Proceedings of the Workshop on Natural Language Processing in
the Biomedical Domain, Philadelphia, July 2002, pp. 69-76.
Association for Computational Linguistics.

Analyzing the Semantics of Patient Data to Rank Records of Literature Retrieval

Eneida A. Mendonça
Department of Medical Informatics
Columbia University
em264@columbia.edu

Stephen B. Johnson
Department of Medical Informatics
Columbia University
sbj2@columbia.edu

Yoon-Ho Seol
Department of Medical Informatics
Columbia University
seol@dmi.columbia.edu

James J. Cimino
Department of Medical Informatics
Columbia University
jjc7@columbia.edu

Abstract

We describe the use of clinical data present in the medical record to determine the relevance of research evidence from literature databases. We studied the effect of using automated knowledge approaches as compared to physician's selection of articles, when using a traditional information retrieval system. Three methods were evaluated. The first method identified terms and their semantics and relationships in the patient's record to build a map of the record, which was represented in conceptual graph notation. This approach was applied to data in an individual's medical record and used to score citations retrieved using a graph matching algorithm. The second method identified associations between terms in the medical record, assigning them semantic types and weights based on the co-occurrence of these associations in citations of biomedical literature. The method was applied to data in an individual's medical record and used to score citations. The last method combined the first two. The results showed that physicians agreed better with each other than with the automated methods. However, we found a significant positive relation between physicians' selection of abstracts and two of the methods. We believe the results encourage the use of clinical data to determine the relevance of medical literature to the care of individual patients.

1 Introduction

The practice of evidence-based medicine, which gained popularity in the last decade, has encouraged clinicians to understand and utilize critically appraised published research evidence. The tremendous increase of biomedical knowledge resources in electronic form, particularly on the World Wide Web, has generated a great deal of interest. The increased availability of information does not make it easy for clinicians to filter large amounts of information and incorporate evidence to clinical practice. Although the number of clinicians and medical students who routinely perform their own searches has increased, they still have difficulty keeping-up-to-date with advances in medical science. (Gorman and Helfand, 1995)

Decision support tools designed to provide relevant and current evidence to clinicians promise to substantially improve health care quality (Haynes, Hayward, and Lomas, 1995 ;Rodrigues, 2000 ;Sim, et al., 2001) and potentially reduce medical errors.(Bates, et al., 2001) Such tools include those that facilitate the access to, extraction of, and summarization of evidence. The Evidence and Decision Support track of the 2000 AMIA Spring Symposium examined the challenges in the development and adoption of clinical decision support systems for evidence-based practice.(Sim, et al., 2001) The speakers for the Evidence and Decision Support track described five central areas of activity as essential for the adoption of those systems. Two of the areas were a) the capture of both literature-based and practice based research evidence into machine-interpretable form, and b) the establishment of a technical and methodological foundation for applying research evidence to individual patients at the point of care.

Our goal is to improve the way retrieved medical literature is presented by identifying critical information in the individual medical record that is useful for determining the relevance of literature data, also called research evidence. We describe an automated knowledge based approach that uses case-specific evidence present in patient's medical record to rank research evidence from literature databases.

2 Background

The integration of information with clinical applications may facilitate the access to scientific evidence, clinical guidelines, and other decision tools, in a way that information retrieved from these sources is personalized based on the context of individual needs.(Cimino, 1996) One of many challenges in building such systems is to understand what information in the individual medical record is important to the user and therefore potentially useful in search, retrieval, summarization, and presentation processes. Identifying the important terms, their semantic types, and common relationships maybe an interesting solution to the problem. The approach we describe here is based on previous research on automated methods to extract information from medical literature, and the use of natural language processing techniques to analyze free text clinical reports. Natural language processing techniques have been used to analyze free text reports in order to provide data for applications, such as automated encoding, decision support, patient management, quality assurance, outcomes analysis, and clinical research.(Baud, et al., 1995 ;Fiszman, et al., 2000 ;Friedman, et al., 1994 ;Friedman, et al., 1999 ;Gundersen, et al., 1996 ;Sager, et al., 1995) Data mining and knowledge discovery techniques have been used to interpret data from natural language processing output of narrative reports.(Wilcox and Hripcsak, 2000)

