Processing and Access Issues for Full-Text Journals

William H. Mischo and Timothy W. Cole

INTRODUCTION

The University of Illinois at Urbana-Champaign (UIUC) was one of six sites awarded a four-year federally funded grant in 1994 under the first phase of the Digital Library Initiative (DLI). The DLI grants, jointly funded by the National Science Foundation (NSF), the Defense Advanced Research Project Agency (DARPA), and the National Aeronautics and Space Administration (NASA), were awarded, in addition to UIUC, to Stanford University, the University of California at Berkeley, Carnegie Mellon University, the University of California at Santa Barbara, and the University of Michigan. A detailed description of the Illinois DLI project, along with links to the other five projects, can be found at http://dli.grainger.uiuc.edu/ and is described in Schatz et al. (1999) and Schatz et al. (1996).

Activities on the $4 million UIUC DLI grant have been carried out by a multi-departmental research team comprised of individuals from the university's Graduate School of Library and Information Science, the University Library, the National Center for Supercomputing Applications (NCSA), and the Department of Computer Science. The project also includes important in-kind contributions in the form of full-text articles in SGML format from a number of professional society publishers and significant equipment and software grants from several companies. The Illinois DLI project includes research, testbed, and evaluation components.

This article will describe the design and implementation of the Illinois DLI testbed. It will focus on the issues addressed, the problems encountered, and the lessons learned in the course of deploying the testbed. This work was carried out by the testbed team headquartered in the Grainger Engineering Library Information Center, a $22 million facility that opened in 1994 and is dedicated to the exploration of emerging information technologies.

The testbed is constructed from source text journal articles in the areas of engineering, physics, and computer science contributed by several professional society publishers. This contributed testbed source material is comprised of full-text articles in SGML format and bit-mapped images of figures of sixty-three journal titles presently containing over 60,000 articles from 1995 forward. The full-text articles for the testbed have been contributed by the American Institute of Physics (AIP), the American Physical Society (APS), the American Society of Civil Engineers (ASCE), the Institute of Electrical and Electronics Engineers Computer Society (IEEE CS), and the Institution of Electrical Engineers (IEE). Several other publisher collections were examined but not included in the testbed.

TESTBED GOALS AND ISSUES

The overarching focus of the DLI testbed team has been on the design, development, and evaluation of mechanisms that can provide effective access to full-text engineering, physics, and computer science journal articles within an Internet environment. The primary goals of the Illinois testbed have been:

1. the construction and testing of a scalable multi-publisher heterogeneous SGML-based full-text testbed employing flexible search and rendering capabilities and offering extensive links to local and remote information resources;
2. the development of procedures for the local processing, indexing, normalization (through use of metadata), retrieval, and rendering of full-text journal literature in marked-up format as provided by contributing publishers;
3. the integration of the testbed (and other full-text resources) into the continuum of information resources offered to end-users by the library system;
4. determining the efficacy of full-text article searching vis-à-vis document surrogate searching, and exploring end-user full-text searching behavior in an attempt to identify user-searching needs; and
5. identifying models for the effective publishing and retrieval of full-text articles within an Internet environment and employing these models in the testbed design and development.

On an overarching level, the project has addressed the issues connected with the migration from a print-based journal collection to an Internet-based model. The testbed team has identified the staff and hardware requirements necessary for the local processing, loading, and retrieval of full-text data. In addition, the team has generated mechanisms for providing access (via links) to testbed materials and other publisher repositories through standard retrieval tools such as Abstracting and Indexing (A & I) services. Both of these approaches focus on defining retrieval mechanisms to optimize user access to full-text journals.

At the onset of the DLI Project, the testbed team faced a number of clearly defined design and development issues. These issues related to collection procurement and utility, the identification of standards for the format of the collection materials, accurate rendering of materials, the development of processing procedures, optimum database search engine retrieval capabilities, and the determination of the appropriate mix of off-the-shelf software versus locally developed code.

All full-text retrieval test systems face limitations relating to the breadth and the depth of the collection. Test systems cannot typically include all the subject-related journals needed to meet the needs of researchers in the covered discipline. Likewise, the typical full-text test collections will not provide the years of coverage necessary to completely meet the needs of researchers. (Interestingly, these two problems have been made more acute by the discrete publisher-based full-text retrieval system model we see today.) The Illinois testbed, comprised of sixty-three journals, has addressed these issues through expanded links from full-text references and by providing simultaneous searching of testbed resources and A & I service databases. The integration of distributed full-text repositories will continue to be addressed within the testbed.

