 5.5; median location was 6. Of the 30 sets derived with the advanced search algorithm, mean location of the first relevant result was 2.4, with a median of 2. We observed that placement of the first relevant result was elevated by a mean of four citations when retrieved using the advanced search algorithm.

The overall number of relevant results in the top ten citations retrieved increased from a mean of 0.4 relevant citations in the top ten using the baseline algorithm to a mean of 2.3 relevant citations in the top ten using the advanced search algorithm.

We attribute better performance of the advanced search to reduction of the number of search terms and establishing better relations between the terms due to dependency parsing and rules (as opposed to ANDing all terms found in the note.)

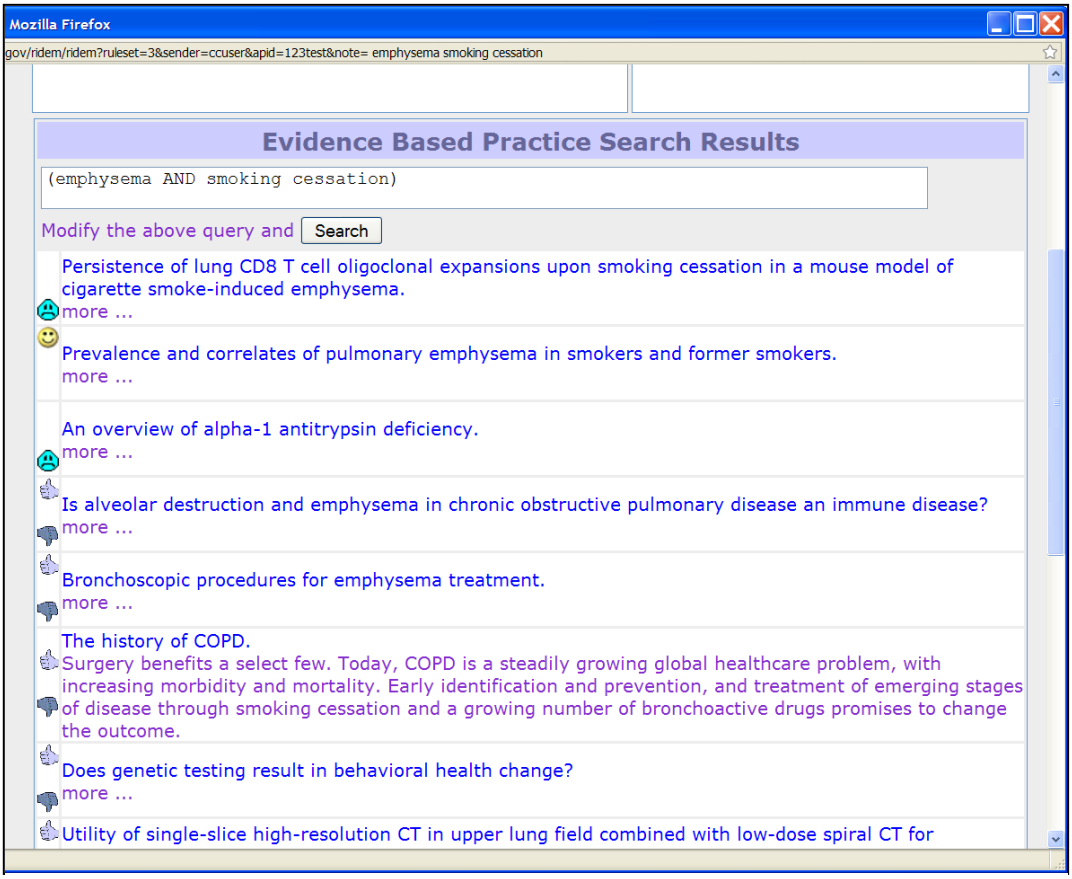

Figure 4: Obtaining expert judgments at the point of care

5. Obtaining Relevance Judgments at the Point of Care

Obtaining relevance judgment by the intended consumers of evidence at the point of care is a non-trivial task. Ideally, it has to involve minimal effort and be perceived as part of the workflow. We hope to achieve this goal by providing a service that delivers key facts extracted by the above described tools directly to an electronic patient record (EHR).

The next section provides an overview of the system that delivers evidence to an EHR and, at the same time, provides information to the system that automatically builds the repository for informed decision making.