2.1 Automated extraction from medical literature

Research studies have introduced approaches to facilitate knowledge extraction from MEDLINE (Cimino and Barnett, 1993 ;Mendonça and Cimino, 2000) and the Unified Medical Language System (UMLS).(Zeng and Cimino, 1998) MEDLINE is the National Library of Medicine (NLM) premier bibliographic database covering the fields of medicine, nursing, dentistry, veterinarian medicine, the health care system, and the preclinial sciences. MEDLINE contains bibliographic citations and author abstracts from more than 4,600 biomedical journals published in the United States and 70 other countries. MEDLINE citations are indexed with Medical Subject Headings (MeSH) terms. MeSH (1999) is the NLM's controlled vocabulary used specifically for medical bibliographic indexing. Terms from MeSH are manually assigned to each document. The UMLS project was initiated in the mid-1980s by the National Library of Medicine.(Humphreys and Lindberg, 1993) The main goal was to provide a mechanism for linking diverse medical vocabularies as well as sources of information. There are currently three components of the UMLS Knowledge Sources: the Metathesaurus, Semantic Network, and SPECIALIST Lexicon.

We based our method on the approach described by Mendonça and Cimino. The researchers described an automated knowledge extraction method from MEDLINE citations, based on the ideas introduced by Zeng and Cimino (Zeng and Cimino, 1998), using the search strategies by Haynes and colleagues.(Haynes, et al., 1994) The approach involved the use of hierarchical and semantic links in the Medical Entities Dictionary (MED)(Cimino, et al., 1994) to identify additional terms which could be used to build specific patient-oriented queries. The MED uses a frame-based semantic network that includes a classification hierarchy to represent medical concepts and the relationship among them. The authors identified semantic associations in literature citations of four basic clinical tasks: etiology, prognosis, diagnosis, and therapy. These associations were based on the co-occurrence of MeSH terms in 4,000 MEDLINE citations.

The results of the study showed that only 7 to 8% of the semantic pairs generated in each task group differ significantly from random chance. A pilot study to assess the clinical validity of the associations showed a relative good specificity and sensitivity for their intended purpose, information retrieval, except in one group(prognosis). Performance was especially good in the therapy group.

Generic representation of a microbiology laboratory culture and sensitivity test:
[LPRO] -> (AE) -> [ANTB] -> (DS) -> [PFUN] -> (PO) -> [ORGM] <- (PP) <- [OATT]

Interpretation: A procedure assesses the effect of an antibiotic which disrupts a physiologic function which is a process of an organism which has an attribute (sensitive/resistant).

Example: Culture & Smear Site
Specimen description: catheterized urine
Culture: > 100K col/ml E coli.
Organism: E. coli
Method: Microscan MIC.
Sensitivity test: Ampicillin 2S, Sulfamethoxazole R, Cephalexin 2S (partial result)

Conceptual graph representaion of the test:
[LPRO: Culture & Smear Site] -> (AE) -> [ANTB: Ampicillin] -> (DS) -> [PFUN: undefined] -> (PO) -> [ORGM: E. coli] <- (PP) <- [OATT: sensitive]
[LPRO: Culture & Smear Site] -> (AE) -> [ANTB: Sulfamethoxazole] -> (DS) -> [PFUN: undefined] -> (PO) -> [ORGM: E. coli] <- (PP) <- [OATT: resistant]
[LPRO: Culture & Smear Site] -> (AE) -> [ANTB: Cephalexin] -> (DS) -> [PFUN: undefined] -> (PO) -> [ORGM: E. coli] <- (PP) <- [OATT: sensitive]

Figure 1. Conceptual representation of a culture and sensitivity test

3 Research Question

The work we describe here focused on the clinical data present in patients' medical records, and the use of these data to determine the relevance of research evidence. The main research question was “What is the effect of using the automated knowledge based approach compared to a physician’s selection of articles when using a traditional information retrieval system?”

