Methods of Information in Medicine


**J. J. Cimino**

# Linking Patient Information Systems to Bibliographic Resources

Department of Medical Informatics, Columbia University, New York, NY, USA

**Abstract:** Medical informatics researchers have explored a number of ways to integrate medical information resources into patient care systems. Particular attention has been given to the integration of on-line bibliographic resources. This paper presents an information model which breaks down the integration task into three components, each of which answers a question: what is the user's question?, where can the answer be found?, and how is the retrieval strategy composed? Twelve experimental systems are reviewed and their methods for addressing one or more of these questions are described.

***Keywords:*** Clinical Information Systems, Bibliographic Systems, Literature Searching, Information Retrieval

## Introduction

One area of active research in medical informatics has been the development of methods for providing a variety of information sources to health-care providers, including clinical information systems and bibliographic resources. Initial efforts resulted in sophisticated hardware and operating system solutions which allowed diverse applications to be accessible from the same terminal, computer or workstation. More advanced approaches have sought to integrate these information sources at the application level so that, for example, patient data can be used to drive literature retrieval strategies. This paper reviews these latter approaches and examines the various ways in which such systems determine the users' questions, identify appropriate information sources, and conduct queries to these sources.

### *Linking Clinical and Bibliographic Systems*

Bibliographic retrieval systems are often represented as valuable tools for the practicing health care worker [1], and there is some evidence to support this belief [2]. The belief that clinicians have difficulty using such resources effectively appears to be equally valid [3-5]. In the past, potential users have had difficulty getting access to retrieval software, because of limited availability of terminals in convenient settings. However, initiatives such as the National Library of Medicine's (NLM) Integrated Advanced (formerly Academic) Information Management Systems (IAIMS) are making these resources available through the same workstations that are being used to access clinical information. Typically, systems are "integrated" using an "IAIMS Menu" which serves as a central access point for a variety of applications, including clinical and bibliographic systems [6]. This centralized approach improves access, but it does little to help users with two remaining difficulties: navigating retrieval software and composing effective search strategies. This paper discusses a variety of techniques which are being employed to address these problems for users of clinical systems wishing to retrieve medical information. The focus of this review, like the vast majority of work published to date, is on strategies for conducting bibliographic searches using information from clinical systems.

Bibliographic and clinical systems intersect at a particularly interesting point in patient care information management. Health-care workers using a clinical system will invariably encounter patient information which raises questions that can be answered through a survey of recent medical literature [7]. For example, a laboratory test result might indicate some condition for which new therapies are being devised. Having a bibliographic retrieval system available at that moment offers an opportunity to educate the user about the particular problem in the particular patient – the case-based teaching method in action. Thus, the potential exists for bibliographic systems, available at the bedside, to have an impact on both the care of the particular patient and on the education of the user in preparation for future encounters.

A number of researchers are exploring ways to integrate systems to improve bibliographic search results. The basic idea is that if clinical data trigger the need for additional information, those same data can be used in some automated way to improve the search process, both in terms of ease of use and quality of result. For example, if a laboratory test result is the trigger, it might be incorporated automatically in the search strategy to provide a simple way to conduct a relevant search. Formal information models are used to identify the specific role of the data in the retrieval.

## Questions to Address in an Information Model

Creating a model for integrating clinical data into the literature-retrieval process first requires a model of the process of responding to an information need. The user's request for information can be decomposed into three distinct questions:
- What is the question?
- Where can the answer be found?
- How is the retrieval strategy composed?

If a system can answer correctly one or more of these questions by itself, it is plausible that less interaction with the user will be required, making the system easier to use. Figure 1 shows a number of possible ways in which clinical data can be used to help a system answer these questions.

In the IAIMS Menu approach (A), the user of a clinical system is allowed to access a bibliographic system at will, but the system does not use clinical data in the retrieval. Tighter integration is accomplished (B) by extracting information, usually as text, and then incorporating it into a search strategy (C). The system might also use clinical information to help in identifying the particular topic of interest (D) which can then be used to produce a more relevant search strategy (E). Finally, the system might provide the capability, based on the topic selected by the user, to select an information source other than a bibliographic database, if appropriate (F).

