DXplain on the Internet

G. Octo Barnett, MD, Kathleen T. Famiglietti, Richard J. Kim,
Edward P. Hoffer, MD, Mitchell J. Feldman, MD
Harvard Medical School
Laboratory of Computer Science
Massachusetts General Hospital
Boston, Massachusetts

DXplain, a computer-based medical education, reference and decision support system has been used by thousands of physicians and medical students on stand-alone systems and over communications networks. For the past two years, we have made DXplain available over the Internet in order to provide DXplain's knowledge and analytical capabilities as a resource to other applications within Massachusetts General Hospital (MGH) and at outside institutions. We describe and provide the user experience with two different protocols through which users can access DXplain through the World Wide Web (WWW). The first allows the user to have direct interaction with all the functionality of DXplain where the MGH server controls the interaction and the mode of presentation. In the second mode, the MGH server provides the DXplain functionality as a series of services, which can be called independently by the user application program.

INTRODUCTION

DXplain[1], a computer-based medical education, reference and decision support system, was developed by the MGH Laboratory of Computer Science (LCS) fifteen years ago and has been used by thousands of physicians and medical students on stand-alone systems, and over communications networks. DXplain has the characteristics of an electronic medical textbook, a medical reference system and a decision support tool. In the role of a medical textbook, DXplain can provide a comprehensive description and selected references for over 2,000 different diseases, emphasizing the signs and symptoms that occur in each disease, the etiology, the pathology, and the prognosis. As a decision support tool, DXplain uses its knowledge base of the crude probabilities of over 5,000 clinical manifestations associated with over 2,000 different diseases. The system uses an interactive format to collect clinical information and makes use of a modified form of Bayesian logic to produce a ranked list of diagnoses which may be associated with at least some of the clinical manifestations. DXplain uses this same knowledge base and logic to list other findings that, if present, would support a particular disease, and also lists what findings entered by the user are not usually found in a particular disease. DXplain has been the subject of a careful evaluation by Berner[2], who analyzed the performance of DXplain (and several other computer-based diagnostic systems) on 105 cases and found that the DXplain knowledge base contained 91% of the "correct diagnosis" (as determined by a panel of experts), and that in 69% of the time, DXplain included the "correct diagnosis" in the top 20 suggested diagnoses.

Five years ago, the DXplain team began the conversion of the interface from a character-based, command oriented approach to a graphical user interface which would facilitate data entry and access to the vast amount of information in the knowledge base. At the same time we wanted to respond to the increasing demand to provide DXplain's knowledge and analytical capabilities as a resource to other applications within MGH and at outside institutions. In addition, many issues related to the distribution of DXplain remained a constant concern: platform dependence, installation support, database updates and integrity, copyright infringement, and user isolation with the resulting difficulty of obtaining feedback from users.

METHODS

The growing availability of Internet access offered an attractive option for DXplain distribution. Two years ago we began making DXplain available over the Internet on a licensed, no-charge trial basis. Both the DXplain users and the developers maintain platform independence and revised versions can be available immediately. Over 70 medical institutions and over 500 physicians now access DXplain over the Internet.

We encountered a number of design considerations for the Web environment: 1) reducing the amount of dialog as DXplain attempts to match the user's narrative text entries; 2) maintaining a session state even though the interaction is a series of separate connections with the server; 3) selecting the appropriate browser techniques for presenting information and simplifying user input; 4) taking advantage of the browser's capabilities to offer alternative protocols for the user to explore the knowledge base; and 5) providing the appropriate degree of security by restricting access to licensed users or licensed institutions.

