

Information Management and Delivery via a Networked Scientific Repository

*Laura M. Bartolo, Collection Development Librarian and ALCOM Principal Investigator and Robert Casson, Graduate Research Assistant, School of Library and Information Science
Kent State University Libraries and Media Services*

Abstract

This paper describes on the development of a Web-based information management and communication tool to facilitate work remotely among research scientists, government agencies, and industrial partners. Implications of the information retrieval database for the ALCOM/NIST project is discussed in relations to digital libraries applications and the Internet. The paper addresses the development of a domain-specific vocabulary for polymer liquid crystal and phase separation research as well as the impact assessment of the database on the research project. As a vital part of research teams, librarians can impact the creation of new information tools and market their expertise and services.

Overview

My paper focuses on the collaborative working relation among government, academic, and industrial partners to test the robustness of a prototype networked repository for managing and delivering scientific knowledge rapidly in Ohio.

Scientific enterprise represents a pivotal area in research collaboration because of its prolific use of emerging information technology and its impact on technology transfer. The Digital Libraries Initiative (a recent funding focus of the National Science Foundation) and the National Information Infrastructure (a major direction of the National Institute of Standards and Technology) lay the foundation for the construction of an electronic scientific network of distributed repositories where multimedia objects -- data, program codes, simulations, formal and informal literature -- can be stored, located and disseminated rapidly to the industrial sector.

Information Management Project Objectives

We want to expand and test an existing Web-based scientific repository - the pilot ALCOM/NIST Repository on phase separation in liquid crystals - to examine:

- who creates and contributes information?
- what types and formats of scientific information are generated?
- how is the information located by project participants?
- what are relevant terms to index and search information across the sciences?
- what information is selected by university researchers? by government agencies? by industrial partners?
- what are project participants' information delivery preferences?

The ALCOM/NIST Repository has begun to store, organize and distribute information generated by its participants on phase separation in liquid crystals via the Internet. The scientific knowledge in the Repository will be a standard suite of theoretical models and computer programs for modeling polymer-dispersed and polymer-stabilized liquid crystalline materials used in the optical display industry. The pilot ALCOM/NIST Repository is available at URL <http://cpip.kent.edu/PSP> (<http://cpip.kent.edu/PSP>) (login and password available upon request).

Collaborative Partners

The pilot ALCOM/NIST Repository on phase separation in liquid crystals is a working association involving the NSF Center ALCOM, NIST Center for Theoretical and Computational Materials Science (CTCMS) and Ohio Department of Development. The pilot database was included among NIST CTCMS efforts in the area of information infrastructure. SLIS has been limited to the extent to which it has been able to participate in the Repository.

ALCOM

One of 25 Science and Technology Centers established by the National Science Foundation nationwide, ALCOM is located in Ohio and is a consortium of Kent State University, Case Western Reserve University, and the University of Akron, with the Liquid Crystal Institute at Kent State University as its hub. The Center melds basic studies of liquid crystals with applied science and has lead to the technological advances and new applications, such as display tablets, optical shutters, variable transmission windows, projection display devices, improvements for active matrix displays, and flexible displays. In 1989 the National Science Foundation established Science and Technology Centers (STCs) to help maintain the position of the United States as a world leader in science and technology. The National Science Foundation and Congress have called upon universities and research institutes to play an active role with U.S. industrial partners. The STCs' provide an important link with the industrial community to meet the changing needs of society, science and economic development.

The author is an ALCOM Principal Investigator in the area of information management.

NIST

NIST's mission as a government agency is to promote more rapid technological innovation by creating linkages between industry and university scientists.

Academe, Government & Industrial Participants

The scientists working with the Repository are physicists, chemists, and mathematicians from:

- universities: NSF Center for Advanced Liquid Crystalline Optical Materials (ALCOM - Kent State University, Case Western Reserve University, University of Akron), New York University, University of California at Los Angeles and University of Toronto;
- government agencies: National Institute of Standards and Technology (NIST) the Ohio Department of Development & Wright Patterson Air Force Base, and
- industries: IBM, dpiS (Xerox), General Motors, and Raychem Corporation.

Information Scientists

Kent State University's School of Library and Information Science (SLIS) have begun working with the author, an ALCOM PI and information scientist on the ALCOM/NIST Repository, and the directors of the project on phase separation in liquid crystals, to build a networked information retrieval and communication system. The collaboration of SLIS and ALCOM is a strong union bridging the latest in information technology with a premier basic and applied research center in Ohio.