### *Question 1: Identifying the User's Question*

The simplest possible question is of the form "Tell me about X." Simple integration between clinical and bibliographic systems can be accomplished by allowing the user to access the retrieval system (Path A in Fig. 1) and use it in the normal manner. The user is free to ask any desired question (within the constraints of the system's user interface), but gets no help with navigation or search formulation. One system which improves on this method is the Medical Desktop, a Microsoft Windows-based application developed at the State University of New York at Buffalo [8]. This system uses standard Windows features (cut-and-paste and dynamic data exchange) to allow the user to select desired character-based information from the clinical system and paste it into a Term Linker. The Term Linker attempts to identify a corresponding concept in the NLM's Unified Medical Language System (UMLS) Metathesaurus [9] to allow the user to convert the information into a standard form such as the NLM's Medical Subject Headings (MeSH). The concept is then transferred to a Resource Database program from which a list of appropriate sources (based on the semantic type of the concept) is composed. The user selects a source and the concept is transferred to that source as part of a search strategy.

A similar approach can be found in the Meta-1 Front End at Yale University [10]. This system displays a patient report and allows selection of a word or phrase of interest. The text is then translated to MeSH, using the UMLS Metathesaurus. From the Metathesaurus, associated concepts are located and displayed to the user. The user can then choose desired MeSH terms to construct a Medline search strategy which is then passed on to Grateful Med for use with Medline.

No attempt is made in either of these systems to determine information about the kind of retrieval that is of interest to the user. Researchers are now exploring ways in which the information being viewed in the clinical system can help the system decide what questions are being asked.

One of the first systems to explore this idea, developed at Yale University, was Hepatopix (He'-pa-top'-ix) [11] which dealt with topics in liver disease, followed by Psychtopix [12] with its focus on psychiatric disorders. Each of these systems scanned text reports (pathology reports and psychiatric histories, respectively) from clinical record systems for "topics of interest". The user would then choose the topic of interest and the corresponding search, stored in a database, was then sent to Medline, again using Grateful Med. This approach was a revolutionary step in the use of medical bibliographic systems: by attempting to use the clinical record to anticipate the users' needs, these systems provided users with high-quality search strategies. Even when a topic was "irrelevant" with respect to the record at hand, it might still be of interest to the user. The topic sets were presented in a hierarchical manner so that browsing them was simple and logical.

The "topix" approach is not without drawbacks. One problem is that the specificity found in the patient record

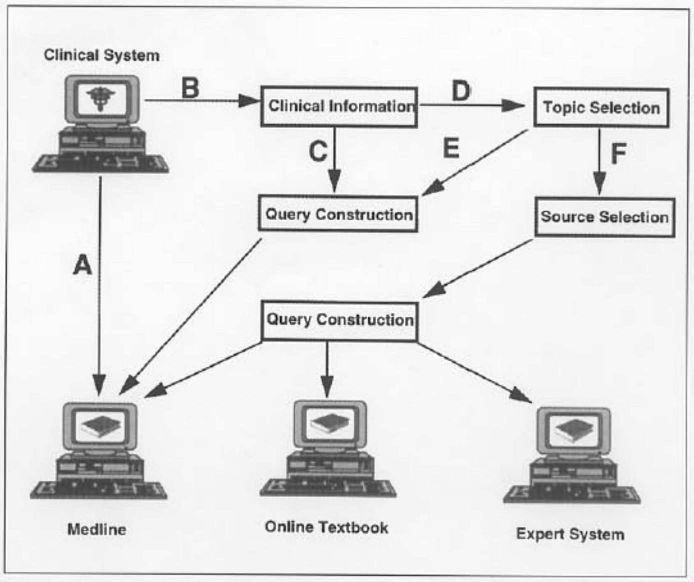


**Fig. 1** Various approaches for integrating clinical systems with on-line information sources. The letters refer to different paths by which clinical information might be processed for use in queries. See text for detailed descriptions of each path.

