Contextual Models of Clinical Publications for Enhancing Retrieval from Full-Text Databases

Gretchen P. Purcell and Edward H. Shortliffe
Section on Medical Informatics
Medical School Office Building, Room X215
Stanford University School of Medicine
Stanford, California 94305-5479

ABSTRACT

Conventional methods for retrieving information from the medical literature are imprecise and inefficient. Information retrieval systems employ unmanageable indexing vocabularies or use full-text representations that overwhelm the user with irrelevant information. This paper describes a document representation designed to improve the precision of searching in textual databases without significantly compromising recall. The representation augments simple text word representations with contextual models that reflect recurring semantic themes in clinical publications. Using this representation, a searcher may indicate both the terms of interest and the contexts in which they should occur. The contexts limit the potential interpretations of text words, and thus form the basis for more precise searching. In this paper, we discuss the shortcomings of traditional retrieval systems and describe our context-based representation. Improved retrieval performance with contextual models is illustrated by example, and a more extensive study is proposed. We present an evaluation of the contextual models as an indexing scheme, using a variation of the traditional inter-indexer consistency experiments, and we demonstrate that contextual indexing is reproducible by minimally trained physicians and medical students.

INTRODUCTION

"Information overload" is a significant problem encountered throughout medical training and practice. Medical students struggle to master an enormous body of basic science knowledge while practicing physicians strive to keep abreast of an expanding collection of published medical research. With the current trend toward the practice of evidence-based medicine, more health care providers are routinely searching the primary literature for research that supports individual medical decisions (1). As high speed networks connect universities, libraries, and medical centers throughout the country, the amount of information that is accessible will only continue to increase.

Unfortunately, several studies suggest that physicians are not keeping pace with the exponentially growing biomedical literature. Clinicians are often unaware of important clinical advances, such as the use of glycosylated hemoglobin in monitoring diabetic control, months or even years after the publication of definitive studies (2, 3). Most physicians who are knowledgeable about state-of-the-art medical research learn of the current standards of care from journals (4). Although individuals from a variety of specialties express a preference for journals as references, physicians overwhelmingly consult other, frequently outdated information sources (5, 6).

The reasons cited for this discrepancy point to the inadequacies of current technology for searching the medical literature. A surprising number of physicians still rely on search intermediaries although online systems could provide them with convenient and timely access to the medical literature. Physicians find the volume of literature overwhelming, the task of sorting out irrelevant information too difficult, and the time to search for information too long (2, 5). Some medical libraries provide photocopying and document delivery, but such services are costly and still require at least 24-48 hours (7). Until easier and more timely methods for finding information are available, physicians may continue to have unmet information needs that directly affect patient care.

CONVENTIONAL RETRIEVAL SYSTEMS

Each of the numerous tools developed for retrieving information from the medical literature (8) performs the task illustrated in Figure 1. An electronic retrieval system is created by translating documents into simplified representations of their content. To retrieve documents from the collection, users describe their information needs in a query representation or language. A matching process measures the similarity of document and query representations and returns articles above a certain threshold. The performance of information retrieval systems is often measured by the metrics *recall* and *precision*. Traditionally, information retrieval systems represent the content of documents with descriptors known as index terms. Human indexers or automatic text-analysis programs construct document representations by selecting appropriate index terms from either (1) a controlled vocabulary that describes the concepts in a document collection or (2) words from the text of the documents. Representations derived from either source have inherent shortcomings that limit their ability to support effective information retrieval. The following paragraphs describe the representations employed by conventional information retrieval systems and compare their performance in terms of precision and recall.