When work on the UIUC DLI project began in 1994, the World Wide Web (WWW) was in a nascent stage. At that time, NCSA's Mosaic 2.0 beta was the browser of choice, the HTML 2.0 standard was still under development, Netscape had yet to release its first Web browser, and Microsoft Windows 3.1 was the standard PC operating system. Early studies of the effectiveness of full-text retrieval were necessarily limited in scope, primarily because of the breadth/depth problem and the lack of figures, mathematics, and tables for the article display (Tenopir & Ro, 1990).

The initial task of the testbed project team was to identify technologies that were both of sufficient maturity to be usable at once and of sufficient potential to evolve over the life of the project. As the project evolved, two clear trends emerged: the WWW has become the standard for text retrieval and display and, as a direct corollary, publishers have taken advantage of emerging Web technologies to establish their own full-text repositories.

TESTBED TECHNOLOGIES: SGML FORMAT AND RENDERING

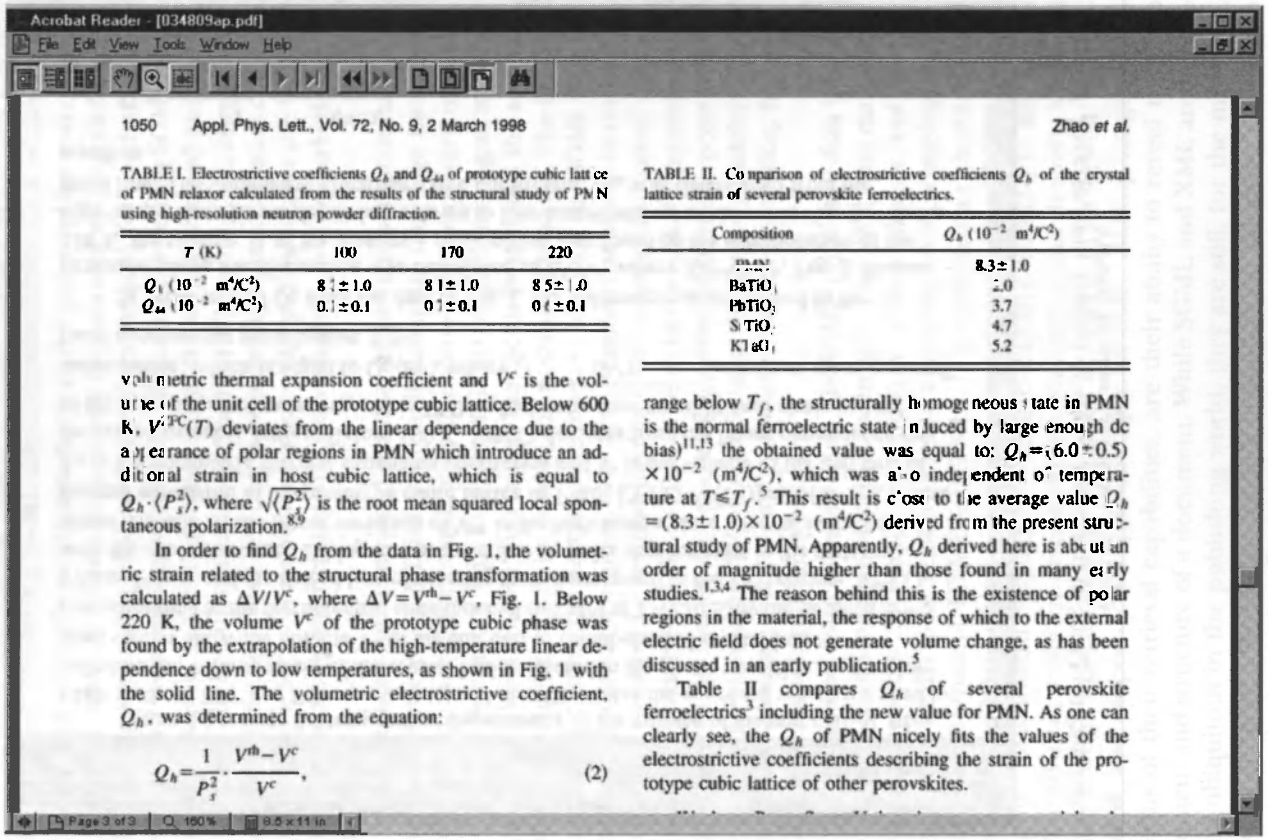
Critical to the ultimate success of the project was the determination of the testbed document format standard. The ideal format would support full-text indexing; high-granularity field-specific search and retrieval; and robust platform-independent rendering. No existing format matched all criteria, and it was immediately obvious that HTML 2.0 fell far short of desired structure and rendering functionality. The Standard Generalized Markup Language (SGML), a nonproprietary international standard, was clearly the best of the formats available for exposing the intellectual content and structure of documents. The Text Encoding Initiative (TEI) (Sperberg-McQueen, 1994) was built around SGML, and pilot full-text journal publishing projects using SGML were then underway at OCLC (Weibel, 1994). However, rendering engines for SGML were limited and required separate executables or plug-ins.

