QUEST--Query Environment for Science Teaching

Ben Shneiderman[1], Azriel Rosenfeld[2], Gary Marchionini[3], William G. Holliday[4], Glenn Ricart[5], Christos Faloutsos[6], and Judith P. Dick[7]

University of Maryland, College Park

[1] Professor of Computer Science

[2] Research Professor and Director, Center for Automation Research

[3] Associate Professor of Library and Information Services

[4] Professor of Curriculum and Instruction

[5] Director, Computer Science Center

[6] Associate Professor of Computer Science

[7] Assistant Professor of Library and Information Services, dick@glue.umd.edu,

FAX: 301-314-9145, telephone: 301-405-2048

Abstract

QUery Environment for Science Teaching (QUEST) is a proposed digital library implementation consisting of a set of research projects dealing with data capture and organization, content analysis, information seeking and visual interfaces.The QUEST team includes a large number of renown technical collaborators and prominent source collaborators, as well as a significant number of contributors in the University of Maryland, the central, co-ordinating agency.

A large collection of multidisciplinary materials in visual and textual formats, made accessible to us by our source collaborators, will be organized to allow integrated access by users from the science education community, that is elementary school through college level teachers. QUEST will be structured so as to provide seamless access to widespread resources on disparate subjects. We intend to provide first-rate subject analysis and representation in order to provide ready access.

QUEST will be accessible nationally by means of Mosaic. We propose to provide highly sophisticated querying, browsing and information investigation facilities which will handle integrated textual and visual materials without difficulty. They will be augmented by online reference and referral services, immediately accessible by the user. QUEST will provide a comprehensive information resource for science education accessible through a dynamic, visual user interface.

Keywords: Interfaces, VLDB, databases, IR, science education.

Introduction

The University of Maryland, and a strong team of partners, propose a coordinated set of research projects in the implementation of a digital library, which will support a nationally-oriented testbed effort in science

education. Our project is called Query Environment for Science Teaching, or QUEST. We will accumulate and make available the most comprehensive collection of science education materials ever offered in any form and make it searchable in a dynamic visual user environment.

The resource will contain multidisciplinary information, both textual and visual, derived from many multimedia formats. The software tools to build, access, and extend the collection will be developed by the QUEST Team at the University of Maryland, with the collaboration of its technical and testbed partners located in the Washington, DC area. We will provide QUEST users with the capacity to browse a thousand times faster than they can in traditional libraries and to retrieve wanted information with ease.

Project development will involve the representation of science education knowledge and the building of mechanisms to support easy access. Multiple levels of representation will be available for users of varying experience, interest levels, and skill. Robust querying interfaces will support the work of our widely diverse testbed user population. Our dynamic querying techniques will allow direct manipulation users of animated visual information displays. (Shneiderman in press).

All the elements of QUEST will be subjected to rigorous testing. Evaluations will be based on the responses of real users in science education environments. Their judgements will determine the direction of efforts at improvement as work progresses.

The partners

Six faculty members from Maryland's Computer Science Department will work with collaborators in seven University units including the Colleges of Library and Information Services, and of Education, the Center for Automation Research, the Institute for Systems Research, the Computer Science Center, the Center for Space Data and Information Systems, and the University of Maryland Libraries.

We have, as well, joined with eleven partner organizations in order to guarantee high-quality research, driven by leading-edge applications of the digital library concept. Our partners include UNISYS (research, testbed and equipment), DEC (equipment) and Bell Atlantic (network development). In addition, the National Science Teachers Association (NSTA), the American Association for the Advancement of Science (AAAS), the American Institute of Physics (AIP), Library of Congress, National Library of Medicine, Prince George's County (MD) Public Schools, Montgomery County (MD) Public Schools and the Center for Renewable Energy and Sustainable Technology have joined in our effort to provide a leading-edge application of the digital library concept.

The research

We propose to attack the problem of building a digital library with a research program that advances the techniques of page decomposition and content analysis, while constructing new tools for data capture and data formatting. Using the collection resulting from the acquisition work, our querying team will conduct research on problems in the retrieval of non-textual materials and index searching for the location of distributed materials. We shall also investigate improved methods of tailoring views resulting from searches and of graphical browsing and querying. Our position is that optimal quality reference services are essential to effectiveness of a digital library. We shall provide manual references services to start, and develop automatic services of increasing sophistication as we proceed. It is our goal that, ultimately, only very difficult queries will require human intervention, and then it will be very high caliber personal aid integrated with the automated services.