Economic Vitality of Ohio

The prototype scientific repository has strong economic implications. Information management and technology serve as an essential link among academics, government, and industry. A distributed information repository and delivery system will improve cross disciplinary collaboration in all areas of scientific research. The ALCOM/NIST Repository model can be applied to other areas, such as the biomedical field, which encounter similar retrieval problems. The Repository will provide a workable utility to gather and archive data, search and deliver innovative research to the marketplace.

Rapidly transferring innovative ideas to the commercial sector can lead to mutually rewarding research and licensing agreements. ALCOM research has led to the development of startup industry in Ohio. Companies such as Crystalloid and Kent Display Systems (KDS) will benefit from an electronic repository linking information, such as newly registered patents. Though recently begun, the pilot ALCOM/NIST Repository has received considerable interest from ALCOM industrial partners in and outside of Ohio, such as IBM, dpiS (Xerox), General Motors, and Raychem Corporation.

A further economic benefit will be training future information specialists in Ohio to meet the information needs of science and industry. ALCOM information specialists and SLIS researchers are working together to extend the initial work on the Repository. Kent's School of Library and Information Science is Ohio's only school of Information Science, attracting students throughout the state and beyond. Through involvement with projects, such as the Repository, information specialists from SLIS will graduate with a strong information and computing background ready to work with Ohio's scientific and industrial sectors.

Scientific Repositories and the Internet

Recent publishing practices show that the fastest, most economical method of distributing timely documents is through the Internet. Scientific communities need to construct networked repositories with detailed classification of information and vocabulary transfer across the scientific disciplines. These databases could deliver customized scientific results, enabling researchers in academe or industry to readily search and solve problems by correlating information across these databases on the Internet.

Prototype: The ALCOM/NIST Repository

How the Repository Functions

The ALCOM/NIST Repository is a test site with implications reaching beyond liquid crystal research. The Repository's electronic delivery of scientific knowledge among its government, academic, and industrial participants will act as a model information management system for other future research repositories. Participants will contribute original research to the database from their desktop through web-based submitting forms. The database will automatically index, abstract and update project information submitted by researchers, such as the contents of scientific presentations, codes, experimental data, and simulation results. Electronic whiteboard and videoconferencing will be incorporated for drawing, writing, and

conferencing to approximate face-to-face interactions among researchers. The web server will be constructed to support a distributed, open system framework to share research information as well as modeling, simulation, and visualization applications.

Communicating Scientific Information

The repository will provide communication mechanisms for rapid site-to-site collaboration in finding and delivering research results among project participants. Researchers will profile their individual research interests to reflect continuing and new information needs. Based upon these preferences, participants will be presented a personalized database when they visit the Web site. An automated SDI (Selective Dissemination of Information) current awareness mechanism via e-mail based upon the the profiles will alert individuals to recent submissions. An electronic newsletter will cover recent project developments in abstract form. If readers want more information, they will link to the full text. If they have specific questions, they will connect directly to research teams in the project.

Indexing and Retrieving Cross-Disciplinary Information

Scientific collaboration requires information across the sciences to solve research problems. A new and significant contribution of the ALCOM/NIST Repository will be the development of a cross-discipline thesaurus for polymer liquid crystal research. The methodology for developing the cross-disciplinary vocabulary is not limited to liquid crystal research and can be applied to other interdisciplinary science projects. Experts in computer and information science recognize indexing as the primary source of information retrieval problems. Because of semantic vocabulary differences in scientific disciplines, the need to search indexed collections of information with cross-disciplinary vocabulary will increasingly grow.

For example, scientific literature in polymer liquid crystal research can be found in separate databases for physics, chemistry, and allied subject disciplines, such as mathematics. However a detailed thesaurus specific to polymer liquid crystals and phase separation research does not currently exist. A cross-discipline vocabulary will be constructed to ensure an efficient information indexing and retrieval system. This thesaurus will not be vocabulary for polymer liquid crystals research used solely in physics, chemistry or mathematics. Rather it will identify and link similar terms developed by scientists in these three disciplines so that researchers can correlate scientific knowledge across the sciences. A three-pronged measurement of word frequencies, term co-occurrence, and term discriminator values will be used to find the optimum level of specificity and precision in the vocabulary and indexing. A full-text retrieval software will be used to automatically calculate these measurements.