Two other contending document formats were also examined. TeX and LaTeX are well established in the mathematical sciences academic community and support extremely robust rendering of mathematics, but the available authoring and display tools were limited and were largely UNIX-based. Also, exposure of document structure in TeX as used in real-world applications is limited.

PDF, an Adobe-proprietary format, provided the best emulation of the printed page. Adobe Acrobat reader was free and available for multiple platforms. However, PDF lacked (as of 1994) important hyperlink functionality and vital (for our project) cross-collection indexing features. It also was then, and remains today, a primarily appearance-oriented format.

SGML was chosen for the testbed document format standard because it was nonproprietary and inherently best both for indexing and for search and retrieval. This decision was consistent with the publishing world's identification of SGML as the emerging standard for document representation and transmission. While all publishers contributing source materials for the project had experience with the three formats under consideration, it was clear that SGML or SGML with embedded TeX for mathematical equations was the preferred format.

However, the use of SGML in a Web environment presents some formidable challenges. The lack of suitable SGML renderers has hindered the project from the very beginning. It was the hope of the testbed team that, as the technologies evolved, there would be advances in SGML rendering engines, but these improvements have not materialized. To compensate for immediate SGML rendering limitations, several of the publishers provided PDF versions of articles in addition to SGML versions. Figure 1 shows an excerpt of a sample article in PDF format displayed by Adobe Acrobat. Figure 2 shows the same article excerpt in SGML format rendered by the SoftQuad Panorama viewer. Note that the Panorama SGML viewer has problems with the accurate rendering of display mathematics, in particular with kerning operations, fraction bars, radical length, and line breaks. Note also that PDF rendering imitates the published page layout while the SGML rendering results in a less structured and continuous display.


Acrobat Reader - [034809ap.pdf]
File Edit View Tools Window Help
Page 3 of 3
150%
8.5 x 11 in

1050 Appl. Phys. Lett., Vol. 72, No. 9, 2 March 1998 Zhao et al.

TABLE I. Electrostrictive coefficients Q_h and Q_{44} of prototype cubic latt ce of PMN relaxor calculated from the results of the structural study of PM N using high-resolution neutron powder diffraction.

T (K)	100	170	220
Q_h (10^{-2} m^4/C^2)	8.1±1.0	8.1±1.0	8.5±1.0
Q_{44} (10^{-2} m^4/C^2)	0.1±0.1	0.1±0.1	0.1±0.1

TABLE II. Comparison of electrostrictive coefficients Q_h of the crystal lattice strain of several perovskite ferroelectrics.

Composition	Q_h (10^{-2} m^4/C^2)
PMN	8.3±1.0
$BaTiO_3$	2.0
$PbTiO_3$	3.7
$SrTiO_3$	4.7
$KTaO_3$	5.2

volumetric thermal expansion coefficient and V^c is the volume of the unit cell of the prototype cubic lattice. Below 600 K, $V^{FC}(T)$ deviates from the linear dependence due to the appearance of polar regions in PMN which introduce an additional strain in host cubic lattice, which is equal to $Q_h \cdot \langle P_s^2 \rangle$, where $\sqrt{\langle P_s^2 \rangle}$ is the root mean squared local spontaneous polarization.[8,9]

In order to find Q_h from the data in Fig. 1, the volumetric strain related to the structural phase transformation was calculated as $\Delta V/V^c$, where $\Delta V = V^{th} - V^c$, Fig. 1. Below 220 K, the volume V^c of the prototype cubic phase was found by the extrapolation of the high-temperature linear dependence down to low temperatures, as shown in Fig. 1 with the solid line. The volumetric electrostrictive coefficient, Q_h, was determined from the equation:

$$Q_h = \frac{1}{P_s^2} \cdot \frac{V^{th} - V^c}{V^c}, \qquad (2)$$

range below T_f, the structurally homogeneous state in PMN is the normal ferroelectric state induced by large enough dc bias)[11,13] the obtained value was equal to: $Q_h = (6.0 \pm 0.5) \times 10^{-2}$ (m^4/C^2), which was also independent of temperature at $T \leq T_f$.[5] This result is close to the average value $Q_h = (8.3 \pm 1.0) \times 10^{-2}$ (m^4/C^2) derived from the present structural study of PMN. Apparently, Q_h derived here is about an order of magnitude higher than those found in many early studies.[1,3,4] The reason behind this is the existence of polar regions in the material, the response of which to the external electric field does not generate volume change, as has been discussed in an early publication.[5]

Table II compares Q_h of several perovskite ferroelectrics[3] including the new value for PMN. As one can clearly see, the Q_h of PMN nicely fits the values of the electrostrictive coefficients describing the strain of the prototype cubic lattice of other perovskites.