The testbed--science education

Using a large collection of materials (in both digital and traditional forms) from the American Institute of Physics (AIP), the American Association for the Advancement of Science (AAAS), the National Science Teachers Association (NSTA), and major national libraries (LoC, NLM, NASA), we intend to create an invaluable resource. QUEST will be essential to practising science teachers of elementary through college-level students. Such teachers are working toward motivating and teaching ALL students to become scientifically literate, and they are striving to achieve the educational goal of AAAS and NSTA. We have the enthusiastic cooperation of AAAS and NSTA, the two organizations dominating national efforts at improvement in science education. We include under our umbrella as well two local school systems which truly reflect urban America. By the end of the project, materials will be easily accessible to science teachers throughout the nation by means of a Mosaic interface.

The challenge

The emergence of vast electronic libraries was anticipated in 1945 by Vannevar Bush in his bold "Memex" concept of a desktop hypertext environment with linked documents (Bush 1945). Elaborate digital libraries were described by Licklider (1965) in "Libraries of the Future", based on the work of The Council on Library Resources, Inc. which had been established by the Ford Foundation in 1956.

". . . the trouble stems from what we may call the 'passiveness' of the printed page. When information is stored in books, there is no practical way to transfer the information from the store to the user without physically moving the book or the reader or both. Moreover, there is no way to determine prescribed functions of descriptively specified informational arguments within the books without asking the reader to carry out all the necessary operations." (fn Licklider 1965, 5.) Now we can conceive of means appropriate to respond to this challenge!

QUEST

QUEST (QUery Environment for Science Teaching) is a design proposed to meet the challenge of these visionary thinkers. While we cannot claim to be able to build a perfect digital library, within four years, even with concerted effort by our multidisciplinary team, we strongly believe we can produce an exceptionally valuable resource for national development. Furthermore, we intend that our ten research projects, ambitious testbed development, and efficient implementation demonstrate the viability of the digital library concept. Finally, we anticipate that testing by active users and rigorous evaluation will help us identify fruitful paths for future development.

QUEST's design emerged from a novel concept of libraries of the future, and the communities they will serve along with a penetrating theory of visual information seeking. We are dedicated to universal access and to usage by diverse groups. Patrons will be able to contribute to QUEST's content and will be able to facilitate retrieval by sharing their techniques and experiences. We anticipate that QUEST will contain massive multimedia resources. Moreover, it will be readily accessible by the many and varied networks in the National Information Infrastructure.

QUEST's contents will be analyzed in detail and indexed largely automatically. Search will be facilitated by every means possible. Advanced user interfaces will be developed in order to promote the investigation of serious subject matter in greater depth than state-of-art retrieval systems allow. Users who achieve mastery over the advanced user interfaces we develop will have a decided advantage in advancing their knowledge.

In order to realize our goals, we have formulated ten interlocking QUEST research projects: four deal with building and six deal with querying. These projects take advantage of established research expertise and reach beyond current paradigms. Our choice of research projects was guided by a desire to produce foundational results with widespread applicability. To validate the outcome of these research projects, we will build an extensive testbed and evaluate its efficacy using our own user community.

QUEST will meet the critical library needs of students enrolled in the major testbed site--Prince George's Public Schools. Prince George's County, Maryland is located next to the Northeast sector of Washington, DC. and has one of the most diverse populations in the country. The populace is both multicultural to a high degree, and of widely divergent socio-economic strata. With this group, we anticipate providing trained science teachers with the means to engage their students in problem-solving projects beyond the capabilities of their present school resources. As a result, we expect to witness a large-scale increase in motivation for further work in scientific discovery. QUEST will provide both networking and query subsystems to support such investigations.

Partnerships

We believe the digital library will payoff handsomely in better science education. For that reason, we have focused on serving the needs of science teachers in the elementary grades through college.

Our secondary users will include the science students themselves, adult distance learners, the continuing education population, journalists, policy makers, and other adults with interest in scientific information. Recognized professional societies, including such partners as the American Association for the Advancement of Science (AAAS) and the National Science Teachers Association (NSTA), will guide the selection of materials and the evaluation of the QUEST testbed design. Science teachers, our primary users, will be able to, for example, access climatic data from NASA documents and then perhaps choose to associate it with population and hunger statistics from NSTA documents, while using interdisciplinary tools from AAAS's science curriculum Project 2061. Students using such a source are challenged to learn science by analyzing data and attacking realistic world problems. In so doing, they develop skill in reasoning and in constructing arguments as their education progress.