5.1 System for Evidence Based Practice Support

Delivery of evidence to the point of care starts when a

clinician requests evidence[4]. The EHR generates a request to our service. The request sent to our service contains the Chief Complaint and the de-identified patient's note (similar to the one shown in Figure 1). Our system then extracts the PICO elements and MeSH terms found in the note using the CQA modules described in Section 3.1; constructs the query following rules described in Section 3; searches MEDLINE; and responds with an overview of the retrieved evidence to be displayed in the EHR. Figure 4 shows the part of the information dashboard delivered to the EHR that contains the overview of evidence. An overview of the evidence delivery system is provided in (Demner-Fushman et al., 2008).

The evidence provided to clinicians in the information dashboard consists of the titles of MEDLINE citations (linked to the citation and full text paper, if available) and the summary of key facts extracted from the citation using the CQA system.

Logging of the clinicians' navigation of the dashboard and their judgments is described next.

5.2 Capturing Expert Judgments

The key facts are displayed under the title of a MEDLINE citation on demand, The "thumbs up" and "thumbs down" icons to the left of each article allow for a quick one-click judgment. When a clinician clicks one of the thumbs icons, the smiley- or sad-face icons are displayed to illustrate judgment results. At the same time, judgments linked to the citation PubMed unique identifier and the unique identifiers of the clinical scenario are stored in the repository, maintained as MySQL database. The system also registers if the judgment is based on viewing the title and the summary alone, or after following the link to the full citation.

5.3 Preliminary Results

The system is under evaluation through delivering evidence to an EHR at a major clinical center since August 2009. Although viewing evidence is the third popular activity (after accessing information about drugs and the specifics of the patient's case), the absolute number of viewed citations is small (the 350 followed links constitute less than 0.02% of all interactions with the information dashboard). At the same time, we obtained expert relevance judgments for 267 citations (116 positive and 151 negative). The relatively large proportion of papers with judgments compared to the total number of viewed papers is not surprising. The primary goal of judging is to improve other interdisciplinary team members experience when looking for evidence support: papers judged negatively for a given scenario are lowered in rank (based on the cumulative judgment scores), whereas papers judged positively are promoted in rank at the subsequent evidence deliveries.

Recently, clinicians requested that the system allows modifying the automatically constructed query and repeating the search. The search box shown in the top part of Figure 4 is about to be deployed to the EHR. We are looking forward to augmenting our collection with manually corrected searches.

6. Related Work

To the best of our knowledge, the proposed approach to building a collection is new. The mechanism for capturing experts' judgments is related to collaborative filtering widely used in commercial systems to predict a user's interest by collecting information about many users' taste in music, books, etc. (Goldberg et al., 1992), and adaptive information retrieval (Jose et al., 2008).

Our capturing of relevance judgments provided to improve colleagues experience is related to secondary use of biomedical literature, such as using inclusion or citation of a paper by the American College of Physicians (ACP) Journal Club as an indication of the paper's high quality and relevance to a specific clinical task (Aphinyanaphongs et al., 2005); use of MeSH heading indexing as the reference standard in information extraction task (Aronson et al., 2008); and use of journal descriptors for word sense disambiguation (Humphrey et al., 2005).

We also believe to have developed a new approach to automatic query creation. Several approaches have been previously taken to the process of automating query construction using contextual information while filtering for the best quality information.

Of these systems, many retrieve relevant information by relying on the categorization of information within the EHR to automatically identify terms and use those to populate the search fields of an appropriate evidence resource. KnowlegeLink system of drug information retrieval provides users with a search button within the EHR record where medication names appear (Maviglia et al., 2006). The system identifies drug names using text parsing and automatically populates the URL of one of two pre-specified drug databases (MicroMedEx or SkolarMD) with the name of the drug. Within the individual drug databases, users are able to select the type of information about the drug they wish to search for (therapy, adverse effects, etc.) Cimino's (2007) InfoButtons/InfoButton Manager system links the EHR to a repository of clinical questions and answers based on the user's selection of a limited number of different categories of information from the electronic record. A system designed by Rosenbloom et al. (2005) for use with Vanderbilt University's CPOE system similarly provided users with a method of automatically retrieving information relevant to patient care. Using terms derived from the active diagnosis and medication categories in the CPOE system, basic keyword searches could be constructed in PubMed at the point and moment of care.

[4] The approach to initiating the process largely depends on the EHR. In our current setting, the request is issued when a clinician clicks on the EBP tab of the EHR.

7. Conclusions

This paper presents an approach and an ongoing effort towards building a collection of clinical scenarios (that serve as topics of interest), combined with documents automatically retrieved to augment the scenarios with evidence, and expert judgments on the relevance of retrieved documents to the clinical scenario. The benefits of the proposed approach are in the relatively low cost and minimal supervision in construction of the collection, as well as obtaining expert judgments at the point of care.