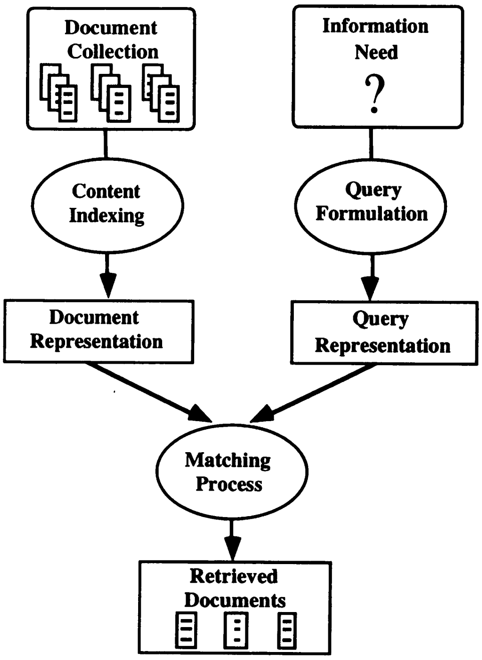

Figure 1. The Information Retrieval Task

Controlled Vocabularies

Controlled indexing vocabularies, such as MeSH (Medical Subject Headings), that encompass concepts from the domain of medicine are enormous. The 1994 MeSH vocabulary contains over 17,000 terms and underwent over 1000 changes in a single year. To search effectively, both indexers and searchers must understand a controlled vocabulary. Professional indexers and medical librarians periodically receive extensive training, but clinicians cannot keep abreast of frequent changes in a large and growing vocabulary for searching (9). MEDLINE indexing is considered state-of-the-art, but the NLM indexers who assign MeSH terms achieve an overall consistency of less than 50% when assigning subject headings to a common set of articles (10, 11). These numbers are troublesome because the consistency of indexing is directly correlated with performance of a retrieval system (12). If professionals cannot manage the indexing vocabulary, even physicians with searching experience are likely to encounter significant difficulties.

Furthermore, vocabularies that represent rapidly growing disciplines such as immunology or molecular biology must evolve with the field. The areas that change rapidly contain those topics most frequently sought in the primary literature (13), yet indexing consistency decreases with frequency of changes in the indexing vocabulary (11). Thus, the concepts that physicians are most likely to seek in literature are described by the parts of the indexing vocabulary that are most likely to produce poor retrieval performance. The unmanageable nature of controlled vocabularies renders them inadequate representations for consistently effective searching (14, 15).

Full-Text Representations

Alternatively, documents can be represented by terms from the text of the articles. Automatic text-analysis programs usually select and weight index terms according to their frequency in the document collection or their ability to discriminate one document from another (16). Representations derived from text words assume that statistical measures of term frequencies can sufficiently reflect the importance of terms in a document. In practice, the presence or frequency of a word cannot precisely measure the relevance of a document to a particular query. Relevance to a specific request depends on the way those terms are used in the text. For example, the presence of a drug or procedure name in the methods section of a clinical research article might be interpreted quite differently from a similar occurrence in the background. Traditional document representations cannot adequately express such distinctions, thus leading to imprecise searching in full-text systems.

Retrieval Performance

Retrieval performance studies have demonstrated that both novice and professional searchers generally achieve better recall when searching in full text than when searching with controlled vocabulary representations, but consistently, the problem with searching in full text is achieving adequate precision. Early research comparing full-text and controlled vocabulary representations of the Harvard Business Review database demonstrated significantly greater relative recall using full-text representations (0.74 vs. 0.28), but somewhat less precision (0.18 vs. 0.34) (17). A similar study using a database of American Chemical Society journals provided more evidence that precision was a substantial problem in searching full text (18). The MEDLINE/Full-Text Research Project (19) compared searches in three databases: the MEDIS full-text database from Mead Data Central, the Comprehensive Core Medical Library (CCML) full-text database from BRS Information Technologies, and MEDLINE. Again, searches from full-text databases produced significantly greater recall and significantly less precision. More recently, comparison of searching with MeSH terms versus full-text abstracts of the AIDSLINE database demonstrated that graduate students, physicians, nurses, and librarians inexperienced with MEDLINE achieved significantly greater recall and statistically equivalent, although absolutely better, precision when searching in full text versus using MeSH terms. Differences in precision and recall for experienced librarians were not statistically significant (20). A study of novice physician and housestaff searchers of MEDLINE demonstrated that recall improves with experience, but precision does not (15). The well-known Blair and Maron experiment (21) suggested that full-text systems retrieve only a small and inadequate fraction of relevant articles in a large databases, but this study was conducted on an unstructured collection of documents including office memoranda and personal communications in a legal setting where attorneys have months to examine evidence for a trial. The results are probably not applicable to the clinical setting where medical decisions must be made in a timely manner, and the literature contains substantial redundancy.

In summary, research shows that full-text systems achieve better recall than systems with controlled vocabularies, but can only occasionally produce equal precision. If the precision of searching could be increased without compromising recall, full-text systems would provide unparalleled retrieval performance. Below, we propose a method for achieving this goal.