4 Methods

We evaluated the application of semantic algorithms to data in an electronic medical record for sorting abstracts of articles (citations) retrieved from medical literature databases.

4.1 Semantic Approaches

Data from an individual’s medical record was retrieved from the clinical repository using the latest entry of each laboratory test and narrative reports, if within one month from the retrieval data, to create a “map” or summary of the medical record. Discharge summaries were an exception to this rule. The latest discharge summary was always retrieved independently of the time constraints.

The selected narrative reports were parsed by AQUA - A QUery Analyzer,(Johnson, et al., 1993) a natural language parser that translates text into a standard notation: conceptual graphs. (Sowa, 1984) AQUA’s lexicon is based on the UMLS Metathesaurus. The UMLS Semantic Net recommends which concepts and relations can be sensibly combined.

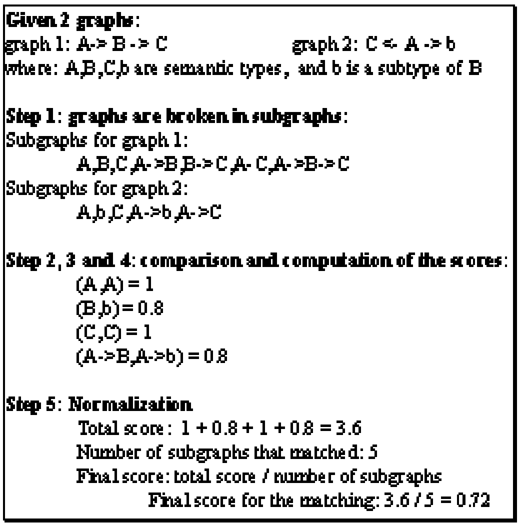
Coded data (e.g., laboratory tests) were also represented as conceptual graphs. We used the MED to infer knowledge when appropriate. For instance, when a glucose measure of 150 mg/dl was retrieved, the information in the MED allowed us to infer that the result could also be interpreted as hyperglycemia. The MED was also used to map concepts in the electronic medical record to UMLS concepts in order to obtain their semantic types. Figure 1 shows an example of a test result extracted from the medical record and its conceptual graph representation.

Three semantic algorithms are used. The first algorithm is based on graph matching techniques. The second method identifies associations between terms in the medical record, assigning them semantic types and weights based on the co-occurrence of these associations in citations of biomedical literature. The method is applied to data in an individual’s medical record, and scored citations according to this information. The last method combines the first two.

The graph matching algorithm is based on assumption that the similarity of two representations is a function of the amount of information they share. (Maher, 1993 ;Poole and Campbell, 1995) It worked as follows:

1. graphs on both sides (clinical data and citations) are broken into subgraphs;
2. subgraphs of clinical data are then compared to subgraphs of the citations;
3. if a perfect match is found (semantic type and relationship) a score of 1 is given. If not, points are reduced for each type of relation that did not match. Points are reduced based on the UMLS semantic types and relationship hierarchy (UMLS Semantic Net);
4. indirect matches are searched;
5. the score is then normalized based on the number of subgraphs generated by each graph, and the number of graphs in the document.

Figure 2 shows how the similarity between two graphs is computed.