Figure 1. Excerpt from Testbed Article as Rendered in Adobe Acrobat Reader (PDF Format)

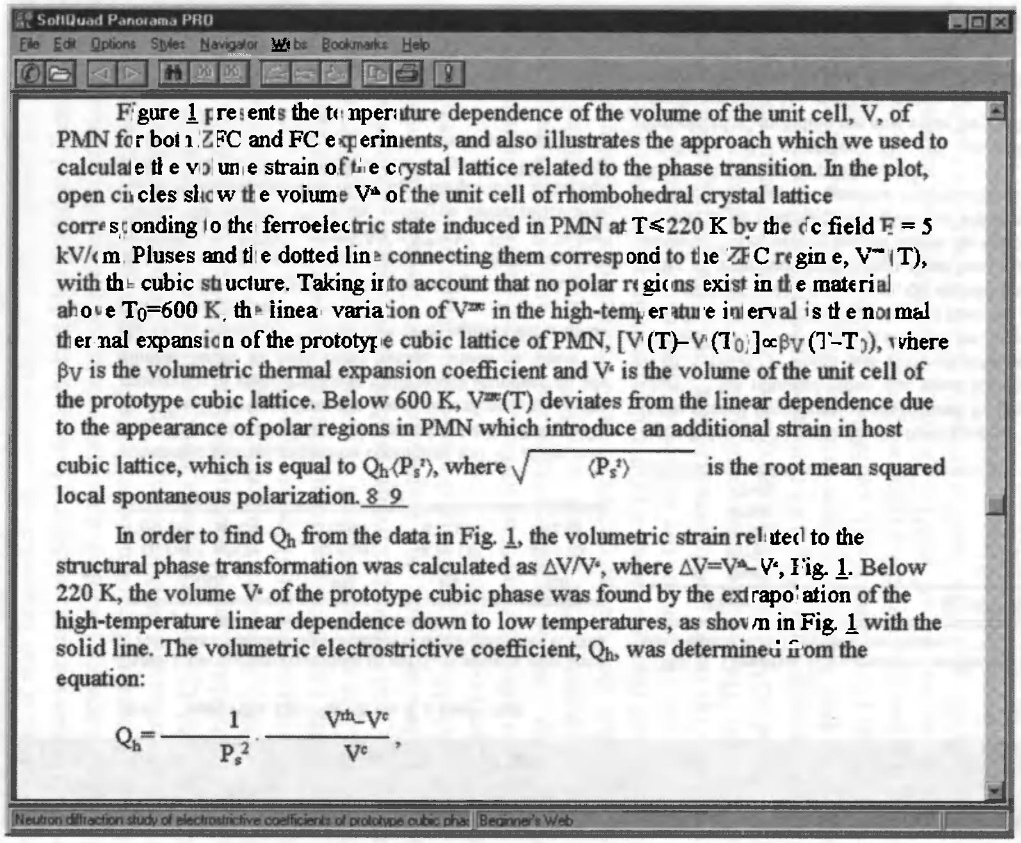

SoftQuad Panorama PRO
File Edit Options Styles Navigator Webs Bookmarks Help
Neutron diffraction study of electrostrictive coefficients of prototype cubic phas
Beginner's Web

Figure 1 presents the temperature dependence of the volume of the unit cell, V, of PMN for both ZFC and FC experiments, and also illustrates the approach which we used to calculate the volume strain of the crystal lattice related to the phase transition. In the plot, open circles show the volume V^{a} of the unit cell of rhombohedral crystal lattice corresponding to the ferroelectric state induced in PMN at $T \leqslant 220$ K by the dc field E = 5 kV/cm. Pluses and the dotted line connecting them correspond to the ZFC regime, $V^{zfc}(T)$, with the cubic structure. Taking into account that no polar regions exist in the material above $T_0=600$ K, the linear variation of V^{zfc} in the high-temperature interval is the normal thermal expansion of the prototype cubic lattice of PMN, $[V(T)-V(T_0)] \propto \beta_V (T-T_0)$, where β_V is the volumetric thermal expansion coefficient and V^c is the volume of the unit cell of the prototype cubic lattice. Below 600 K, $V^{zfc}(T)$ deviates from the linear dependence due to the appearance of polar regions in PMN which introduce an additional strain in host cubic lattice, which is equal to $Q_h \langle P_s^2 \rangle$, where $\sqrt{\quad \langle P_s^2 \rangle \quad}$ is the root mean squared local spontaneous polarization. 8 9