Why Scientific Repositories will Enhance Collaboration and Technology Transfer

The research of the ALCOM/NIST Repository will generate a new model and tools to support all types of interdisciplinary scientific information, retrieval and dissemination. An automatic classification scheme will define the subject hierarchy and identify terms commonly occurring together in documents in the digital collection. Enhanced searching capabilities will interactively map the researcher to alternative subject terms to significantly enhance retrieval effectiveness. Specialized terms, once identified within the Repository, will link researchers to corresponding terms in different subject disciplines. Development of the ALCOM/NIST Repository will follow standards to support the Internet's information-organization. The Repository will reach out to the scientific community of expert specialists analyzing emerging research as well as deliver to industrial partners advanced materials for the development of new or improved products.

Research Methodology

The ALCOM/NIST Repository will construct a workable prototype to manage and deliver scientific information. We will test the Repository to determine the following areas:

- Unobtrusive tracking mechanisms will be used to measure how the researchers acquire, organize, maintain and retrieve project-generated and project-related information throughout the life of the ALCOM/NIST repository.
- Tracking usage data such as database contributions, SDI current awareness, database search engine, and transaction logs will document how a user was directed to scientific information archived in the Repository and what was selected as relevant information. The Repository will observe how scientific distributed digital work impacts technology transfer.
- In building the phase separation-liquid crystal cross-discipline vocabulary, a Liquid Crystal Research Thesaurus developed prior to this current project and the INSPEC Classification and INSPEC Thesaurus will serve as the base to extend the vocabulary.

Further details of the transaction and tracking mechanisms used in the Repository are detailed in an appendix to this paper prepared by Robert Casson, Graduate Assistant, School of Library and Information Science, Kent State University.

Challenges for Collaborative, Multidisciplinary Projects

The primary mission of universities is to increase society's knowledge and understanding through rigorous scholarly investigation. Working closely with government and industrial partners can complement academic research activities. To ensure the interactions among the three groups are successful, it is important to be aware of factors, such as intellectual property rights, publications, and conflicts of interest, which may adversely affect collaboration. Obligations regarding intellectual property need to consider if the collaborating partners cooperatively developed an invention or idea or if work was solely invented by university researchers. Patent agreements have generally established well defined criteria for determining who are inventors or co-inventors. Publishing timely research results can mutually benefit the three parties. First, the peer review process by the scientific community helps to insure accuracy of the research work. Second, preserving the rights of intellectual property will be maintained by promptly filing patentable ideas. Commercial ventures can carry potentially conflict of interest concerns for academics and government partners. Full disclosure of sponsorship at the beginning of new research can reduce the possibility of conflicts. Similarly, the peer review system of the scientific community has proven to be an effective mechanism for curbing the existence of biased or misrepresented finds.

External Funding

Start up money for the pilot project was initially provided by ALCOM through the Ohio Department of Development and the Center for Theoretical and Computational Materials Science at NIST. Scientific repositories are a major focus of the Digital Libraries Initiative (National Science Foundation) and the National Information Infrastructure (National Institute of Standards and Technology). Both of these national efforts emphasize broad collections of digital scientific literature. The federal government through NSF and NIST sponsor information technology initiatives especially to promote scientific collaboration and technology transfer. The industrial community, such as telecommunication companies, publishing companies and computer companies have an interest in and will benefit from the development of scientific repositories.

Resources Required for the Collaboration

The ALCOM/NIST Repository completes its work with the following resources:

- Computer equipment and software to enhance the pilot repository.
- Release time for faculty members, staff and research assistant help.
- Travel expenses to give presentations at professional and scholarly conferences in information and computer science, and the respective sciences involved

Evaluation and further research implications

In the ALCOM/NIST Repository project we will conduct observations of scientific work, technology transfer, and how they intersect with the use of distributed, digital information. Individual and group interviews will be conducted with a range of testbed users from the government, universities, and industrial participants. Experiments involving the effects of economic models and charging mechanisms could be investigated in the future.