Given 2 graphs:
graph 1: A-> B -> C   graph 2: C <- A -> b
where: A,B,C,b are semantic types, and b is a subtype of B

Step 1: graphs are broken in subgraphs:
Subgraphs for graph 1:
A,B,C,A->B,B->C,A- C,A->B->C
Subgraphs for graph 2:
A,b,C,A->b,A->C

Step 2, 3 and 4: comparison and computation of the scores:
(A,A) = 1
(B,b)= 0.8
(C,C) = 1
(A->B,A->b) = 0.8

Step 5: Normalization
Total score: 1 + 0.8 + 1 + 0.8 = 3.6
Number of subgraphs that matched: 5
Final score: total score / number of subgraphs
Final score for the matching: 3.6 / 5 = 0.72

Figure 2. Simplified graph matching representation

The second method studied is based on the semantic associations between concepts in the medical record. A knowledge base containing the statistically significant semantic type associations found in MEDLINE by Mendonça and Cimino was built. In addition to the semantic types, the knowledge base also stores the number of times the association occurred in the citations, the MeSH terms that originated the association, and the *P* values generated by the significance test. The knowledge base contains three groups of associations: therapy, etiology and diagnosis. The associations are grouped based on the type of questions the citations were retrieved to answer. In this method, we identify all possible associations between semantic types in the medical record. Semantic relationships are not taken in consideration. If the same associations are found in the citations retrieved, we consider it a match. Only the associations present in the knowledge base are weighted. The weights for each citation depend on the type of question that originated the citation.

The algorithm may be best understood through an example. Assume a clinician sees Mr. Ocean, and has a question about how to treat Mr. Ocean's migraine. The clinician searches the literature and finds two citations, one published in the Annals of Internal Medicine and the second, in the New England Journal of Medicine. In the semantic approach described, if a pair of semantic types is found in Mr. Ocean's medical record (e.g., Disease or Syndrome – Pharmaceutical Substance) and also in the citations retrieved, and the association is present in the knowledge base for questions on therapy, then that association receives a certain weight. The association weights are based on the co-occurrence of these associations in citations of biomedical literature. Two values are used in the scoring process: a) number of associations that are present in the medical record and citation, b) the logarithm of the sum of the inverse of *P* values of each association found.

The third semantic algorithm combines features from the previous two. For each association that matches the medical record, 0.1 point is added to the graph matching score for that citation.

4.2 Evaluation studies

We performed a study in order to assess the effect of using the automated knowledge approach compared to a physicians' selection of articles when using traditional information retrieval systems.

Three patients consented to the use of anonymized versions of the data stored in their electronic medical records. We randomly selected one admission of each patient to build the clinical cases. Data from these individuals' medical records were retrieved from the clinical repository as previously described. Narrative reports were parsed differently depending on the algorithm in evaluation. The "maps" of the three medical records were created. For each case, four clinical questions were selected from a database of generic questions based on the work of Ely and collaborators.(Ely, et al., 2000) Nonclinical questions (e.g., What are the administrative rules/considerations in <situation y?>) were eliminated from the database before the selection. Each question selected was also eliminated before the next random selection, so that we had a total of 12 unique questions. A health science librarian generated the search strategy for each question based on the case description. Two information retrieval systems were searched: PubMED (clinical queries using research methodology filters based largely on the work of Haynes and colleagues) (Haynes, et al., 1994) and OVID (Evidence-Based Medicine Reviews)[1]. All search strategies were keyword based with Boolean connectors. The search was time limited (last 3 years). In the cases where no citation was retrieved, the time limit was removed. The time limit was imposed because the time required by an expert to analyze all citations retrieved without this limitation would have been a disincentive to their participation in the study.

Subjects were recruited as follows. Three board-certified internists, one board-certified family physician, and one research physician were selected as experts. Four of the five physicians actively practice medicine in their fields. Participants were given instructions and received the following materials: a) cases' description, b) clinical questions selected for each case, and c) citations retrieved to answer each question. Case descriptions were based on the admission note (chief complaint, history of present illness, past medical and surgery history, and current medications), and the results of laboratory tests performed during the admission. Subjects were asked to score each citation according to the relevance of the article (citation) to the question asked and to the patient the case referred to. We asked each to define a relevant citation as providing information that could be used in the care of that particular patient.