may be lost when making the transition to the topic. For example, if the terms "colon", "sigmoid" or "rect" appear in the text of a case report, the topic related to colorectal carcinoma might be evoked; however, searches specific to the particular anatomic location in the report (e.g., sigmoid vs rectum) could not be performed by this approach. In some cases, there may be no need for such specificity. In this example, for instance, the same citations might be retrieved regardless of the anatomic location. A more significant problem encountered by the Yale researchers was the extreme specialty-specific nature of the topics. A great deal of expert effort was needed to construct the topic sets used in Hepatopix and Psychtopix; moving to new domains required starting again with new sets of experts. Similarly, the search strategies were domain-specific and not easily modified for use in a different medical specialty area, thus limiting the extensibility of the technique.

Researchers at the University of North Carolina and Duke University developed a similar approach to question selection based on specific information found in a clinical system [13]. In this case, the information was of a graphical, rather than textual nature. A pilot project allowed a user viewing blood-gas results to select a data point on an acid-base nomogram. Based on the location on the diagram, a particular medical topic was inferred (acute metabolic acidosis, chronic respiratory alkalosis, etc.). The system could then present the user with a number of specific questions about the disorder, each of which was associated with a predefined search strategy. The system allowed the user to restrict the strategy if desired, and then transmit it to the NLM as a Medline search. The specificity of the selected questions was appropriate for the domain chosen in the pilot project. The real "intelligence" in the system lay in its ability to infer a condition from the raw laboratory data. The applicability of this inferencing mechanism to other domains has not yet been studied.

Researchers at the University of Pittsburgh have developed a system called CHARTLINE [14] which examines a patient record for occurrences of terms which are recognized as concepts in the UMLS Metathesaurus. Topics of interest are proposed by the system using MeSH term co-occurrence data provided with the UMLS. Users then select a "topic" which is actually a search, composed of a term from the record and some other term (which may or may not be in the record) using co-occurrence information, CHARTLINE indicates the number of citations that would be retrieved, even before the search is performed. CHARTLINE provides a fairly good match between the specificity of the original clinical terms and those in the Metathesaurus. The number of potential topics relevant to a clinical record can be quite large, however, and a user can easily become overwhelmed when trying to find the desired one in the list.

The group at Columbia University has applied a different technique in an attempt to identify manageable lists of topics which are nevertheless specific to the particular clinical information at hand [15]. This approach involved the development of a list of "generic queries" which were derived from analysis of a set of questions posed to reference librarians. An analysis of the various clinical applications was also made to determine the types of medical concepts which might appear in any particular setting. The intent was to allow users to select one or more relevant data items from the clinical system and then, based on the types of terms selected, match-up a relatively small list of generic queries. The specific concepts chosen from the clinical system were then incorporated into the generic question to make it specific to user's question. In the first implementation of this approach (the Medline Button), an application displaying ICD-9 codes for diseases and procedures related to hospital admissions was linked to Medline [16]. The system, running on the hospital's mainframe computer, used the UMLS Metathesaurus to translate selected terms into appropriate MeSH expressions of one or more MeSH terms. The system also selected questions, based on the terms selected; for example, one set of questions dealt with a single disease and a second set of questions dealt with a single disease and a single procedure. When a question was selected, the terms were filled into appropriate blanks of a corresponding Medline search strategy which was then sent to the Medline system, running on the same mainframe. More recently, the Medline Button has been adapted for use by clinical applications based on the World Wide Web [17-19]. In this environment, we have been able to expand the available information sources to include full-text sources and expert systems in addition to Medline.