We have developed and support two different protocols through which users can access the DXplain knowledge base and algorithms through the WWW. The first allows the user to access DXplain and take advantage of all of its functionality as though the program were executing on a local computer. In this mode, the user accesses DXplain using a standard Internet browser with the DXplain functions provided through a Common Gateway Interface (CGI) application program developed at, and run on, a server at the LCS. The DXplain interface is accessed through a Uniform Resource Locator (URL) and the LCS server controls the interaction and the mode of presentation. The HTML form structure is used to dynamically generate pages. Through this interface, the user can enter sets of clinical findings to generate a suggested disease list, ask DXplain to explain a disease in the context of the findings set, ask DXplain to suggest other findings that could narrow the disease list, request a description and references for any disease in the knowledge base and request a list of diseases associated with any finding. This protocol is a connectionless one with no state maintained automatically. Each new request for information requires a connection to the server. In order to provide the context of state and maintain a user "session", the application passes the appropriate variables with each connect including an assigned session identifier used for logging and security purposes.

In the second mode, the MGH server provides the DXplain functionality as a series of services which can be called independently by a user application program. In this mode, the user application controls the style and format of the presentation. We provide a DXplain API (application programming interface) for execution of individual DXplain functions using hypertext transfer protocol (HTTP) to connect to our server over the Internet. Each connect transmits a password, a function name and parameters related to the particular function invoked. As with the first protocol described, there is no state maintained and every call requires a new connection to the server.

A list of examples of the functions and the purpose of each follows; a complete example is provided for the "Match Finding Name" function.

- **Match Finding Name**

  Purpose: Match a text string to the DXplain vocabulary of findings (signs, symptoms, lab data).

  Example:

  Parameters passed to DXplain:
  password: access code assigned by LCS
  function: "matchToFindings"
  finding: text string to be matched against the DXplain vocabulary of finding names

  Returned by DXplain:
  a) If text is an exact match to a DXplain finding name or synonym, the DXplain finding name is returned preceded with "@@".
  Finding text "fever" returns
  "@@FEVER"
  Finding text "joint pain" returns
  "@@ARTHRALGIA"

  b) If text is not an exact match to a DXplain finding name or synonym, a list of DXplain findings is returned. Finding text "arthitis" returns a list including:
  "ARTHRITIS"
  "PULMONARY ARTERITIS"
  "OSTEOARTHRITIS"
  .
  .
  .

- **Match Disease Name**

  Purpose: Match a text string to the DXplain vocabulary of disease names.

- **Get Disease Info**
  Purpose: Retrieve DXplain Disease Information text and references for a specific disease.

- **Get Finding Info**
  Purpose: Retrieve lists of diseases, which are associated with a particular finding.

- **Generate Disease List**
  Purpose: Produce a ranked list of DXplain diseases which might explain or be associated with the list of findings passed.

- **Generate Disease List with Focus**
  Purpose: Allow the user to specify certain findings to be key elements to the case. All disease descriptions, which do not contain the FOCUS findings, will be excluded from the pool of diseases to be considered for scoring.

- **Explain Disease**
  Purpose: Produce an explanation of how/why a particular diagnosis has been scored in DXplain and which findings are considered to be high information items for this diagnosis. The function will return the findings which support the diagnosis, those unexplained, and findings which would make this diagnosis more likely.

- **Get Question User List**
  Purpose: Retrieve list of findings whose presence or absence can clarify the diagnostic possibilities. List includes an explanation as to why each finding may be useful.

RESULTS

We report on the experience we have had with the two access protocols with users both within our own institution and in other institutions.

Starting in 1996, we made our WWW-based version of DXplain available on a licensed, no-cost, trial basis to both medical institutions and individual physicians. In calendar year 1997, there were 476 3-4 yr. medical student users with 671 sessions (315 of which were identified as real cases - versus using DXplain as a teaching exercise on a hypothetical case); 165 residents with 316 sessions (225 real cases); and 315 staff physicians with 743 sessions (474 real cases). A questionnaire was administered by the system at regular intervals so that we could collect user feedback as an evaluation of the system. In response to the item: "Did DXplain understand the findings you entered", the response was "Always or frequently" for 58% of the 3-4 yr. medical students, 63% of the residents, and 76% of the staff physicians. In response to the item: "Did DXplain's Disease List suggest plausible diseases which you had not considered", the response was "Always or Frequently" for 64% of the 3-4 yr. students, 44% of the residents, and 64% of the staff physicians. In response to the item: "As a result of using DXplain in a real case, did you pursue other diagnostic possibilities or change your management", the response was “Always or Frequently” for 30% of the 3-4 yr. students, 25% of the residents, and 36% of the staff physicians. The questionnaire responses support our belief that using a decision support resource such as DXplain can benefit the intended audience by providing decision support and clinically useful information.