The score used by the physicians was:

1 – completely nonrelevant
2 – almost entirely nonrelevant
3 – somewhat relevant
4 – very relevant
5 – completely relevant

Each participant analyzed all questions.

The automated methods also scored each citation. The scores were based on how well the abstract and title in the citation matched the case's summary. The computer scores are described in the previous section. We used the inverse chronological order in which the citations were provided by their respective programs as an additional method for comparison (control).

The main outcome measure in my study was the distance of averaged correlation coefficients between subjects and the average of the raters. For each physician, we calculated the average distance from the average of the other 4 physicians, and for each automated method, we calculated the average distance from the average of all 5 physicians. The null hypotheses were: a) that each subject was no more distant from the average of the physicians than the physicians were from each other and b) that there was no correlation between the average of the physicians' scores and the average of the subjects' scores. We used bootstrapping to estimate variance directly from the data.

We used Pearson's product-moment correlation to calculate the strength of the association between subjects and the average of the raters. In order to accommodate the fact that questions had a different number of citations associated with them, we calculated a weighted average r_ of correlation coefficients r_i given weights w_i as follows:

$$r\_ = r_i * (w_i / \Sigma(w_i))$$

$$w_i = (n_i - 1)^{½}$$

where n is the number of citations retrieved in question *i*.

5 Results

The 3 clinical cases and 12 questions generated a set of 219 citations: 111 from PubMED and 108 from EBM reviews. The number of citations per question varied from 1 to 28. The four questions that retrieved only one citation were removed from the statistical analysis. Thus, the total number of citations analyzed was 215.

[1] EBM Reviews includes the following databases : ACP Journal Club (ACP), Cochrane Database of Systematic Reviews (COCH), and Database of Abstracts of Reviews of Effectiveness (DARE)

The correlation coefficient between subjects and the average of raters varied from -0.07 to 0.52. The weighted correlation coefficient for each subject is listed in Table 1. A significant positive correlation was found between the average of physicians' scores and the scores given by the graph matching and the combined algorithms.

The main outcome measure, the difference between subject correlations minus average physician correlations, is shown in Table 2. Positive numbers imply worse performance (more unlike the average physician). No physicians differed significantly from other physicians. The automated methods did differ from physicians with significant *P* values.

Table 1. Correlation coefficients and significance of the correlation

Subject	Correlation	*P* Value
Physician 1	0.46	< 0.0001
Physician 2	0.44	< 0.0001
Physician 3	0.52	< 0.0001
Physician 4	0.52	< 0.0001
Physician 5	0.48	< 0.0001
Graph matching	0.19	0.0098
Graph matching + associations	0.15	0.046
Number of associations	-0.07	> 0.05
Associations value	-0.03	> 0.05
Inverse chronological order	0.04	> 0.05

Table 2. Average subject correlations minus average physician correlations

Subject	Difference (95% CI)	*P* Value
Physician 1	-0.03 (-0.08 to 0.14)	0.60
Physician 2	-0.05 (-.08 to 0.18)	0.43
Physician 3	0.04 (-0.06 to 0.14)	0.41
Physician 4	0.05 (-0.07 to 0.17)	0.40
Physician 5	-0.01 (-0.11 to 0.13)	0.86
Graph matching	0.29 (0.24 to 0.54)	0.0002
Graph matching + associations	0.33 (0.29 to 0.58)	< 0.0002
Inverse chronological order	0.44 (0.32 to 0.56)	< 0.0002

6 Discussion

Our main goal in this project was to assess the effect of the use of clinical data to improve presentation of medical literature. We evaluated three semantic methods.

The level of association between pairs of subjects ranged from -0.07 to 0.52. The level of association associations among physicians seemed to be similar to levels of agreement between 2 independent raters reported in the literature.(Wilczynski, McKibbon, and Haynes, 2001) No single physician stood out as significantly different from the others.

