MedWeaver: Integrating Decision Support, Literature Searching, and Web Exploration using the UMLS Metathesaurus

William M. Detmer[1], G. Octo Barnett[2], and William R. Hersh[3]

[1]Section on Medical Informatics, Stanford University School of Medicine*
[2]Laboratory of Computer Science, Massachusetts General Hospital
[3]Biomedical Information Communication Center, Oregon Health Sciences University

Integrating functions from disparate and widely-distributed information systems has been an interest of the medical informatics community for some time. Barriers to progress have included the lack of network-accessible information sources, inadequate methods for inter-system messaging, and lack of vocabulary translation services. With the advent of the World Wide Web (WWW) and the evolution of the National Library of Medicine's Unified Medical Language System (UMLS), it is now possible to develop applications that integrate functions from diverse, distributed systems. In this paper we describe one such system, MedWeaver, a WWW application that integrates functions from a decision support application (DXplain), a literature searching system (WebMedline), and a clinical Web searching system (CliniWeb) using the UMLS Metathesaurus for vocabulary translation. This system demonstrates how application developers can design systems around anticipated clinical information needs and then draw together the needed content and functionality from diverse sources.

Studies have shown that information needs that arise in clinical practice are frequently unmet [1, 2]. Some of the barriers to satisfying these needs include lack of up-to-date information resources, poor organization of available information, ignorance of the availability of relevant information, and lack of time for searching.

Progress could be made if up-to-date information relevant to the need was rapidly available at the point of care. Current electronic resources such MEDLINE, full-text journal articles, decision support systems, and clinical Web sites begin to address users information needs. However, these resources have the disadvantage of focusing on single functional areas such as decision support, literature searching, or Web exploration. Users whose information needs bridge these functional areas need to access each system separately and integrate the results for themselves. More desirable are applications that are designed to meet broad information needs and that integrate for the user information from many sources.

In this paper, we describe MedWeaver, a World Wide Web (WWW) application that integrates functions from a decision support application, a literature searching system, and a clinical Web search system. We also discuss the advantages and disadvantages of this integration approach.

BACKGROUND

The goal of integrating biomedical information for clinicians is not a new one. In fact, the National Library of Medicine's (NLM) Unified Medical Language System (UMLS) was developed with this as one of its main goals [3, 4]. Investigators have used the UMLS as the centerpiece of information-integration models [5], resource selection systems [6, 7], and systems that retrieve bibliographic or decision support advice triggered by data in the electronic medical record [8, 9]. Investigators in the fields of database theory [10] and artificial intelligence [11] have also proposed models that support data and information integration.

MedWeaver differs from previous systems in that it integrates functions from three stand-alone applications—DXplain, WebMedline, and CliniWeb—and uses the UMLS for vocabulary translation. As background for understanding MedWeaver's architecture and features, each of these systems is briefly described below.

DXplain

DXplain is a diagnostic decision support system developed at Massachusetts General Hospital [12]. DXplain can generate a differential diagnosis for a list of clinical finding, explain why a particular disease is triggered by a set of findings, or provide textbook information on a disease or finding.

DXplain stores medical knowledge as disease profiles. A profile consists of a set of clinical findings found in the disease and weights for how frequently the finding is found, how strong the finding evokes the disease, and how consequential is the finding. DXplain uses these weights to generate a list of possible diagnoses using a pseudo-probabilistic algorithm and then displays the highest ranked diagnoses divided into "common diseases" and "rare or very rare diseases."

DXplain contains information on more than 2000 diseases, 4700 clinical findings, and 65,000 interrelationships. In stand-alone form, DXplain is

---

* Current affilitation: Department of Health Evaluation Sciences, University of Virginia. For more information on MedWeaver visit http://www.med.virginia.edu/~wmd4n/

available on floppy disk, via Telnet, or via the WWW [13] and is used by more than 200 individuals and institutions in the US. DXplain also has an application program interface (API) that enables remote applications such as MedWeaver to use DXplain's functions.