The second mode of accessing DXplain allows other applications access to specific information in the DXplain knowledge base and algorithms via connections using HTTP protocol. Each connection transmits a password, a function name and parameters to the DXplain server. The server program returns results appropriate to the particular function called and characteristics of the parameters passed.

During 1997, approximately 6,200 direct function calls were made to DXplain’s server. Over 35% of the calls were made to generate a suggested disease list from findings collected and submitted by the outside application; 25% of the calls were made to request DXplain disease information and references for a specific disease.

For the past year, we have used DXplain as an integrated educational resource with a laboratory test reporting system made available over an MGH Intranet. The Referring Physician Information Access (RPIA)[3] is a Web-based clinical information system that allows physicians not on the MGH campus to access the hospital's clinical information repository through a dial-in connection to a password-controlled server. The report accessed by the physician is an HTML page where abnormal laboratory test results are flagged, and accompanied by a tag indicating whether the value is high or low. When such an abnormal test result is included in the DXplain database, the RPIA application places a button beside the laboratory test abnormality. If the user clicks on this button, the RPIA application uses an internal table to transform the particular lab test name and value into an appropriate DXplain term and then uses the DXplain API to make a function call to the "Get Finding Info" function. The call returns a list of diseases which may be associated with the particular laboratory abnormality. This list is tagged by the RPIA application and presented as an HTML page to the user. The list is an ordered list in that the diseases are separated into categories according to how common is the disease, and how frequently the laboratory abnormality is found in the particular disease. The user can click on any of these diseases to request that DXplain provide a description of the particular disease.

We now are deploying a Web-based application which provides access to the COSTAR[4] medical record which is used in primary care clinics at MGH by over 30 staff physicians and approximately 70 residents. In the prototype version of this application[5] we are providing links to practice guidelines developed by MGH and to DXplain for both diagnoses and abnormal laboratory tests. These links are accessed by clicking buttons next to the diagnosis or laboratory test on the COSTAR page and provide full access using Web technology to all the DXplain functions.

Collaborators from outside of MGH are now accessing the DXplain knowledge base and capabilities to enhance their own applications. In the past two years, we have had several provocative collaborations in providing DXplain resources to other institutions where the DXplain server provided services over the Internet but did not control the user interaction. For example, the staff of the Department of Medical Informatics of Columbia University have developed an experimental linkage of integrating DXplain interpretations with their laboratory reporting module.[6] In this collaboration, the MGH DXplain server is providing access to the DXplain knowledge base, but Columbia is controlling the user interaction. The Columbia experience demonstrated the importance that the remote institution must assume of assuring vocabulary matching. The Columbia investigators took advantage of their powerful MED vocabulary server to transform a MED abnormal laboratory finding into an acceptable DXplain term (e.g., "elevated blood glucose" must be transformed into the DXplain term: "hyperglycemia").