Acknowledgements

Financial support for this project was provided by the National Science Foundation Center for Advanced Liquid Crystalline Optical Materials (ALCOM) and the Center for Theoretical and Computational Materials Science at the National Institute of Standards and Technology (NIST) Award 60NANB6D0175

References

- Baldwin, R.G. and Autin, A. E. Toward Greater Understanding of Faculty Research Collaboration. *Review of Higher Education* 19: 45-70 (1995).
- Beard, Kate. Digital Spatial Libraries: A Context for Engineering and Library Collaboration. *Information Technology & Libraries* 14: 79-85 (1995).
- Citron, Paul. Research Interactions Between Industry and Academia: A Corporate Perspective. *Physiologist* 39: 81, 90-92 (1996).
- Fox, M. F. and C.A. Faver. Independence and Cooperation in Research. The Motivations and Costs of Collaboration. *Journal of Higher Education* 55: 347-59 (1984).
- Jones, S., M. Gafford, S. Robertson, M. Hancock-Beaulier, J. Secker, S. Walker. Interactive thesaurus navigation: Intelligent Rules OK? *Journal of the American Society for Information Science* 46: 52-59 (1995).
- Paepcke, A., S. B. Cousins, H. Garcia-Molina, S. W. Hassan, S. P. Ketchpel, M. Roscheisen, & T. Winograd. Using Distributed Objects for Digital Library Interoperability. *Computer* 29: 61-68 (1996).
- Pao, M. L. Global and Local Collaborators: A Study of Scientific Collaboration. *Information Processing & Management* 28: 99-109 (1992)
- Salton, G (1989) *Automatic Text Processing*. Reading, MA: Addison-Wesley (1989).
- Schatz, B., H. Chen. Building Large-Scale Digital Libraries. *Computer* 29: 22-26 (1996).
- Schatz, B., W. H. Mischo, T. W. Cole, J. B. Hardin, A. P. Bishop, & H. Chen. Federating Diverse Collections of Scientific Literature. *Computer* 29:28-36 (1996).

Shoval, P. Principles, Procedures and Rules in an Expert System for Information Retrieval. *Information Processing and Management* 21: 475-487 (1985).

Stiles, H. E. The Association Factor in Information Retrieval. *Journal of the Association of Computing Machinery*, 8: 271-279 (1961).

Taubes, G. Indexing the Internet. *Science* 269:1,354-1,356 (1995).

Appendix: Transaction and Tracking Mechanisms

Robert Casson, Graduate Assistant

School of Library and Information Science, Kent State University

Using one or more tracking mechanisms, we seek to improve our understanding of the flow of information and to develop new tools to better serve our users.

The very nature of the World Wide Web serves as both a benefit and an obstacle to a project like ours. The current speed and relative ease-of-use of the Internet greatly reduce the impact of geographical boundaries, but this increased efficiency comes with some sacrifices. Transactions on the web involve a client (or browser) and an HTTP server. Browsers, like Netscape or Internet Explorer, send requests to servers, specifying a file name and a desired action. This request establishes a connection between the client and server, often a simple request to view an HTML file. Once the server has responded, the connection between client and server is terminated. This is referred to as a "stateless connection." In this model, the client side offers a necessary, but minimal, amount of information to the HTTP server: the object requested and the client's host IP address. This information is stored on the server in log files. No state information is saved about the user. State information is any information that is maintained regarding ongoing interaction between a client and a server. Each request is recorded in these log files as unique and unrelated to any other selection on other pages. This process of request response termination improves the efficiency of the machines involved, and increases the number of clients that can be supported simultaneously. An example of a stateful connection is a library's online catalog; a user's requests are remembered by the OPAC, and this state information can be modified across searches.

Tracking the movements of visitors from page to page, however, requires the saving of more state information. This is in direct conflict with the current operation of the web. To save any additional state information about a user, a place must be designated to store the data, whether on the client's browser, on the server, or through a cooperative effort between the two. We have discussed some of the limitations of server log files; a stateless server ends connections between requests, saving no information between a single user's selections. Another option is to save state information to the client's browser, and this brings us to the idea of Persistent Client State Cookies. Saving state information could greatly improve the way transactions are carried out on the web, but the idea of an outside server writing to a personal hard drive should be met with some skepticism. Many of the security issues will be addressed as our description of cookies continues.

Cookies are small pieces of state information that are created by server-side programs called CGI scripts. These programs are activated when a client requests an object from a particular path on a server. The response, along with the cookie, arrives through the existing connection, which is then closed. The new cookie, and its state information, will now be stored on the client. If a client requests another page in the particular path, the cookie value will be returned to the server, which can recognize the value and perform some operation; in our case, this could aid in the analysis of our web site's information flow.