WebMedline

WebMedline is a Common Gateway Interface (CGI) application developed at Stanford University that facilitates searching of the medical literature via a standard Web browser [5, 14]. WebMedline uses the Hypertext Markup Language (HTML) to display both bibliographic data from MELVYL MEDLINE, the University of California's implementation of the MEDLINE database [15], and hypertext links to corresponding full-text articles on the Internet.

In stand-alone mode, users submit queries via HTML forms and WebMedline modules transform user input into legal MELVYL MEDLINE search statements, retrieve data from MEDLINE, find links to full-text articles on the Internet by querying a local link database, and mark up the results in HTML. WebMedline also has a URL-based API that MedWeaver and other remote applications can use to retrieve MEDLINE data.

WebMedline has been in daily use at the University of California-San Francisco and Stanford University since February 1995. In two years, over 400,000 sessions have been logged. WebMedline provides links to the journals *Science* and the *Journal of Biological Chemistry*, the literature-review publication *ACP Journal Club*, and guidelines from the National Institutes of Health. WebMedline has also been licensed to Ovid Technologies, Inc. and has been incorporated into their Ovid Web Gateway product.

CliniWeb

CliniWeb is a retrieval system developed at Oregon Health Sciences University to help health practitioners find useful medical information on the WWW [16, 17]. Human indexers trained in medicine explore the Web, select quality medical resources, assign to them appropriate controlled-vocabulary terms from the Medical Subject Headings (MeSH), and place them in a database. MeSH term assignment is facilitated by SAPHIRE [18], a concept-matching system that suggests MeSH terms for free-text terms entered by the indexer.

In stand-alone mode, clinicians can explore CliniWeb in two ways. First, they can browse the MeSH hierarchy by disease or anatomical site. For instance, they can find Internet resources on "heart attack" by traversing the disease hierarchy Cardiovascular Disease/ Heart Disease/ Myocardial Ischemia/ Myocardial Infarction and exploring associated Internet links. Second, they can enter terms into the CliniWeb search interface and retrieve matching MeSH terms ranked by likelihood. As with the indexing step, CliniWeb uses SAPHIRE to map users' free-text entries to MeSH. Users then select the desired MeSH term and receive a list of clinically useful URLs. CliniWeb also has an API that allows remote applications such as MedWeaver to retrieve CliniWeb information.

Currently, CliniWeb contains over 10,000 clinically-useful Internet sites is used approximately 10,000 times per month.

UMLS Metathesaurus

The UMLS Metathesaurus is one of four knowledge sources in the NLM's UMLS project [3, 4]. It is a collection of medical vocabularies tied together by the concepts they share. The Metathesaurus preserves the names, meanings, hierarchical contexts, attributes, and inter-term relationships present in its source vocabularies; adds certain basic information to each concept; and establishes new relationships between terms from different source vocabularies. The Metathesaurus also includes the names of selected databases in which the concept appears, and information on the co-occurrence of concepts in MEDLINE.

The UMLS Metathesaurus is available on an experimental basis from the NLM and is provided on CD-ROM or over the Internet via a command-line, API, or Web interface. MedWeaver uses the UMLS API interface to interact with the Internet-based UMLS server.

METHODS

While the three stand-alone systems described above may be useful to clinicians, they are limited because they perform only a specific task—diagnosis of a patient, search of the literature, or search of the Web. We developed MedWeaver, an application built using a model of clinical query management described elsewhere [5], to demonstrate how functions from the three systems can be combined. The result is a decision support system that performs assisted searches of the medical literature and directs users to useful Internet sites.

MedWeaver Architecture

MedWeaver is a Common Gateway Interface (CGI) application that resides at Stanford University, but gathers information from systems in Portland, Oregon (CliniWeb); Boston, Massachusetts (DXplain); and Bethesda, Maryland (UMLS). MedWeaver is made up of several modules (Figure 1) that receive requests from the user (Interface Manager), compose appropriate queries (Query Formulator), retrieve information from various sources (Retrieval Manager), and integrate and display the results for the user (Interface Manager). MedWeaver is written in the Practical Extraction and Reporting Language (Perl) and runs on UNIX servers.