IMPRECISION IN SEARCHING FULL TEXT

To better understand the problems of full-text searching, we examined many imprecise searches and discovered a consistent pattern of failure. Irrelevant articles are retrieved when the search terms appear in an inappropriate context. Consider the following example of an information need defined by a physician at the Stanford University Medical Center:

> I need information about the effectiveness of radiation therapy when compared with surgical therapy for treating in situ ductal or lobular carcinoma of the breast (DCIS or LCIS). Effectiveness includes any discussion of survival, morbidity, mortality, or prognosis.

A medical librarian performed a search on the CCML full-text collection of approximately 80 medical journals and constrained the retrieval set to clinical research articles as requested by the physician. The following search was conducted:

```
[(BREAST WITH (DCIS OR LCIS OR SITU OR
        INSITU))OR (BREAST WITH EARLY
        WITH (CANCER OR CARCINOMA))]
WITH    [RADIATION OR IRRADIAT$ OR
        RADIOTHER$]
WITH    [LUMPECT$ OR MASTECTOM$ OR
        SURGERY OR SURGICAL$5]
AND     [SURVIV$ OR MORBIDITY OR
        EFFECTIVE$ OR OUTCOME$ OR
        PROGNOS$ OR MORTALITY]
```

This search retrieves documents in which (1) the terms for early breast cancer occur WITH (i.e., in the same sentence as) the words describing the therapies of interest, and (2) any one of the terms indicating the outcomes of interest are mentioned. This full-text search identified 22 articles, but a physician expert determined only 1 of these articles to be truly relevant. Thus, the precision of this search designed by an experienced searcher is 1 / 22 or 4.5%.

The irrelevant documents from this search demonstrate the ambiguity of search terms taken out of context. Eight articles describe studies of advanced rather than early breast cancer. In these papers, the terms for early breast cancer often occur in the background section or in the exclusion criteria. Six papers evaluate treatments other than surgical or radiation therapy. Since these latter therapies are common in the management of breast cancer, many papers mention them in a discussion of previous work or in describing results for comparison. Six papers are observational studies that do not evaluate any intervention, so the terms for surgery and radiation therapy appear, but not in the context of the interventional methods for the study.

Our description of these erroneously retrieved documents contains references to implicit and sometimes explicit sections of clinical research articles – background, exclusion criteria, interventions, results – and describes terms occurring within the *context* of a particular part of the article. These sections or contexts within an article are familiar constructs to the reader of the medical literature, but they are not explicitly represented in information retrieval systems. We propose a document representation that explicitly models these recurring themes of clinical publications in structures called *context models*. With an explicit representation of document contexts, one can not only search for concepts of interest, but also specify the context in which those terms should occur. The document contexts limit the possible meanings of individual search terms, and thus, facilitate more precise searching in full-text collections.

CONTEXT MODELS

We have developed hierarchically organized context models for four different types of clinical publications: clinical research articles, review articles, case reports, and clinical practice guidelines. The context model for clinical research articles is shown in Figure 2. The primitive element of the model is the *context* or *basic context* which is a proposition that reflects a characterization of the text. For example, the text within the OBJECTIVE context of a clinical research article describes the goal or hypothesis of the research. In a particular article, contexts serve as the basis for interpreting a text, so the same text occurring within different contexts might have different interpretations. For example, the phrase "hemoglobin < 7.5" found in the context of ELIGIBILITY / SELECTION of a clinical research article might be interpreted as a rule for enrolling patients in a clinical trial, but in the context of EXPERIMENTAL FINDINGS, the passage might reflect the value of an experimental variable. The use of context to interpret the meaning of words is analogous to the use of context as the basis for interpretation of clinical data in the temporal reasoning work of Shahar (22).