The Interactive Query Workstation (IQW) from Massachusetts General Hospital and Harvard University uses generic queries with a somewhat different approach [20]. Rather than attempt to determine what questions the user might ask, the developers of IQW have attempted to determine what questions various information sources can answer. The system makes use of a clinical database (COSTAR) and a knowledge base of information sources which contains standardized representations of source characteristics, including the types of queries they can answer. Terms selected from the clinical system are identified in the UMLS and, based on their semantic type, are matched-up to appropriate queries and their corresponding source. The user selects a query of interest and provides additional information, if requested by the system, to complete the query. The system then establishes a connection to the information source, passes it the search strategy and logs the results for subsequent display to the user.

### *Question 2: Identifying the Appropriate Information Source*

When more than one information source is available, a number of approaches are possible for selecting the most appropriate one. The simplest is to go to a default source covering the general domain of medicine such as Medline. Most of the systems described in the previous section use this approach. Another simple approach is to ask the user to pick a source from a list. The Medical Desktop, for example, uses this method, as did an early version of the IQW. However, if the user is unfamiliar with the possible choices, a poor selection might be made.

When a predefined list of topics is available, the selection of a particular information source could be done by an expert searcher (such as a medical librarian) and then "hard wired" to the topic. Applications which run on Internet-based systems are able to link to other sources with relative ease. For example, one system at Duke University makes use of the Internet Gopher protocol. This system does not seek to establish user information needs at all, but rather anticipates them by providing hard-wired links from a clinical system (Care Maps) directly to a variety of information resources on the Internet [21]. Care Maps includes information about common patient problems, including standard interventions and the expected outcomes of those interventions. Embedded in particular points in the Care Maps are pointers to specific relevant information sources. For example, a user reviewing a clinical alert based on an abnormal laboratory test result is offered the option of retrieving the full text of a relevant publication, the medical logic which generated the alert, or an appropriate image from an image library.

Some researchers are making use of a second UMLS resource, the Information Sources Map (ISM), to help with source selection. The ISM contains records describing a variety of on-line information sources; the records describe the topics they cover and the indexing vocabulary they use. The information known about the user's question can be matched to the ISM in order to determine the most useful source or sources. If more than one is available, the user can get source descriptions from the ISM to make an intelligent selection. For example, the IQW includes a knowledge base derived in part from the ISM. When a user selects clinical terms and IQW identifies relevant queries, the system can use both sets of information to determine the appropriate information sources for a given generic question which has been filled-in with specific terms. Similarly, the NetMenu system at Yale makes use of the ISM to allow users to select information sources from those available via their medical-center network [22].

*Question 3:*
*Composing the Retrieval Strategy*

The ability for a system to generate high-quality search strategies depends on quantity and quality of the information available about the user's information need. When a system has a list of terms of interest, such as the Medical Desktop or the Meta-1 front-end, the user may use accepted forms of those terms (through linkage by the UMLS) and may get some help from the search-engine interface in finding appropriate syntax for the search, but the inclusion of sophisticated topic-specific semantic information, such as would be provided by a reference librarian, is not possible.

In contrast, those systems which are able to anticipate the user's information needs can develop pre-assigned search strategies. In some cases, the entire search strategy is "canned", such as for the Duke Internet Gopher, the TMR-NLM system, Psychtopix and Hepatopix. Quality of the search results with this approach can be quite high, as has been demonstrated with Hepatopix.

Other systems with predetermined queries developed only a general framework for search strategies. Because users have the freedom to select a wide variety of clinical terms, the pre-compilation of all possible search strategies is impossible. However, by creating frameworks for search strategies, it is possible to "fill in the blanks" once the specific user's terms become known. Systems such as CHARTLINE and the IQW use this approach. The Medline Button makes use of a formal knowledge base of librarian expertise for translating queries into appropriate search strategies [23].