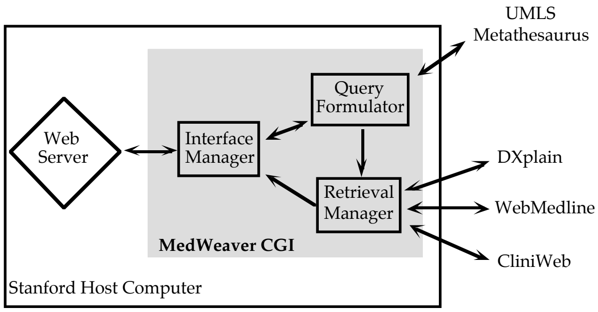

Figure 1. MedWeaver architecture. Users interact with MedWeaver using a standard Web browser. The Web server accepts user requests and hands them off to the MedWeaver CGI. MedWeaver's modules decode users' requests (Interface Manager), formulate appropriate queries using UMLS translation services (Query Formulator), retrieve information from remote systems (Retrieval Manager), and integrate and display the results (Interface Manager).

Query Formulator

The Query Formulator's task is to translate the request of the user to a legal query of the remote information system. The query formulation process is straightforward if the user's need can be fulfilled by a single function call to one system, as would occur if the user asked for the differential diagnosis for a set of findings (a single DXplain function). However, if the user's information need bridges the functionality of two or more systems (e.g., retrieve recent MEDLINE citations on the treatment of a disease listed by DXplain), the query formulation process is more complicated because the vocabulary of one system needs to be translated in the vocabulary of the other.

MedWeaver uses the UMLS Metathesaurus to translate among the vocabularies of DXplain (DXplain vocabulary), WebMedline (MeSH), and CliniWeb (MeSH). MedWeaver first queries the Metathesaurus for the concept code that corresponds to the term in the base vocabulary and then requests the closest matching Metathesaurus entry from the target vocabulary.

For example, if a user wishes to perform a MEDLINE search on the treatment of the DXplain disease, the Query Formulator would first request the Metathesaurus concept code for the DXplain term and then check to see if there is a MeSH term associated with the same concept code. If so, the resulting MeSH term is used by the Query Formulator to create a query statement for WebMedline. If the Metathesaurus concept does not contain a MeSH entry, then MedWeaver requests broader concepts for the root term and then extracts the best matching MeSH term.

Retrieval Manager

The Retrieval Manager's job is to manage the retrieval of information from the information sources. At the network level, the Retrieval Manager establishes and maintains connections with remote hosts using agreed upon protocols (Sockets, Telnet, or HTTP). On the application level, the Retrieval Manager performs user authentication, query submission, and data retrieval. To do this successfully, the Retrieval Manager must know characteristic of the remote system such as interaction style, output formats, and error handling. For instance, the Retrieval Manager knows that to retrieve from DXplain a newline-separated list of diagnoses ranked by likelihood it should package the query in the following format: *getDDx("password", "finding1", "finding2", "finding3").*

Interface Manager

The Interface Manager handles interaction with the user both by processing incoming requests and by integrating results and displaying them to the user. On the input side, the Interface Manager converts user requests from hypertext links or HTML forms into attribute-value pairs and passes them to the Query Formulator. On the output side, the Interface Manager extracts desired data from the Retrieval Manager's output, establishes which functions to provide in the output, and then marks up the selected output in HTML.

Embedded in the Interface Manager is an implicit model of clinicians' information needs and how they can be anticipated and satisfied given the available resources. This model is use to determine which functions are provided at each step in the interaction with the user.