Contexts that share a common, higher-level characterization form *compound contexts*. For example, text from the STUDY POPULATION METHODOLOGY context describes all procedures used to construct a study group for an experiment. Text within the ELIGIBILITY / SELECTION, EXCLUSION / WITHDRAWAL, AND STUDY GROUP ASSIGNMENT contexts contribute to this characterization, and thus, form the compound STUDY POPULATION METHODOLOGY context. In Figure 2, compound contexts are shown in uppercase, and the component contexts are indented and listed below their parent context. Compound contexts allow a hierarchical organization for the context model, with the root context indicating the type of publication.

```
CLINICAL RESEARCH ARTICLE
    Title
    Authors
    Background
    Objective
    METHODS
        Study Type
        Study Setting
        STUDY POPULATION METHODOLOGY
            Eligibility / Selection
            Exclusion / Withdrawal
            Study Group Assignment
        EXPERIMENTAL PARAMETERS
            Risk / Prognostic Factor Assessment
            Intervention Evaluated
            Concomitant Interventions
            Diagnostic Test Evaluated
            Gold Standard
            Outcome Measures
        Consent / Ethics Procedures
        Statistical Methods
    RESULTS
        Experimental Findings
        Adverse Effects
    Conclusions
    Limitations / Biases
    Future Work
    Acknowledgments / Collaborators
    References
```

Figure 2. Context Model for Clinical Research Articles

The context models were derived from literature describing the content of clinical publications (13), including recommendations for structured abstracts (23-25) and from common section headers that appear in each type of article. We refined the models by examining hundreds of clinical articles and selecting only the contexts that consistently recurred.

Many of the contexts are identical to section headings that occur in clinical publications, but the text under a particular heading will not always be part of the corresponding context. Although clinical publications contain predictable categories of information, information of a particular type is often scattered throughout a document. Section headings represent these typical categories of information, such as METHODS and RESULTS, but these delimiters dictate format rather than content, and thus, often inadequately reflect the nature of the text they precede. For example, experimental methodology is often described under the heading of RESULTS in order to juxtapose the description of a methodology and the results produced. BACKGROUND information, such as the results of previous studies, is also frequently found under the heading of RESULTS, in order to facilitate comparison.

Advocates of structured abstracts have suggested that such headings might be useful for information retrieval (26, 27). As the previous examples illustrate, these headings offer a framework for the format of a paper, but do not accurately represent the semantic content of the text. Mandating strict adherence to the headings is probably not an acceptable solution since flexibility in constructing prose is often necessary for producing interesting and understandable articles. For example, a complex experiment with multiple steps may be best explained by interspersing METHODS and RESULTS in the text; a strict separation of these categories of information might produce confusing descriptions.

Context models offer a method for representing semantic structure without restricting the presentation style of the author. To employ this representation, documents must be analyzed, and the contexts that apply to each section of the document must be indicated. This process is called *context markup*. Currently, we ask humans to perform this process manually, but ultimately, text-analysis programs might automatically complete this task. Context markup occurs at the level of the sentence since this grammatical construct allows the expression of a complete thought. Each sentence may have multiple characterizations (e.g., a sentence may describe the objective and setting), so more than one context may be assigned to a sentence. When an entire collection has been marked, searchers may conduct a *context-based search* by specifying both keywords and the contexts in which they should appear.

The context-based search statement shown below addresses the information need presented earlier:

```
In STUDY POPULATION METHODOLOGY:
    [(BREAST WITH (DCIS OR LCIS OR SITU
    OR INSITU))OR (BREAST WITH EARLY WITH
    (CANCER OR CARCINOMA))]
AND In INTERVENTION EVALUATED: [(BREAST
    WITH (DCIS OR LCIS OR SITU OR
    INSITU))OR (BREAST WITH EARLY WITH
    (CANCER OR CARCINOMA))] WITH
    [(RADIATION OR IRRADIAT$ OR
    RADIOTHER$) WITH (LUMPECT$ OR
    MASTECTOM$ OR SURGERY OR SURGICAL$5)]
AND In OUTCOME MEASURES: [SURVIV$ OR
    MORBIDITY OR EFFECTIVE$ OR OUTCOME$
    OR PROGNOS$ OR MORTALITY]
```

This search requires that the terms for breast cancer occur in the STUDY POPULATION METHODOLOGY context, that the terms for breast cancer and the therapies of interest appear within the same sentence in the INTERVENTION EVALUATED context, and the terms describing outcome measures occur within the OUTCOME MEASURES context. This search constitutes a subset of the full-text search, so the context-based search cannot retrieve any articles that were not part of the original retrieval set. Therefore, to determine retrieval performance, we asked an individual blinded to the search topic to mark the 22 articles retrieved by the full-text search and performed the context-based search on this set. The search retrieved only 3 articles, including the 1 article previously determined to be relevant. Thus, the context-based search improved the precision of the full-text search from 4.5% to 33.3% without a loss of recall.