Our industrial partners will provide a practical perspective in the research and testbed aspects, including access to the latest hardware and software technology. They also provide destinations for technology transfer that could lead to the commercial operation of digital libraries. We are pleased to have a major commitment of research effort from UNISYS Corporation. As well, we welcome contributions of equipment, material, and services from UNISYS, Digital Equipment Corporation and Bell Atlantic of Maryland.

The American Institute of Physics (AIP) is the foremost publisher of peer-reviewed, scholarly journals in physics and physics education. It has close ties to the American Association of Physics Teachers (AAPT), and is involved in pursuing digital publishing. AIP will contribute contemporary materials of interest to science educators and analyze issues relating to the compensation of authors and editors for access to materials.

Our research community is additionally enriched by partnerships with the Library of Congress and the National Library of Medicine, which will provide demonstration sites. We continue our long and close cooperation with NASA and will enjoy access access to their facility. We have the advantage of being located in close proximity to all these renowned associates, and so are able to interact with them quickly and easily.

Electronic libraries

Libraries in the electronic age will retain some traditional roles related to the acquisition and preservation of resources, but increasingly they are taking on new roles. A necessity, as the number and diversity of patrons increase, is the support of information access and reference services. As libraries grew tremendously during the current century, patrons increasingly became aware of needing help in finding their way around the enlarged collections and growing facilities. Librarians came to provide a wide range of personal information services as well as technical expertise in organizing sources by cataloging and classifying them. Then, industry experienced a dramatic increase in its' need for knowledge and librarians began to anticipate users' wants. They alerted patrons to the receipt of new material before it was requested.

Alerting services of varying degrees of sophistication were begun. Profiles of users' interests were used to select and disseminate copies of current title pages, abstracts of journal articles, news stories, new technical data, standards of various kinds, legal enactments and decisions, scientific and engineering data, legal memoranda for litigation support and competitive intelligence briefing books for executives, not to mention up-to-the-minute news on financial affairs, securities, corporate matters and so on. Librarians became less involved with custodial duties in book collections and much more concerned with the analysis of content information as their patrons needs increased in volume and sophistication.

In the electronic age, when the potential for communication within communities of users is even greater, the role of libraries seems inevitably linked to their capacity to provide better developed in-depth services at great speed, thus supporting their burgeoning information requirements and promoting collaboration among users.

Users of digital libraries in the future may be amused when they comprehend that users in our time were required to travel significant distances to reach a library; that a book could be read by only one person at a time and that misplaced books were often as good as lost in large collections. On the other hand, unless we take care now, future users may be frustrated by being confronted with an overwhelming flood of unorganized data of indifferent value. They may be alienated by the necessity of using superficial search mechanisms, which are imprecise and unforgiving. If interfaces are unduly complex or rigid and uncommunicative and information seekers find themselves isolated at their terminals without access to a considerate helper, we may lose an opportunity to make real progress in advancing scientific knowledge. Thus our vision of a digital library is user centered and applies technology to allow users to effectively manage digital information.

Research projects

QUEST combines research efforts in the following areas:

* automation of data capture, interpretation

and organization

* content analysis, and representation

* query, browsing and data investigation

* online reference and referral services

* visual user interfaces, highly adaptable for

* individual use in searching and preparing

materials.

Our first challenge in creating QUEST is to capture the information and to format it appropriately. The approach we propose is semiautomatic and interactive, incorporating an incremental use of scanning technology. Documents will be decomposed into segments and the segments then labeled and clustered. Salient sections will be further analyzed to provide structural and semantic information. The marked-up segments will afford a base for developing generic models of a wide range of document types, leading eventually to a degree of automatic understanding. (Doermann 1992, in press)

Users will be able to query QUEST sources at variant levels. Automatic term analysis and frequency data will be used in conjunction with case and clausal analyses in order to provide a robust, hybrid indexing unit. The goal throughout will be to support sophisticated and comprehensible querying at whatever level needed. We anticipate the necessity of including some conceptual representations as well. Semantic information is to be derived from both the decomposition and indexing operations. (Doermann, in press; Dick 1992) Parallel performance of tasks will result in the production of keywords, phrases, and some primitive conceptual representations, all of which will combine to support the retrieval function. In addition, when traditional descriptors such as classification numbers and subject headings are available, they will be used in combination, and their potential exploited to the fullest in enhancing document representations.