## Discussion

The systems reviewed in this paper were generally developed as proof-of-concept prototypes and, as such, have had very limited evaluation. Attempts to address the problems of information retrieval for users of clinical systems make use of a variety of creative techniques. A common thread in all of these systems is the attempt to let the clinical data being reviewed serve some role in the information requests that they are presumably responsible for stimulating. We are only beginning to understand how to take best advantage of having this clue to what the user finds interesting. Although work has been done which recognizes the information needs of clinicians in traditional practice [24], very little has been done to understand the specific information needs engendered by clinical systems [25].

The use of Medline as the preferred information source is also common, although not universal. The preference is largely based on Medline's ready accessibility and retrieval capabilities. But, for all of its careful indexing, the Medline database is not capable of satisfying queries such as "what's the answer to X?" but rather "what citations might have the answer to X?". As more is learned about the information needs which occur while the clinician is on line, specific resources may be developed to pinpoint specific questions, much as the Duke group has incorporated specific Gopher pointers into their system.

In addition to having better on-line information, improvements are needed in the interfaces which provide access to the information. Systems which attempt to access multiple information sources must currently struggle with the peculiarities of each, much as would a human user. Standards for access (such as WAIS [26]), retrieval (such as Z39.50 [27]) and presentation (such as the World Wide Web [17, 18]) are now reaching the point where information requests can be carried out successfully without intimate knowledge of the target application.

Finally, it is worth noting that almost all of the aforementioned systems make use of the UMLS in one way or another. This is consistent with the goal of the UMLS to facilitate integrated access to disparate on-line information sources. While many of the initial systems were developed specifically as part of NLM-sponsored UMLS research, the use of the UMLS Knowledge Resources had spread to independently-funded groups. However, further development of the Knowledge Sources is needed to support the linking of clinical and bibliographic systems. At present, the Metathesaurus is primarily used in inter-vocabulary translation when two vocabularies have exact synonyms. More experiments, as well as changes in the Methathesaurus content, will be needed to develop algorithms for reliable translation in other cases. Continued development of the ISM, based on research at Yale and Harvard, may lead to methods for representing information sources in such a way that source selection, query formulation and retrieval can all be automated. Vendors of information sources should be encouraged to provide ISM records to the NLM for all of their products. In addition, the use of standardized vocabularies for indexing on-line materials should be encouraged, with use of existing standards or coordination of new standards, where needed, with the on-going UMLS Metathesaurus development work.

The last decade witnessed rapid expansion of the availability of on-line clinical systems and information sources. While such systems initially were found in isolated, stand-alone environments, the next decade promises to be one in which they move through the continuum of being linked, to being coordinated with "cut and paste" approaches, to being truly integrated [6]. Additional work is needed to understand the information needs of different users in different settings and how to satisfy those needs through more sophisticated selection of information resources and translation of concepts from the clinical application to these resources. Integration of these systems is the essence of medical informatics: understanding the flow of information itself and discovering ways to improve it.

**Acknowledgments**

This work was supported in part by Unified Medical Language System (UMLS) and High Performance Computing and Communication (HPCC) contracts from the National Library of Medicine.