This example illustrates the type of improvement in retrieval performance that context-based searching might provide. With only a single relevant document identified by both searches, we did not observe a decrease in recall by constraining the search terms with contexts, but with more relevant items, some loss of recall is possible. We believe that context-based searching will produce a significant increase in precision without a substantial decrease in recall when compared with the performance of traditional full-text searching, and we are currently testing this hypothesis in a comparative study.

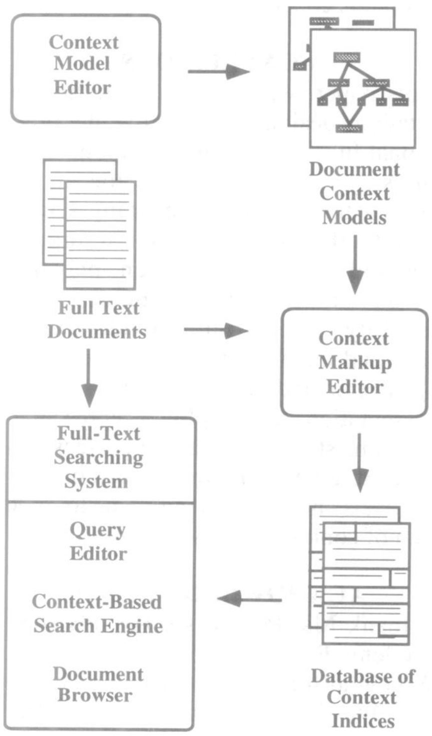

Figure 3. Context-Based Searching Tools

CONTEXT-BASED SEARCHING TOOLS

We have developed an environment for context-based searching illustrated in Figure 3. A Context Model Editor allows the user to define context models for clinical publications. Once the context models are specified, a Context Markup Editor facilitates the markup of clinical articles and creates a set of indices that indicate the locations of individual contexts with the documents. Finally, the Full-Text Searching System guides the user in constructing context-based searches and in viewing the results.

EVALUATION

Our research proposes an innovative document representation for information retrieval. Documents are represented by both text words and the contexts in which they occur, and searchers may indicate combinations of terms and contexts in their queries to the system. This representation is expected to facilitate improved retrieval performance for several reasons. Firstly, the controlled part of the representation, the context model, describes the *structure of the literature* and thus, is relatively small compared with the indexing vocabularies such as MeSH that encompass all of medicine. Furthermore, although the domain of medicine is evolving rapidly, the literature changes more slowly. Texts describing the structure and content of scientific publications that were written ten to twenty years ago remain valid today (13, 29). The small and stable nature of the context models renders the tasks of indexing and searching with this representation more manageable.

In addition, the contexts provide a method for eliminating irrelevant articles without losing pertinent material. Since physicians generally achieve adequate recall when searching in full text, but are overwhelmed with irrelevant documents, context-based searching offers a method for effectively narrowing these searches.

To evaluate the context-based representation, the first challenge is to test the utility of the context models as an indexing scheme. Traditionally, MeSH has been evaluated through inter-marker studies of professional indexers that report an average percentage of agreement among indexer pairs (10, 11). This methodology has several shortcomings because (1) the experiments do not assess the ability of the expected searchers (i.e., librarians or physicians) to understand the indexing scheme, and (2) the evaluation metric does not account for chance agreement among indexers.

We are currently evaluating our context models using an inter-marker consistency study similar to those conducted at the NLM, but with several important differences. Our subjects consist of senior medical students and physicians, the expected end users of the context models. Second, our evaluation metric is the kappa coefficient of agreement for nominal scales (30, 31); this metric represents the average agreement among markers *beyond the agreement due to chance.*

For each type of clinical publication, we asked 5 senior medical students and physicians to mark 5 articles with the appropriate context model. The subjects were selected from available volunteers with no prior exposure to the context models. Each group was instructed for 1.5 hours on the use of the context model, and the 5 test articles were chosen randomly from the most recent issues of the *New England Journal of Medicine*, *The Annals of Internal Medicine*, and the *Journal of the American Medical Association*. Subjects assigned one or more contexts to each sentence, and a kappa was computed for each context. This method not only provides an overall picture of the reproducibility of the model, but also indicates the most commonly confused contexts.