QUEST integrates graphs, maps, and images along with text. We have adopted two unique approaches to image understanding and indexing. The first approach will leverage any written description and will also apply scene recovery techniques to distinguish foreground from background, and to extract structural attributes from still and motion pictures. The second approach will be to use whole-image and sub-pattern feature matching functions to select candidate images, enabling users to focus their searches through relevance feedback. Thus, once a user, for example, an astronomy teacher, has located an image of interest, say a NASA photograph of Saturn, the user will be able to ask for "similar" images, in order to retrieve more photographs of Saturn or of other planets as well, as he decides. (Faloutsos 1985, 1987, 1988)

QUEST will provide multiple levels of query support. A superindex to resources and sites will give users rapid access to a variety of sources while minimizing storage space. System response will be expedited even in a wide area network environment. Users will also be able to browse cross-referenced indexes. Client-server architectures will support dynamic downloading of material and caching. Users will be able to create their own custom views, some of which may be added as valuable extensions of the QUEST corpus.

QUEST will provide two types of reference services to users. First, a frequently asked question (FAQ) service will be developed and maintained. Functions for seeding, updating, and weeding questions will be developed and evaluated by means of studies of user populations. Second, a human consultant network will be implemented in order to deal with user problems which require human interpretation and analysis. This network will put users directly in multi-channel contact with consultants and all interactions will be captured, analyzed, and classified for integration into QUEST.

The main entry point for QUEST will be a dynamic query interface. A visual overview is presented to the user and s/he can rapidly filter content by adjusting sliders, buttons, or other easy-to-use devices. Finally, the user can make selections in order to obtain details at will. In developing QUEST's dynamic interface, the main challenges will be mapping information needs to cogent visual scales, developing a tool kit for dynamic query interfaces and developing custom screen displays. (Shneiderman 1983, in press) The final challenge will be to thoroughly test the interface's performance by means of users evaluations.

QUEST will allow information seekers to browse up and down through levels of representation by applying

zooming and panning mechanisms. Levels of representation appropriate for text and images will be identified and mapped onto the dynamic query interface. Mechanisms for continuous zooming and panning will be developed and tested in user studies. (Marchionini 1992, 1994) These mechanisms will allow users to move gracefully from high level storage structures, such as directories, to specific document segments with the same interface tools.

We will provide advanced graphic user interfaces with dynamic two or three dimensional animations that represent the conceptual space of our libraries. These novel interfaces will allow exploration and sifting of very large information resources. QUEST will enable first-time users to succeed in answering basic queries. It will provide occasional users with comprehensible overviews, in order to orient them correctly, and will offer tools for use in navigation. Regular users will have powerful tools to facilitate serious research in greater depth than the others. Finally, experienced, frequent users will have flexibility to fashion customized procedures for themselves or the communities they serve.

Each aspect of our work will be validated by testing. We will measure our progress regularly, and develop metrics to enable us to compare alternative approaches, and to document our successes and our difficulties.

The testbed for science education

In order to demonstrate the viability of our research projects we will accumulate and make available the most comprehensive collection of science education materials offered in any form. We will also prepare the software tools to build, access, and extend the collection. We will provide training to instructors in approaches to teaching the use of the new online library. QUEST will be available for teaching support in public schools. We shall monitor its utilization and assess the impact of the digital library use on science teaching, teachers and students. The American Institute for Physics, our publishing partner, along with our technology partners, UNISYS, Digital Equipment, and Bell Atlantic, will look for ways to commercialize ideas validated in the testbed.

In cooperation with our professional society partners, especially AAAS and NSTA, we have identified key resources to be included in QUEST. At the core is a set of materials carefully chosen from leading science publishers in the Project 2061 library. We have identified eight organizational areas of the K-16 science education digital library as priorities for teachers:

1. Organizations as Resources

2. Science Activities and Projects

3. Conventional Publications

4. Physical and Social Science Data Banks

5. Assessments

6. National and State Curriculum Programs

7. Grants and Other Funding Sources

8. Connections, Conversations, and

Conferences

In cooperation with our industrial partners, and with support from our government library partners, we will build QUEST from multimedia sources including textual resources of all kinds and a variety of CD-ROMs, videotapes, software, scientific datasets, electronic graphics and images. This rich environment will provide a resource for collaborative learning events including: team projects among students, electronic discussions with students in distant states, and discussions with professional scientists or mentors otherwise unaccessible. With prominent national partners, we will advance the cause of universal access and diverse usage of current, high-quality, scientific and technical information from a range of disciplines.