RESULTS

A typical MedWeaver session begins with the user entering clinical findings with the goal of receiving a differential diagnosis. MedWeaver uses DXplain functions to map the entered clinical terms to DXplain controlled vocabulary terms and then to produce a ranked list of possible diagnoses from the list of DXplain terms (Figure 2). MedWeaver can then display for any diagnosis on the list: (1) a summary of that disease (a DXplain function), (2) an explanation as to why the diagnosis appears on list (a DXplain function), (3) an assisted search of the medical literature (a WebMedline function - Figure 3), or (4) a

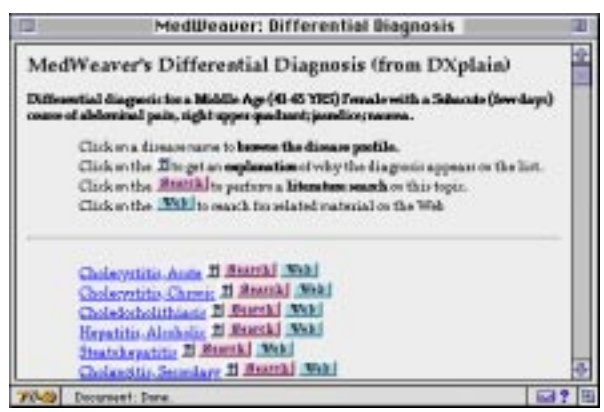

Figure 2. Diagnoses generated by DXplain. MedWeaver uses a DXplain function to retrieve a list of possible diagnoses for the clinical findings "middle-aged female", "subacute (few days)", "abdominal pain, right upper quadrant", "jaundice", and "nausea." From this page, users can retrieve a disease profile, view an explanation, perform an assisted search of the medical literature, or retrieve a list of related Web resources.

list of clinically relevant Internet sites (a CliniWeb function - Figure 4).

To perform an assisted search of MEDLINE, the Query Formulator retrieves from the Metathesaurus the MeSH term that is conceptually closest to the triggering DXplain disease name. It then retrieves for the target MeSH term its concept definition, a list of co-occurring MeSH terms, and its associated MeSH subheadings. The Interface Manager uses this information to produce an HTML form that contains the concept definition, subclasses of information about the disease created by the developer, and co-occurring terms (Figure 3). The user then selects the desired information and submits the HTML form. MedWeaver processes the form, composes a legal WebMedline query, retrieves citations from MEDLINE with links to their full-text equivalents on the Internet, and displays the results in hypertext.

In the Query Formulation step, MedWeaver composes a Boolean search statement consisting of the central MeSH terms, co-occurring terms that were selected, and other modifiers such as language English and journal subset Abridged Index Medicus (AIM). Information classes that were chosen are mapped to