Some of the drawbacks of the approach are in slow rate of obtaining judgments, sparseness of judgments and lack of reasons for judgments. The last two issues have been extensively studied in the context of TREC (Voorhees and Harman, 2005), but might need re-evaluation in the context of clinical applications. We plan to speed up the collection process through using the system in EBP educational sessions at the clinical center. During the sessions we hope not only to obtain more relevance judgments faster, but also to capture reasoning behind the judgments.

We are less concerned about the quality of obtained judgments – those are provided by members of tightly-knit teams for other members with the purpose of improving provided care, therefore we expect high-quality judgments.

8. Acknowledgements

This work was supported by the Intramural Research Program of the NIH, National Library of Medicine.

9. References

Aphinyanaphongs Y, Tsamardinos I, Statnikov A, Hardin D, Aliferis CF. (2005). Text Categorization Models for High-Quality Article Retrieval in Internal Medicine. *Journal of the American Medical Informatics Association*, 12(2), pp. 207--216.

Aronson AR. (2001) Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp*, pp. 17-21.

Aronson AR, Mork JG, Neveol A, Shooshan SE, Demner-Fushman D. (2008). Methodology for Creating UMLS Content Views Appropriate for Biomedical Natural Language Processing. In *Proceedings of the Annual Symposium of the American Medical Informatics Association (AMIA).* Washington, DC: AMIA, pp. 21--25.

Bond CS. (2007). Nurses and computers. An international perspective on nurses' requirements. *Medinfo*, 12(1), pp. 228—232.

Cimino JJ. (2007). An integrated approach to computer-based decision support at the point of care. *Transactions of the American Clinical and Climatological Association*, 118, pp. 273--288.

Demner-Fushman D, Lin J. (2007). Answering Clinical Questions with Knowledge-Based and Statistical Techniques. *Computational Linguistics*, 33(1) pp. 63--103.

Demner-Fushman D, Seckman C, Fisher C, Hauser SE, Clayton J, Thoma GR. A Prototype System to Support Evidence-based Practice. AMIA Annu Symp Proc. 2008 Nov 6:151-5.

Ely JW, Osheroff JA, Chambliss ML, Ebell MH, Rosenbaum ME. (2005) Answering physicians' clinical questions: obstacles and potential solutions. *Journal of the American Medical Informatics Association*, 12(2), pp. 217--224.

Goldberg D, Nichols D, Oki BM, Terry D. (1992) Using collaborative filtering to weave an information tapestry, *Communications of the ACM*, 35 (12), pp. 61—70

Humphrey SM, Rogers WJ, Kilicoglu H, Demner-Fushman D, Rindflesch TC. (2005). Word Sense Disambiguation by Selecting the Best Semantic Type Based on Journal Descriptor Indexing: Preliminary Experiment. *Journal of the American Society for Information Science and Technology*, 57(1), pp. 96--113.

Ide NC, Loane RF, Demner-Fushman D. Essie: a concept-based search engine for structured biomedical text. *J Am Med Inform Assoc*. 2007 May-Jun;14(3):253-63.

Jose JM, Joho H, van Rijsbergen CJ. (2008). Adaptive Information Retrieval, *Information Processing & Management*, 44(6), pp. 1819—1821

Lindberg DA, Humphreys BL, McCray AT. (1993) The Unified Medical Language System. *Methods of information in medicine*, 32(4), pp. 281--291.

Maviglia SM, Yoon CS, Bates DW, Kuperman G. (2006). KnowledgeLink: impact of context-sensitive information retrieval on clinicians' information needs. *Journal of the American Medical Informatics Association*, 13(1), pp. 67--73.

McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. *Stud Health Technol Inform*. 2001;84(Pt 1): 216–20.

Richardson W, Wilson MC, Nishikawa J, Hayward RSA. (1995). The well-built clinical question: a key to evidence-based decisions. *ACP Journal Club*, 123, pp. A-12.

Roberts PM, Cohen AM, Hersh WR. (2009). Tasks, topics and relevance judging for the TREC Genomics Track: five years of experience evaluating biomedical text information retrieval systems. *Information Retrieval*, 12(1), pp. 81 -- 97.

Rosenbloom ST, Geissbuhler AJ, Dupont WD, Giuse DA, Talbert DA, Tierney WM, Plummer WD, Stead WW, Miller RA. (2005). Effect of CPOE user interface design on user-initiated access to educational and patient information during clinical care. *Journal of the American Medical Informatics Association*, 12(4), PP. 458--473.

Voorhees EM, Harman DK. (2005). *TREC: Experiment and Evaluation in Information Retrieval.* Cambridge, MA: The MIT Press