In order to find Q_h from the data in Fig. 1, the volumetric strain related to the structural phase transformation was calculated as $\Delta V/V^c$, where $\Delta V=V^{a}-V^c$, Fig. 1. Below 220 K, the volume V^c of the prototype cubic phase was found by the extrapolation of the high-temperature linear dependence down to low temperatures, as shown in Fig. 1 with the solid line. The volumetric electrostrictive coefficient, Q_h, was determined from the equation:

$$Q_h= \frac{1}{P_s^2} \cdot \frac{V^{rh}-V^c}{V^c},$$

Figure 2. Excerpt from Testbed Article as Rendered in SoftQuad Panorama Viewer, Version 2.0 (SGML Format)

The most problematic aspect of SGML full-text rendering has been in accurately rendering SGML display mathematics. One of the exciting promises of marked-up mathematics is in the potential for searching and displaying both syntactic and semantic elements of article mathematics. The testbed team has explored several techniques for rendering mathematics in a Web-based environment, including converting SGML display mathematics to TeX (and subsequently bit-mapped images) and also the display of marked-up mathematics within HTML and XML using Cascading Style Sheets (CSS). The accurate rendering of mathematics continues to be a major focus of attention for scientific publishing on the Web. The testbed team is experimenting with search techniques for marked-up mathematics.

SGML is an open standard for document representation and transmission. However, SGML is not in itself a markup language but rather a template or model for marking up the content and structure of a document. The Document Type Definition (DTD) accompanying an individual publisher's SGML is the instrument that actually specifies the semantics and syntax of the tags to be used in the document markup. The DTD also specifies the rules that describe the manner in which the SGML tags may be applied to the documents. One of the major roadblocks in the successful deployment of the testbed has been the overhead involved with processing the heterogeneous DTDs of the publishers. Each publisher DTD has required its own suite of processing software. In the process of creating a viable testbed, the Illinois testbed team developed a number of techniques to address problems and normalize SGML processing, indexing, storage, retrieval, and rendering.

The testbed team has converted the SGML publisher data into well-formed XML (eXtensible Markup Language). The XML data can then be rendered natively (without conversion) in a Web browser and/or converted to HTML to be rendered using emerging Web technologies such as CSS and Dynamic HTML (DHTML). It is clear that a rich markup format such as XML, which is a nearly complete instance of SGML, will become the standard language of open document systems, to be used in Web environments for document representation and delivery. XML and SGML permit documents to be treated as objects to be viewed, manipulated, and output. The major strengths of these markup languages, in terms of their retrieval capabilities, are their ability to reveal the deep content and structure of a document. While SGML and XML are becoming ubiquitous in the publishing world, they are still, for the most part, being generated by publishers as a byproduct for archiving and search engine indexing, rather than serving as an integral, integrated part of the production process.

SGML, HTML, and XML, for document representation and display, offer various levels of maturity. SGML supports powerful indexing, search, and retrieval but requires a sophisticated search engine and a plug-in viewer for display. SGML is generally regarded as difficult to use and, along with the client, delivery, and rendering issues, remains a "Web-unfriendly" technology. HTML is ubiquitous and with HTML 4.0 and CSS provides robust rendering capabilities. However, HTML remains a presentation-oriented language with inadequate semantic tools for the effective indexing and fine-granularity searching needed for academic journals. XML is a distinguished subset of SGML that retains the key features of SGML, including semantic-based tagging. But, XML and the XSL styling language are new technologies still being shaped by the standards process. XML cannot be rendered accurately in the 4.0 browsers but is easy to transform to HTML and can be natively rendered by Internet Explorer 5.0.

Figures 3 and 4 show HTML and XML renderings of the same article excerpt displayed in Figures 1 and 2. Note that the use of CSS provides enhanced mathematics rendering and promises sophisticated display capabilities.

DATABASE SEARCH ENGINES AND ARTICLE SERVERS

Various database management systems capable of indexing and searching SGML-based databases were examined. The OpenText DBMS was chosen for the testbed because of its ability to exploit the strengths of SGML. The OpenText search engine grew out of work done at the University of Waterloo to create and index the SGML version of the Oxford English Dictionary (Terry, 1991). OpenText also had attractive features for indexing document metadata in conjunction with document full-text, for normalizing documents created with different publisher Document Type Definitions (DTDs), and for maintaining multiple discrete database repositories. Additionally, OpenText's architecture allows the integration of third party tools, the implementation of locally developed scripts and code, the capability of bypassing unneeded component modules, and the ability to rapidly change processing procedures in response to dynamic processing needs.