Detmer uses DXplain in a very interesting experiment (MedWeaver[7]) in the integration of functions from diverse, distributed systems. His experiment involves the integration of decision support in literature searching using DXplain and UMLS. One of the capabilities of MedWeaver is to accept a set of user-entered clinical manifestations, pass these to DXplain, retrieve a set of DXplain generated diagnostic hypotheses, and then use the UMLS Metathesaurus to translate from DXplain diagnoses to MeSH terms for literature searching. The information about how to connect, authenticate, interact, and retrieve data from DXplain is embedded in the MedWeaver application. The MedWeaver experiment also encountered vocabulary matching issues when the base vocabulary term and the target vocabulary term did not share the same concept. Detmer argues that this experiment illustrates the advantage of creating integrated applications in that it allows need-based as opposed to source-based design. His work also demonstrates the advantage of allowing content creation and maintenance to be distributed across institutions.

DISCUSSION

Providing DXplain access to a wide audience and diverse institutional settings has been an exciting venture. The Internet and use of HTTP protocol and HTML have provided us with an efficient, well-defined, documented and scalable environment for widespread dissemination of our DXplain interface and knowledge base. The technology offers us platform independence and minimizes issues of client side software installation and support. Knowledge base updates and system enhancements are released automatically to all users.

We are convinced that both of our access protocols are important to serve the needs and interests of the user community. We have learned much from the imagination that other institutions have demonstrated in the integration of DXplain knowledge resources into local applications, and believe that there are numerous other opportunities for other institutions to take advantage of using the DXplain knowledge base and diagnostic algorithms as resource services. In our experience with our own Web-based DXplain application, we encourage user feedback. This interaction permitted by our interface is a highly useful feature of the system. We have found that the ease of point and click encourages users to enter complaints, kudos and suggestions. These are reviewed daily by one of the physicians on the project staff, and when necessary to understand the comment, a print-out of the interaction is available from extensive logs we keep of each user "session". In some cases the user comments have led to substantive changes in the knowledge base, including the addition of new diseases and changes in disease descriptions. In others, it has helped us clarify instructions. At times, comments reflect user misunderstanding of the system, and allow us to offer suggestions for more effective use of DXplain. In addition to user comments, we have found that user surveys, such as the results mentioned in the previous section, provide important feedback.

CONCLUSION

We have found that both forms of WWW access to DXplain - the interface we provide as well as the API to specific DXplain knowledge services - offer exciting opportunities to make the resources of this system more widely available on an international scale. In contrast to the distribution of DXplain as a stand-alone system or via dial-in networks, WWW distribution allows a more controlled and reliable distribution with minimal support overhead. In addition, we can both maintain effective communication with our users and enable integration of the DXplain capabilities with a diverse set of applications at other institutions.

For a demonstration and current information on the availability of DXplain, go to:

http://www.lcs.mgh.harvard.edu

ACKNOWLEDGMENTS

This work was supported in part by NLM research grant LM05854. An equipment grant was provided by Hewlett Packard.

REFERENCES

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

2. Berner ES, Webster GD, Shugerman AA et al. Performance of four computer-based diagnostic systems. N Engl J Med, 1994;330: 1792-1996.

3. Kittredge R, Rabanni U, Melanson F, Barnett O. Experiences in deployment of a web-based CIS for referring physicians. Proceedings of the 1997 AMIA Fall Symposium, 1997 Oct 25-29, Nashville, TN. p. 320-324.

4. Barnett GO. The application of computer-based medical-record systems in ambulatory care. N Engl J Med, 1984;310:1643-1650.

5. Rabbani U, Morgan M, Barnett GO. A COSTAR interface using WWW technology. Submitted for presentation at the 1998 AMIA Fall Symposium, 1998 Nov 7-11, Orlando, FL.

6. Elhanan G, Socratous SA, Cimino JJ. Integrating DXplain into a clinical information system using the world wide web. Proceedings of the 1997 AMIA Fall Symposium, 1997 Oct 25-29, Nashville, TN. p. 348-352.

7. Detmer WM, Barnett GO, Hersh WR. MedWeaver: integrating decision support, literature searching, and web exploration using the UMLS metathesaurus. Proceedings of the 1997 AMIA Fall Symposium, 1997 Oct 25-29, Nashville, TN. p. 490-494.