Conclusion

QUEST is designed to bring us into the world of libraries of the future as envisioned by Licklider and the scientists of the sixties. It is an attempt to provide an ultra modern facility for the most part automated, with only an occasional help from a human intervener. The content of the QUEST collection is highly diverse, combining visual and textual information from scientific resources of all kinds modeled in a way suitable to the teaching of the students from Kindergarten through university, and suitable for adults seeking information for professional and private uses.

The latest in database technology is combined with the most forward-looking interface developments in order to demonstrate that users of all levels and types can be served in a manner which facilitates their work. QUEST's speed and precision will allow information seekers to make the transition from the libraries of today to the library envisioned by Bush and Licklider. Moreover, the user will find herself or himself comfortably using a dynamic interface which allows full attention to be given to the information rather than to the system.

This NSF proposal has stimulated a remarkable interdisciplinary collaboration here at the University of Maryland, allying disparate research groups in an effort to create QUEST, the science library of the future. We are motivated to make an impact on computer science research, library design, science education, commercial practice, and government policy. Our steering committee and the associated advisory committees of leading researchers and practitioners, will help ensure that our decisions are effectively made and applied. The research projects give promise of adding to our knowledge about database systems and highly interactive interfaces which will be generalizable to other domains. We see QUEST as an opportunity to rapidly bring the benefits of advanced technology to science education.

Acknowledgments

The QUEST Team

UNIVERSITY OF MARYLAND

Computer Science Department

Database Research Group

College of Library and Information Services

College of Education

Center for Automation Research

Human-Computer Interaction Laboratory

Document Understanding Group

Institute for Systems Research

University Libraries

Computer Science Center

Center for Space Data and Information Systems

TECHNICAL COLLABORATORS

UNISYS--research, testbed and equipment

DEC--equipment

Bell Atlantic--network development

SOURCE COLLABORATORS

National Science Teachers Association (NSTA)

American Association for the Advancement of Science (AAAS)

American Institute of Physics (AIP)

Library of Congress (LC)

National Library of Medicine (NLM)

Prince George's County Public Schools

Montgomery County Public Schools

Center for Renewable Energy and Sustainable Technology

References

[1] Bush, Vannevar, "As we may think." Atlantic monthly, 176, 101--108, July 1945.

[2] Christodoulakis, S. and Faloutsos, C.,"Design considerations for a message file server". IEEE Transactions on Software Engineering 10 (984) 201--210.

[3] Dick, J.P., A conceptual case-relation representation of text for intelligent retrieval. Technical Report CSRI-265, Computer Systems Research Institute, University of Toronto (1992).

[4] Doermann, D.S. and Rosenfeld, A., "Recovery of temporal information from static images of handwriting". Proceedings of Computer Vision and Pattern Recognition (1992) 162--168; International Journal of Computer Vision, in press.

[5] Doermann, D.S., Varma, V., and Rosenfeld, A.,R "Instrument grasp A model and its effects on handwritten strokes". Pattern Recognition, in press.

[6] Faloutsos, C., "Access methods for text". ACM Computing Surveys 17 (1985) 49--74.

[7] Faloutsos, C. and Chan, R., "Fast text access methods for optical and large magnetic disks: Designs and performance comparison. "Proceedings of the 14th International Conference on VLDB, Long Beach, CA (August 1988) 280--293.

[8] Faloutsos, C. and Christodoulakis, S., "Optimal signature extraction and information loss". ACM Transactions on Database Systems 12 (1987) 395--428.

[9] Faloutsos, C., Equitz, W., Flickner, M., Niblack, W., Petkovic, D., and Baer, R., "Efficient and effective querying image content". Journal of Intelligent Information Systems, in press.

[10] Licklider, J.C.R. Libraries of the future. Cambridge MA: MIT Press, 1965.

[11] Marchionini, G., "Interfaces for end-user information seeking". Journal of the American Society for Information Science 43 (1992) 156--163.

[12] Marchionini, G. and Crane, G., "Evaluating hypermedia and learning: Methods and results from the Perseus Project". ACM Transactions on Information Systems, (1994) 5-34.

[13] Shneiderman, B., "Direct manipulation: A step beyond programming languages". IEEE Computer 16(8) (August 1983) 57--69.

[14] Shneiderman, B., "Dynamic queries for visual information seeking". IEEE Software, in press.