Originally the UIUC DLI project had expected to influence generic Web client development by influencing development of the NCSA Mosaic Web browser. This proved a naive expectation. The testbed team realized that search and delivery of testbed materials needed to be done in a browser-neutral manner. A focus of the testbed project has been the development of server-side scripts and dynamic document merge capabilities. The testbed team has employed NT and UNIX operating system platforms as appropriate to task. Netscape Enterprise and Microsoft Internet Information Server Web servers are used. Web server functionality is extended using both conventional CGI and more advanced techniques such as Microsoft's Active Server Platform (ASP). HTTPS protocols (HTTP with

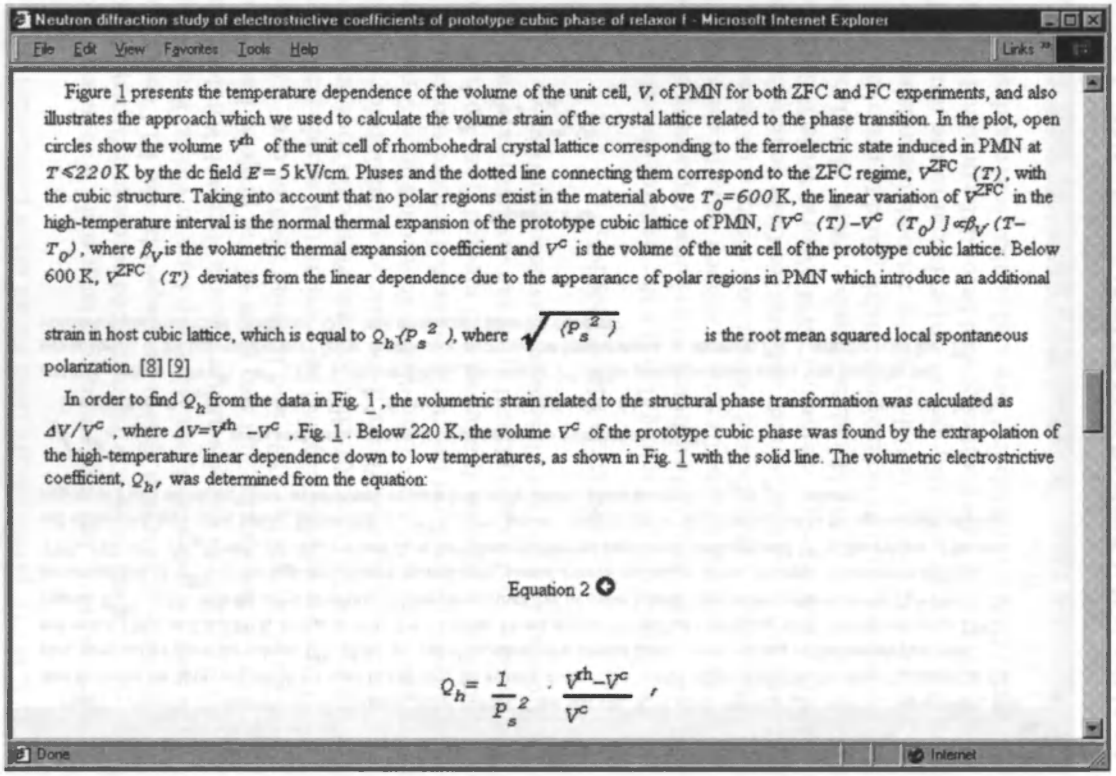

Neutron diffraction study of electrostrictive coefficients of prototype cubic phase of relaxor f - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Links
Done
Internet

Figure 1 presents the temperature dependence of the volume of the unit cell, V, of PMN for both ZFC and FC experiments, and also illustrates the approach which we used to calculate the volume strain of the crystal lattice related to the phase transition. In the plot, open circles show the volume V^{rh} of the unit cell of rhombohedral crystal lattice corresponding to the ferroelectric state induced in PMN at $T \leqslant 220$ K by the dc field $E = 5$ kV/cm. Pluses and the dotted line connecting them correspond to the ZFC regime, $V^{ZFC}(T)$, with the cubic structure. Taking into account that no polar regions exist in the material above $T_0 = 600$ K, the linear variation of V^{ZFC} in the high-temperature interval is the normal thermal expansion of the prototype cubic lattice of PMN, $[V^c(T) - V^c(T_0)] \propto \beta_V (T - T_0)$, where β_V is the volumetric thermal expansion coefficient and V^c is the volume of the unit cell of the prototype cubic lattice. Below 600 K, $V^{ZFC}(T)$ deviates from the linear dependence due to the appearance of polar regions in PMN which introduce an additional strain in host cubic lattice, which is equal to $Q_h \langle P_s^2 \rangle$, where $\sqrt{\langle P_s^2 \rangle}$ is the root mean squared local spontaneous polarization. [8] [9]