Pilot studies were performed for several of the context models using 5 subjects and 3 articles, and the context models were revised based on these preliminary findings. Not all contexts were uniformly marked by the subjects, and the lower kappas pointed to deficiencies in the context models. Contexts that were frequently confused were consolidated or more clearly defined, and several contexts were eliminated or renamed to reflect more appropriately their definitions.

The results of the inter-marker study for the revised model of clinical research articles are shown in Figure 4. Similar experiments for other types of clinical publications are in progress. The graph depicts the number of basic contexts that achieved a given range of kappa values. Compound contexts were not used in the markup process since they are formed from basic context components.

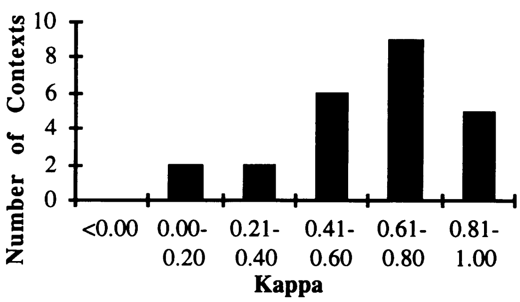

Figure 4. Results of Inter-marker Study for Clinical Research Articles

Although an absolute standard for comparison is not available, these benchmarks defined by Landis (32) provide a useful guide in assessing the agreement achieved by the markers:

Kappa	Interpretation
< 0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

For 20 (83%) of the 24 basic contexts, markers achieved a kappa of 0.40 or better, indicating that the raters had moderate or better agreement, and for 14 contexts (58%), the markers demonstrated substantial or better agreement. Although a direct comparison is not appropriate, we can informally compare this experiment with the performance of NLM indexers. NLM indexers train for over 1 year, and achieve an overall indexing consistency of 61.1% for central concept headings, 48.2% for all subject headings, and less than 45% for heading/subheading combinations, without accounting for agreement due to chance. In contrast, the markers in this study trained for only 1.5 hours, and achieved agreement beyond chance of over 60% for over 58% of the contexts. This results suggests that with brief training, the expected users of a context-based retrieval system can understand and apply the context models as an indexing scheme.

FUTURE WORK

The preliminary results of the inter-marker studies provide some validation of the context models as a useful indexing scheme, but ultimately, the quality of the representation depends on the ability to improve retrieval performance. Currently, we have only a small number of examples, such as the search presented in this article, to support our hypothesis of improved precision in searching without compromising recall. To test definitively our representation, we are preparing a 500-article document collection for markup with the revised context models. We plan to perform a precision / recall study comparing the retrieval performance of context-based searches and traditional full-text searches on this collection.

CONCLUSIONS

In this paper, we present contextual document representations designed to improve the precision of conventional full-text searching without significantly compromising recall. We describe an improved methodology for evaluating indexing schemes, and we illustrate that with minimal training, novice physicians and medical students can apply the context models as an indexing scheme with good reproducibility. This result provides evidence that physicians and medical students can understand the indexing scheme and might more easily search using this representation than employ a representation that cannot be consistently applied by professionals. We present an example of improved searching using the context models and describe our plans for a definitive evaluation of context-based searching.

ACKNOWLEDGMENTS

We thank Ms. Purcell's other advisors, Leslie Lenert, Glenn Rennels, Thomas Rindfleisch, and Terry Winograd for their invaluable guidance in this research. We thank Gloria Linder for her aid in constructing full-text searches and BRS Information Technologies for the use of the CCML database for experimentation. This work is supported by the NLM under grant LM-07033, and the computing facilities are provided by the CAMIS Resource through grant LM-05208 from the National Library of Medicine. Dr. Shortliffe is supported by the Henry J. Kaiser Family Foundation under Grant #84R-2459-HPE.