REFERENCES

1. Guyatt GH. Evidence-based medicine. ACP Journal Club, Ann Intern Med 1991; 114 (Suppl 2): A-16.
2. Lindberg DAB, Siegel ER, Rapp BA, Wallingford KT, Wilson SR. Use of MEDLINE by physicians for clinical problem solving. JAMA 1993; 269: 3124-9.
3. Borgman CL. Why are online catalogs hard to use? Lessons learned from information-retrieval studies. J Amer Soc Inform Science 1986; 37: 387-400.
4. McKibbon KA, Haynes RB, Dilks CJ et al. How good are clinical MEDLINE searches? A comparative study of clinical end-user and librarian searches. Comput Biomed Res 1990; 23: 583-93.
5. Osheroff JA, Bankowitz RA. Physicians' use of computer software in answering clinical questions. Bull Med Libr Assoc 1993; 81: 11-9.
6. Cimino JJ, Sengupta S. IAIMS and UMLS at the Columbia-Presbyterian Medical Center. Med Decis Making 1991; 11 (Suppl): S89-93.
7. Gorman PN, Ash J, Wykoff L. Can primary care physicians' questions be answered using the medical journal literature? Bull Med Libr Assoc 1994; 82: 140-6.
8. Loonsk JW, Lively R, TinHan E, Litt H. Implementing the Medical Desktop: tools for the integration of independent information resources. In: Clayton PD, ed. *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care.* Washington, DC 1991: 574-7.
9. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Meth Inform Med 1993; 32: 281-91.
10. Powsner SM, Miller PL. From patient records to bibliographic retrieval: a Meta-1 front-end. In: Clayton PD, ed. *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care.* Washington DC, 1991: 526-30.
11. Powsner SM, Riely CA, Barwick KW, Morrow JS, Miller PL. Automated bibliographic retrieval based on current topics in hepatology: Hepatopix. Comput Biomed Res 1989; 22: 552-64.
12. Powsner SM, Miller PL. Automated online transition from the medical record to the psychiatric literature. Meth Inform Med 1992; 31: 169-74.
13. Hammond JE, Hammond WE, Stead WW. Information management through integration of distributed resources: the TMR-NLM connection and prototype. In: Miller RA, ed. *Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care.* Washington DC 1990: 719-23.
14. Miller RA, Gieszczykiewicz FM, Vries JK, Cooper GF. CHARTLINE: providing bibliographic references relevant to patient charts using the UMLS Metathesaurus Knowledge Sources. In: Frisse ME, ed. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1992: 86-90.
15. Cimino JJ, Aguirre A, Johnson SB, Peng P. Generic Queries for Meeting Clinical Information Needs. Bull Med Libr Assoc 1993; 81: 195-206.
16. Cimino JJ, Johnson SB, Aguirre A, Roderer N, Clayton PD. The Medline Button. In: Frisse ME, ed. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1992: 81-5.
17. Schatz BR, Hardin JB. NCSA Mosaic and the World Wide Web: global hypermedia protocols for the Internet. Science 1994; 265: 895-901.
18. McKinney WP, Wagner JM, Bunton G, Kirk LM. A guide to Mosaic and the World Wide Web for physicians. MD Comput 1995; 12: 109-14, 141.
19. Cimino JJ, Socratous SA, Clayton PD. Internet as clinical information system: application development using the World Wide Web. JAMIA 1995; 2: 273-84.
20. Cimino C, Barnett GO. Standardizing access to computer-based medical resources. In: Miller RA, ed. *Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care.* Washington DC 1990: 33-7.
21. Hales JW, Low RC, Fitzpatrick KT. Using the Internet Gopher protocol to link a computerized patient record and distributed electronic resources. In: Safran C, ed. *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1993: 621-5.
22. Miller PL, Frawley SJ, Wright L, Roderer NK, Powsner SM. Lessons learned from a pilot implementation of the UMLS Information Sources Map. JAMIA 1995: 102-15.
23. Peng P, Aguirre A, Johnson SB, Cimino JJ. Generating MEDLINE strategies using a librarian knowledge-based system. In: Safran C, ed. *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care.* New York: McGraw-Hill 1993: 596-600.
24. Covell DG, Uman GC, Manning PR. Information needs in the office practice: are they being met? Ann Intern Med 1985; 103: 596-9.
25. Tang PC, Patel VL. Major issues in user interface design for health professional workstations: summary and recommendations. Intern J Biomed Comp 1994; 34: 139-48.
26. Dyson E. WAIS: many ways to do it (wide-area information servers). Release 1.0 1991; 91: 7-8.
27. *Information Retrieval Service Definition and Protocol Specification for Library Applications.* No. Z 39.50-1988. Document Center 1989.

Address of the Author:
James J. Cimino, M. D.,
Atchley Pavilion, Room 1310,
Columbia-Presbyterian Medical Center,
161 Fort Washington Avenue,
New York, New York 10032, USA
E-mail: james.cimino@columbia.edu
Voice: 212-305-8127
Fax: 212-305-3302