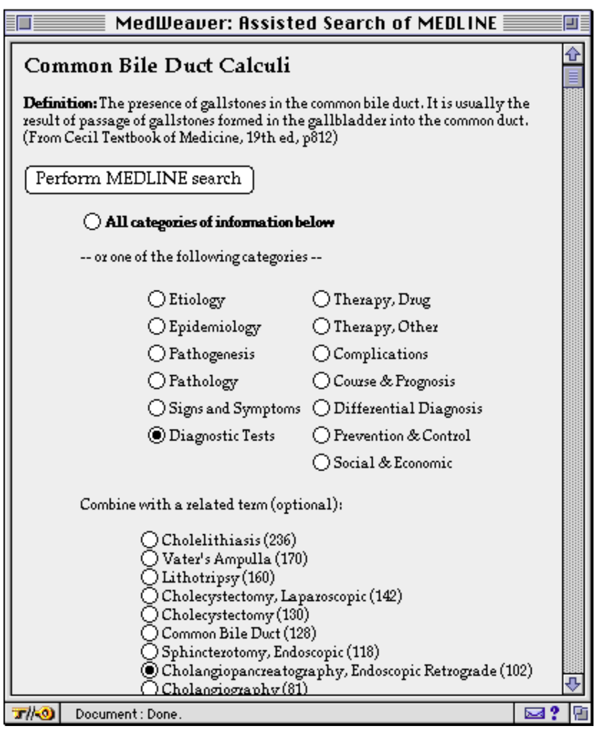

Figure 3. An assisted search of the medical literature. MedWeaver first uses the UMLS Metathesaurus to find the closest MeSH term for the DXplain disease name "Choledocholithiasis." Using this MeSH term ("Common Bile Duct Calculi"), it then retrieves from the Metathesaurus both the term definition and the terms that co-occur in the MEDLINE database. MedWeaver displays this information giving users the ability to limit searches to discrete classes of information such as "diagnostic tests" or "prevention and control." Once a user has specified the desired classes of information and co-occurring terms, MedWeaver generates a sophisticated query of the MEDLINE database and retrieves bibliographic citations and full-text links from WebMedline.

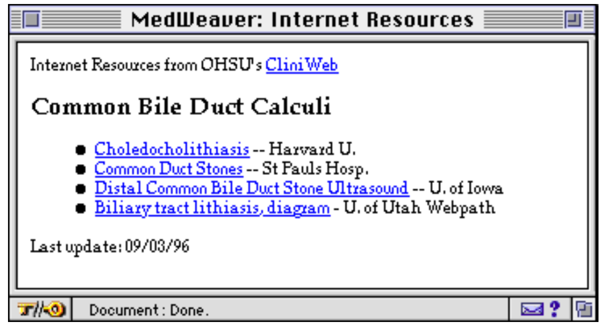

Figure 4. List of clinically useful Internet sites provided by CliniWeb. MedWeaver finds the closest MeSH term for the DXplain disease "Choledocholithiasis," requests from CliniWeb URLs indexed with the specific MeSH term, and then displays the URLs for the user.

related MeSH subheadings. For instance, the "diagnostic tests" choice is mapped to the subheadings "Blood" and "Radiography".

As an example, if a user viewing the screen show in Figure 3 chose "Diagnostic tests" and the co-occurring term "Cholangiopancreatography, Endoscopic Retrograde," the resulting Boolean search statement would be "(Common Bile Duct Calculi/blood OR Common Bile Duct Calculi/Radiography) AND Cholangiopancreatography, Endoscopic Retrograde/ AND Language English AND subset AIM."

To retrieve clinically useful Web URLs from CliniWeb, MedWeaver first identifies the MeSH term conceptually closest to the triggering DXplain name, using the Metathesaurus in a manner similar to what was described for WebMedline. Next the Retrieval Manager requests and retrieves via the CliniWeb API database records that match the given MeSH term. Finally, the Interface Manager extracts from each CliniWeb record the site name, site description, and URL and creates an HTML page of all the results (Figure 4).

DISCUSSION

MedWeaver demonstrates how disparate, widely-distributed information systems can be integrated to produce a new application targeted to meet clinicians' information needs. Because MedWeaver is at present only a prototype, we cannot draw firm conclusions about the utility of this approach. However, we can comment on the potential advantages and disadvantages of the approach.

One of the greatest advantages of creating integrated applications that draw from diverse information sources is that it allows needs-based as opposed to source-based design. In the past, developers created applications by taking an existing information source (e.g., MEDLINE) or by creating a new source (e.g., DXplain) and then developing functionality that maximizes the potential of that individual source. Now, with advent of network-accessible information servers and integration technology, developers can design systems around the information needs of users and then use integration technology to retrieve the needed content and deliver the content at the appropriate point during the interaction with the user.

Another advantage of this approach is that content creation and maintenance are distributed across institutions. Individual organizations have limited resources and talent for producing and maintaining medical content and so are unlikely to produce the vast libraries of information that may be required to meet users information needs. Therefore, tying together distributed content using integration technology is a more parsimonious approach to maintaining content yet meeting users diverse information needs.