In order to find Q_h from the data in Fig. 1, the volumetric strain related to the structural phase transformation was calculated as $\Delta V/V^c$, where $\Delta V = V^{rh} - V^c$, Fig. 1. Below 220 K, the volume V^c of the prototype cubic phase was found by the extrapolation of the high-temperature linear dependence down to low temperatures, as shown in Fig. 1 with the solid line. The volumetric electrostrictive coefficient, Q_h, was determined from the equation:

Equation 2 ⊕

$$Q_h = \frac{1}{P_s^2} \cdot \frac{V^{rh} - V^c}{V^c},$$

Figure 3. Excerpt from Testbed Article as Rendered in Microsoft Internet Explorer, Version 5.0 (HTML Format)

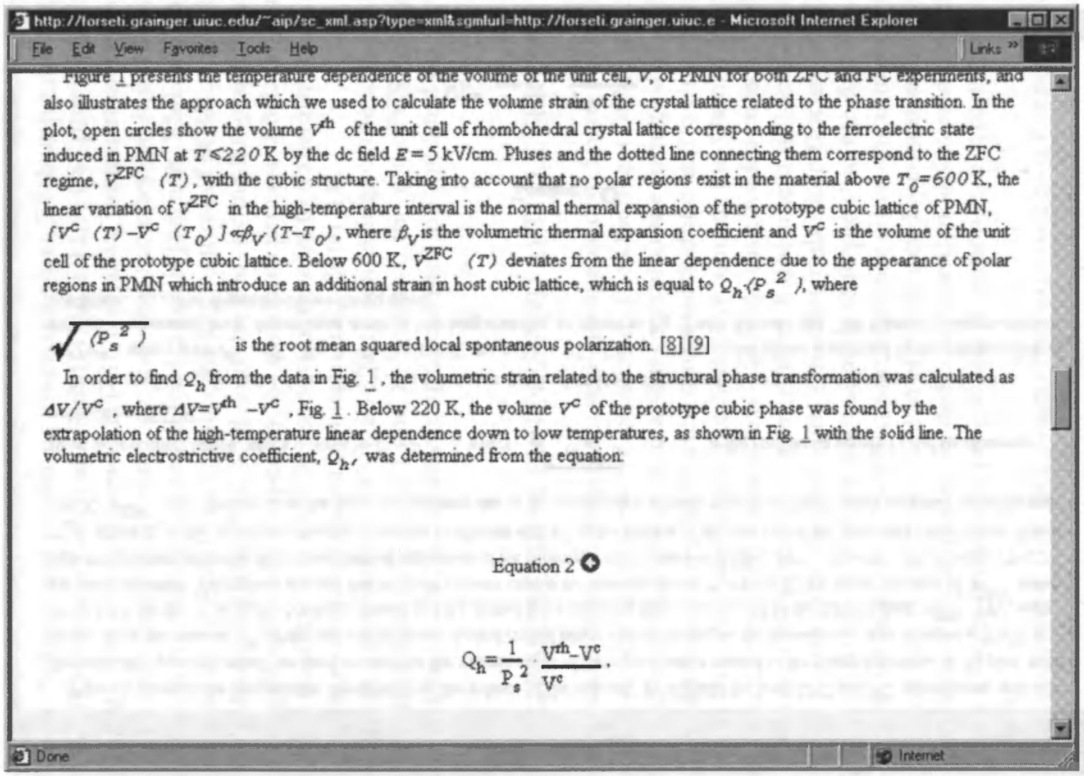

http://forseti.grainger.uiuc.edu/~aip/sc_xml.asp?type=xml&sgmlurl=http://forseti.grainger.uiuc.e - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Links »
Done
Internet

Figure 1 presents the temperature dependence of the volume of the unit cell, V, of PMN for both ZFC and FC experiments, and also illustrates the approach which we used to calculate the volume strain of the crystal lattice related to the phase transition. In the plot, open circles show the volume V^{th} of the unit cell of rhombohedral crystal lattice corresponding to the ferroelectric state induced in PMN at $T \leqslant 220$ K by the dc field $E = 5$ kV/cm. Pluses and the dotted line connecting them correspond to the ZFC regime, $V^{ZFC}(T)$, with the cubic structure. Taking into account that no polar regions exist in the material above $T_0 = 600$ K, the linear variation of V^{ZFC} in the high-temperature interval is the normal thermal expansion of the prototype cubic lattice of PMN, $[V^c(T) - V^c(T_0)] \propto \beta_V(T-T_0)$, where β_V is the volumetric thermal expansion coefficient and V^c is the volume of the unit cell of the prototype cubic lattice. Below 600 K, $V^{ZFC}(T)$ deviates from the linear dependence due to the appearance of polar regions in PMN which introduce an additional strain in host cubic lattice, which is equal to $Q_h \langle P_s^2 \rangle$, where

$\sqrt{\langle P_s^2 \rangle}$ is the root mean squared local spontaneous polarization. [8] [9]