References

1. Evidence-Based-Medicine-Working-Group. Evidence-based medicine: a new approach to teaching the practice of medicine. JAMA 1992;268(17):2420-25.
2. Williamson JW, German P S., Weiss R, Skinner EA, Bowes F. Health sciences information management and continuing education of physicians. Annals of Internal Medicine 1989;110:151-160.
3. Stross JK, Harlan WR. The dissemination of new medical information. JAMA 1979;241:2622-24.
4. Stross JK, Harlan WR. Dissemination of relevant information on hypertension. JAMA 1981; 246:360-62.
5. Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met? Annals of Internal Medicine 1985;103:596-99.
6. Woolf SH, Benson DA. The medical information needs of internists and pediatricians at an academic medical center. Bulletin of the Medical Library Association 1989;77(4):372-80.
7. Broering NC, King E. Meeting the urgency for document delivery in clinical medicine. Seventeenth Annual Symposium on Computer Applications in Medical Care. Washington, D.C. McGraw-Hill, Inc., 1993;576-80.
8. Hersh WR, Greenes RA. Information retrieval in medicine: state of the art. M.D. Computing 1990;7(5):302-11.
9. Humphrey SM. Indexing biomedical documents: from thesaural to knowledge-based retrieval systems. Artificial Intelligence in Medicine 1992;4:343-71.
10. Lancaster FW. Evaluation of the MEDLARS demand search service. U.S. Dept. of Health, Education, and Welfare, Public Health Service, 1968.
11. Funk ME, Reid CA. Indexing consistency in MEDLINE. Bulletin of the Medical Library Association 1983; 71(2):176-83.
12. Leonard LE. Inter-indexer consistency and retrieval effectiveness: measurement of relationships. Ph.D. Thesis. Univ. of Illinois, Champaign, 1975.
13. Warren KS. Coping with the biomedical literature. New York: Praeger Publishers, 1981.
14. Mitchell JA, Johnson ED, Hewett JE, Proud VK. Medical students using Grateful Med: analysis of failed searches and a six-month follow-up study. Computers and Biomedical Research 1992;25:43-55.
15. McKibbon KA, Haynes RB, Dilks CJW, et al. How good are clinical MEDLINE searches? A comparative study of clinical end-user and librarian searches. Computers and Biomedical Research 1990;23:583-93.
16. Salton G, McGill MJ. Introduction to modern information retrieval. New York: McGraw-Hill, Inc., 1983.
17. Tenopir C. Contributions of value added fields and full-text searching in full-text databases. Sixth national online meeting. Medford, NJ: Learned Information, Inc., 1985: 463-470.
18. Abbott JP, Smith CR. Full-text and bibliographic ACS databases: rivals or companions. Sixth National Online Meeting. Medford, N.J.: Learned Information, Inc., 1985: 5-9.
19. McKinin EJ, Sievert M, Johnson ED, Mitchell JA. The MEDLINE/full-text research project. Journal of the American Society for Information Science 1991;42(4):297-307.
20. Hersh WR, Hickam DH. A comparison of retrieval effectiveness for three methods of indexing medical literature. American Journal of the Medical Sciences 1992;303:292-300.
21. Blair DC, Maron ME. An evaluation of retrieval effectiveness for a full-text document retrieval systems. Communications of the ACM 1985;23(3):289-299.
22. Shahar Y, Purcell GP. The context-sensitive pattern-matching task. Proceedings of the IJCAI Workshop on Context. Montreal, Canada, August 1995. In press.
23. Mulrow CD, Thacker SB, Pugh JA. A proposal for more informative abstracts of review articles. Annals of Internal Medicine 1988;108:613-15.
24. Ad-Hoc-Working-Group-for-Critical-Appraisal-of-the-Medical-Literature. A proposal for more informative abstracts of clinical articles. Annals of Internal Medicine 1987;106(598-604).
25. Hayward RSA, Wilson MC, Tunis SR, Bass EB, Rubin HR, Haynes RB. More informative abstracts of articles describing clinical practice guidelines. Annals of Internal Medicine 1993;118:731-37.
26. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. Annals of Internal Medicine 1990;113(1):69-76.
27. Huth EJ. Structured abstracts for papers reporting clinical trials. Annals of Internal Medicine 1987;106(4):626-27.
28. NLM. Medical subject headings – annotated alphabetic list. Bethesda, MD: US Department of Health and Human Services, Public Health Service, NIH, 1994.
29. Bailar JC, Louis TA, Lavori PW, Polansky M. A classification for biomedical research reports. NEJM 1984;311(23):1483-1487.
30. Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 1960;20(1):37-46.
31. Fleiss J. Measuring nominal scale agreement among many raters. Psychological Bulletin 1971;76(5):378-82.
32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174.