The graph matching algorithm highly correlated with physicians' average, although it did not perform as well as individual physicians. This finding encourages the use of clinical data to determine the relevance of medical literature to the care of individual patients. In an integrated system (medical record with information resources) this positive correlation suggests that our method can facilitate presentation of online biomedical literature. For instance, if the electronic medical record is integrated to an existent information retrieval, findings from an individual medical record can be used to rearrange the way retrieved information is presented; in a way that literature matching that individual's medical record will be presented first, rather than the usual presentation in reverse chronological order.

The combined method also correlated significantly with physicians' average, although its performance was not as good as of the simple graph matching. This result may be due to a negative effect of the associations in the knowledge over the matching. There was no correlation between the methods that use the co-occurrence of semantic types in medical literature citations and the average of physicians. The automated method based on the chronological order of articles did not correlate with physicians' average.

The poor results of the method which used the knowledge base of semantic co-occurrences in Medline citations may be due to several aspects. The terms used for indexing medical citations may not correspond well to data usually found in medical records. Approaches using the UMLS Semantic Net may be also somewhat limited by the fact approximately one fourth of the Metathesaurus concepts are assigned several semantic types, which makes it difficult to get a precise understanding of the co-occurrences.(Burgun and Bodenreider, 2001)

We believe enhancements can still be made. The graph matching algorithm is highly dependent on the output of the natural language processor. The general language processor used to parse both clinical data and citations was never validated for this use. AQUA was designed to translate user's natural language queries into a conceptual graph representation. It was developed on a corpus of clinical queries. Prior to this study, the parser was trained with only a few sentences from the medical literature. The complexity of the clinical data and medical literature involved in the study generated a significant number of "broken" graphs. The similarity found between the graphs was usually at the level of single nodes. It was also observed that the parser had difficult with very long sentences and sentences in the results section of the abstract. An example of a sentence partially parsed is "Furthermore, patients treated with aprotinin had significantly less total postoperative blood loss (718 +/- 340 ml vs 920 +/- 387 ml, p =0.04)". With enhancements to the natural language processor, we believe we could obtain a better representation of the data, and consequently more accurate results.

The use of UMLS Semantic Net may have also contributed to the elevated incidence of "broken" graphs. Mendonça and Cimino (Mendonça and Cimino, 2001) found that only 22.99% of the associations of semantic types based on MeSH terms retrieved from the medical literature had a direct semantic relationship in the UMLS Semantic Net. A careful appreciation of the missing relationships may help us to understand whether the addition of new semantic relationships can contribute to a better representation of clinical and literature data.

Whether improvements in the parser to allow it to handle medical literature and complex clinical data would improve the performance of the automated methods is unclear; further studies are needed. The use of this method in association with other information retrieval techniques is being investigated by the authors.

7 Conclusion

The goal of the study is to support the use of clinical data to facilitate the information retrieval of biomedical literature. The results of this study support this goal. The use of conceptual graph representation and graph matching techniques correlated significantly with the average of physicians when judging the relevance of citations to the care of an individual patient. Additional studies are needed in order to understand if this performance is acceptable in a clinical environment. A careful evaluation of the parsed reports and careful appreciation of the missing relationships may help us to understand the results and enhance the performance of the algorithms.

References

Medical Subject Headings - Annotated Alphabetical List. Bethesda, MD: 1999. (National Library of Medicine).

Bates DW, Cohen M, Leape LL, Overhage JM, Shabot MM, Sheridan T. Reducing the frequency of errors in medicine using information technology. Journal of the American Medical Informatics Association 2001; 8(4):299-308.

Baud RH, Rassinoux AM, Wagner JC *et al*. Representing clinical narratives using conceptual graphs. Methods of Information in Medicine 1995; 34(1-2):176-86.

Burgun A, Bodenreider O. Methods for Exploring the Semantics of the Relationships between Co- occurring UMLS Concepts. Medinfo 2001; 10(Pt 1):171-5.