Some of the advantages of this approach can also be viewed as disadvantages. For instance, developers of integrated applications must rely on the quality and currency of other's content, quality of wide-area networks, and reliability of information-source APIs. Standards for information exchange, such as the Z39.50 protocol [19], could help make information transfer more standardized and reliable but are at the moment complex and unable to accommodate the wide variety of medical information sources.

There are known limitations of this approach, some of which will be addressed in future work. First, information about how to connect, authenticate, interact, and retrieve data from information sources is embedded in the MedWeaver application. It would be preferred to externalize this information, or better yet retrieve this information from a centralized source such as the UMLS Information Sources Map [7].

Second, the model of information needs that MedWeaver uses to determine which functions to provide at each step of the user interaction is implicit and is embedded in the application. A preferred approach would be to develop an external, sharable model of information needs that could be maintained separately and could be used to determine what type of information a user might need at different points during system interaction.

Lastly, the algorithm for vocabulary translation is insufficient when the base vocabulary term and the target vocabulary term do not share the same concept. Needed is an improved algorithm that consistently finds the closest match when the participating vocabularies contain different levels of term specificity.

CONCLUSION

MedWeaver demonstrates the potential for meeting clinicians' information needs with systems that integrate information from multiple sources. Further research and evaluation studies are necessary to determine the ultimate value of such systems.

Acknowledgments

This work was conducted with the support of the National Library of Medicine under grant LM-07033. Computing facilities were provided by the CAMIS Resource, LM-05305.

References

1. Gorman PN. Information needs of physicians. J Am Soc Inf Sci 1995;46(10):729-36.

2. Smith R. What clinical information do doctors need? BMJ 1996;313(7064):1062-1068.

3. Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. Methods Inf Med 1993;32(4):281-91.

4. Humphreys BL, Lindberg DA. The UMLS project: making the conceptual connection between users and the information they need. Bull Med Libr Assoc 1993;81(2):170-7.

5. Detmer WM, Shortliffe EH. A model of clinical query management that supports integration of biomedical information over the World Wide Web. Proc Annu Symp Comput Appl Med Care 1995:898-902.

6. Miller PL, Frawley SJ, Wright L, Roderer NK, Powsne rSM. Lessons learned from a pilot implementation of the UMLS information sources map. . J Am Med Inform Assoc 1995;2(2):102-15.

7. Rodgers RPC. Automated retrieval from multiple disparate information sources: The World Wide Web and the NLM`s Sourcerer project. J Am Soc Inf Sci 1995;46(10):755-64.

8. Cimino JJ, Johnson SB, Aguirre A, Roderer N, Clayton PD. The MEDLINE Button. Proc Annu Symp Comput Appl Med Care 1992:81-5.

9. Elhanan G, Socratous SA, J. CJ. Integrating DXplain into a clinical information system using the World Wide Web. Proc AMIA Annu Fall Symp 1996:348-52.

10. Wiederhold G. Mediators in the architecture of future information systems. Computer 1992;25:38-49.

11. Fikes R, Engelmore R, Farquhar A, Pratt W. Network-based information brokers. Knowledge Systems Laboratory, Stanford University, 1995:

12. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain. An evolving diagnostic decision-support system. JAMA 1987;258(1):67-74.

13. http://lcs-dxplain.mgh.harvard.edu/lcshome/dxplain.htm.

14. http://www.med.stanford.edu/cgi-bin/med-lookup.

15. Horres MM, Starr SS, Renford BL. MELVYL MEDLINE: a library services perspective. Bull Med Libr Assoc 1991;79(3):309-20.

16. Hersh WR, Brown KE, Donohoe LC, Campbell EM, Horacek AE. CliniWeb: Managing clinical information on the World Wide Web. J Am Med Inform Assoc 1996;3(4):273-280.

17. http://www.ohsu.edu/cliniweb/.

18. Hersh WR, Hickam DH. Information retrieval in medicine: the SAPHIRE experience. J Am Soc Info Sci 1995;46:743-7.

19. National Information Standards Organization. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. NISO Press, 1996.