Cookies have been in existence since Netscape released its Navigator 1.1 browser, but they have just begun to receive wide-scale attention. This is due in most part to browsers now offering a notification when a cookie is being passed to them. A cookie is an object given a name and a value upon its creation. These parameters are established by the server-side CGI script when called into action by a request. Cookies must also include a domain name and a path name. This allows server B to identify its cookies and to retrieve their state information on subsequent visits by the client. This also prevents server A from accessing any cookies created by server B.

A cookie is a very small file (only 4 kilobytes) and cannot be executable. A browser is capable of holding 300 total cookies, with a limit of 20 cookies per server or domain. Some web users are worried that cookies are able to gather user information surreptitiously. The only way for a server to set a cookie containing any personal information is to explicitly request data from the user and place it in the cookie. Information of this type could be stored on a server, regardless of any cookie use. Early versions of Netscape Navigator 2.0 did allow Javascript to access the client's e-mail address, if entered in the preferences of the browser, but this hole has been fixed for quite some time. Other people are concerned at the possibility of cookies reading the contents of their hard drives, but this is also impossible. In fact, client-side browsers freely offer more information than most other web-based applications, telling servers what operating system and browser you are using. In all, cookies cannot compromise security any more than clients do themselves; it is a good general rule to avoid divulging much information over the web to anyone.

We are exploring several applications of cookies for our web site. Cookies are being used to build "shopping cart" applications for many commercial sites. With these programs, a user can navigate a site and select desired items. The chosen items are placed in the "shopping cart" and are recorded into a list. When the client is through browsing, their items are taken through a "checkout lane" and the transaction is completed. This technique could be directly applied to our web site. Instead of placing compact discs or clothing in the "cookie shopping cart," our researchers could choose specific preprints or program codes and have these delivered to them via electronic mail.

Tracking will be the primary use of cookies. The entire user session cannot be stored in the cookie, due to its small size. A cookie can hold an identifying number, which will track a visitor's movements through the site and store the selections as a single unit, rather than unrelated requests and responses. Once the session has ended, the cookie will expire. Monitoring the information behavior in this way will give clues as to new models of web-based tools. How much and what type of information is being submitted to the central database? What materials are being requested and retrieved? How are these items being used by the participants? All of these questions can be addressed with thoughtful use of cookies.

References

- Catchings, Bill and Mark Van Name. (1996, July 29). Web Security and the Cookie Controversy. *PC Week*. [Online]. Available : <http://www.pcweek.com/archive/1330/pcwk0037.htm>
(<http://www.pcweek.com/archive/1330/pcwk0037.htm>)
- Catchings, Bill and Mark Van Name. (1996, August 5). Have Your Cookies and Beat Them Too. *PC Week*. [Online]. Available: <http://www.pcweek.com/archive/1331/pcwk0037.htm>
(<http://www.pcweek.com/archive/1331/pcwk0037.htm>)

Catchings, Bill and Mark Van Name. (1996, August 18). Putting the Lid on Pandora's Cookie Jar. *PC Week*. [Online]. Available: <http://www.pcweek.com/archive/1333/pcwk0041.htm>
(<http://www.pcweek.com/archive/1333/pcwk0041.htm>)

Humes, Malcolm. (1997, February 11). *Malcolm's Guide to Persistent Cookies Resources*. [Online]. Available: <http://www.emf.net/~mal/cookiesinfo.html> (<http://www.novell.com/>)

Kristol, David M. and Lou Montulli. (1997, February). *RFC 2109: HTTP State Management Mechanism*. [Online]. Available: <http://www.cis.ohio-state.edu/htbin/rfc/rfc2109.html> (<http://www.cis.ohio-state.edu/htbin/rfc/rfc2109.html>)

Netscape Communications Group. (1997). *Persistent Client State HTTP Cookies*. [Online]. Available: http://home.netscape.com/newsref/std/cookie_spec.html (<http://www.novell.com/>)

Stein, Lincoln. (1996, Autumn). CGI Scripts and Cookies. *The Perl Journal*. [Online]. Available: http://orwant.www.media.mit.edu/tpj/programs/Vol_1_Issue_3_CGI/cookie.html
(http://orwant.www.media.mit.edu/tpj/programs/Vol_1_Issue_3_CGI/cookie.html)