Cimino JJ. Linking patient information systems to bibliographic resources. Methods of Information in Medicine 1996; 35(2):122-6.

Cimino JJ, Barnett GO. Automatic knowledge acquisition from MEDLINE. Methods of Information in Medicine 1993; 32(2):120-30.

Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. Journal of the American Medical Informatics Association 1994; 1(1):35-50.

Ely JW, Osheroff JA, Gorman PN *et al*. A taxonomy of generic clinical questions: classification study. British Medical Journal 2000; 321(7258):429-32.

Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. Journal of the American Medical Informatics Association 2000; 7(6):593-604.

Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. Journal of the American Medical Informatics Association 1994; 1(2):161-74.

Friedman C, Knirsch C, Shagina L, Hripcsak G. Automating a severity score guideline for community-acquired pneumonia employing medical language processing of discharge summaries. Proceedings of the AMIA Fall Symposium 1999; 256-60.

Gorman PN, Helfand M. Information seeking in primary care: how physicians choose which clinical questions to pursue and which to leave unanswered. Medical Decision Making 1995; 15(2):113-9.

Gundersen ML, Haug PJ, Pryor TA *et al*. Development and evaluation of a computerized admission diagnoses encoding system. Computers and Biomedical Research 1996; 29(5):351-72.

Haynes RB, Hayward RS, Lomas J. Bridges between health care research evidence and clinical practice. Journal of the American Medical Informatics Association 1995; 2(6):342-50.

Haynes RB, Wilczynski N, McKibbon KA, Walker CJ, Sinclair JC. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. Journal of the American Medical Association 1994; 1(6):447-58.

Humphreys BL, Lindberg DAB. The UMLS project: making the conceptual connection between users and the information they need. Bulletin of the Medical Library Association 1993; 81(2):170-7.

Johnson SB, Aguirre A, Peng P, Cimino J. Interpreting natural language queries using the UMLS. Proceedings of the Annual Symposium on Computer Applications in Medical Care 1993; 294-8.

Maher PE. A similarity measure for conceptual graphs. International Journal of Intelligent Systems 1993; 8:819-37.

Mendonça EA, Cimino JJ. Automated knowledge extraction from MEDLINE citations. Proceedings of the AMIA Fall Symposium 2000; (20 Suppl):575-9.

Mendonça EA, Cimino JJ. Content evaluation of a knowledge base. 2001; 974.

Poole J, Campbell JA. A novel algorithm for matching conceptual graphs and related graphs. Ellis G, Levinson R , Rich W, Sowa JF, edts. Conceptual Structures: Applications, Implementation and Theory, Third International Conference on Conceptual Structures, ICCS'95. Springer, 1995: 293-307.

Rodrigues RJ. Information systems: the key to evidence-based health practice. Bulletin of the World Health Organization 2000; 78(11):1344-51.

Sager N, Lyman M, Nhan NT, Tick LJ. Medical language processing: applications to patient data representation and automatic encoding. Methods of Information in Medicine 1995; 34(1-2):140-6.

Sim I, Gorman P, Greenes RA *et al*. Clinical decision support systems for the practice of evidence-based medicine. Journal of the American Medical Informatics Association 2001; 8(6):527-34.

Sowa JF. Conceptual structures: information processing in mind and machine. Reading, MA: Addison-Wesley, 1984.

Wilcox A, Hripcsak G. Medical text representations for inductive learning. Proceedings of the AMIA Fall Symposium 2000; 923-7.

Wilczynski NL, McKibbon KA, Haynes RB. Enhancing retrieval of best evidence for health care from bibliographic databases: calibration of the hand search of the literature. Medinfo 2001; 10(Pt 1):390-3.

Zeng Q, Cimino JJ. Automated knowledge extraction from the UMLS. Chute CG. Proceedings of the AMIA Fall Symposium. Philadelphia: Hanley & Belfus Inc., 1998: 568-72.