In order to find Q_h from the data in Fig. 1, the volumetric strain related to the structural phase transformation was calculated as $\Delta V/V^c$, where $\Delta V = V^{th} - V^c$, Fig. 1. Below 220 K, the volume V^c of the prototype cubic phase was found by the extrapolation of the high-temperature linear dependence down to low temperatures, as shown in Fig. 1 with the solid line. The volumetric electrostrictive coefficient, Q_h, was determined from the equation:

Equation 2 ●

$$Q_h = \frac{1}{P_s^2} \cdot \frac{V^{th} - V^c}{V^c}.$$

Figure 4. Excerpt from Testbed Article as Rendered in Microsoft Internet Explorer, Version 5.0 (XML Format)

Secure Socket Layers) are used for user authentication and authorization as required. HTTP protocols are used for all other interactions with clients.

The testbed team has implemented a large-scale Web-based testbed of full-text journal articles featuring enhanced access and display capabilities. The Web-based retrieval system developed by the DLI testbed and evaluation teams is called DeLIver (Desktop Link to Virtual Engineering Resources). The DeLIver client, which replaced a Microsoft Windows-based custom client in use for the first two years of the project, has been in operation since September 1997 and is being used by over 2,000 registered UIUC students and faculty and also designated outside researchers. Figure 5 shows the DeLIver search interface. Figure 6 shows an abbreviated citation results list and Figure 7 the extended citation for a specific retrieved item. The contents of the display shown in Figure 7 represents the metadata associated with the retrieved item. Detailed transaction log data of user search sessions (gathered and merged from both database and Web servers) are being kept, and a preliminary analysis of user search patterns from some 4,200 search sessions has been performed.

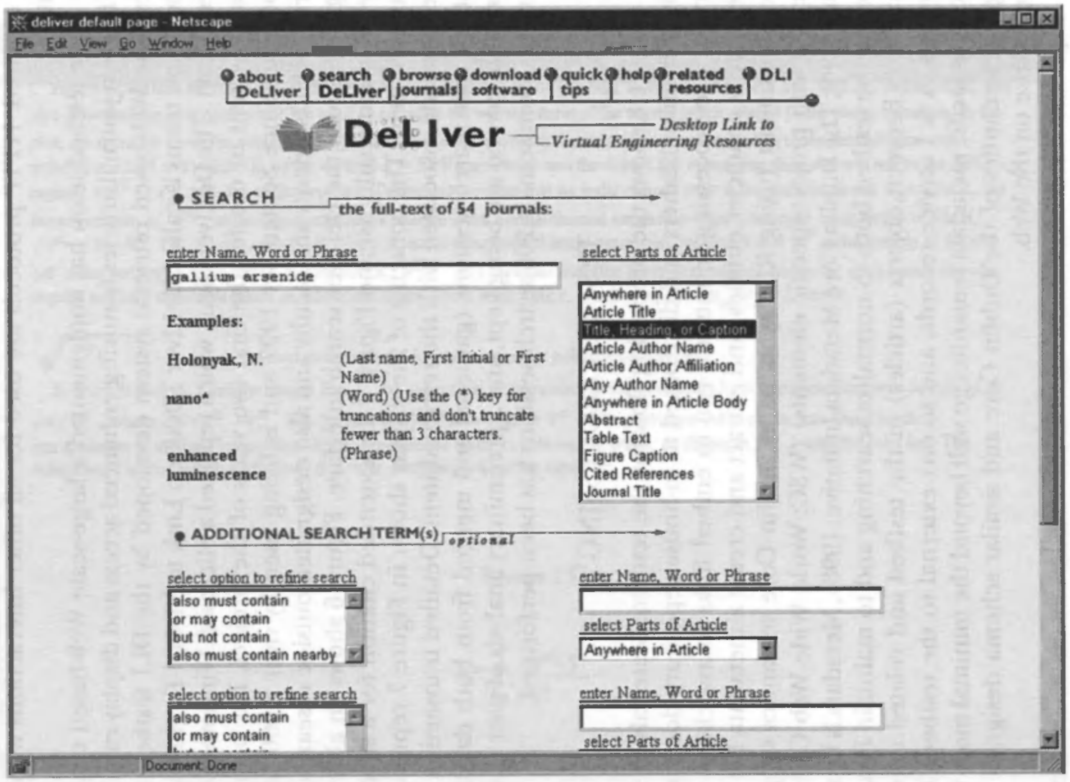
TESTBED PROCESSING

Figure 8 shows the processing flow for the testbed. Materials are received from publishers and distributed to in-house repository document servers. Pre-processing scripts are run to embed links to associated figures, check character entities, and extract and create a metadata file for each document (using RDF syntax and Dublin Core semantics supplemented with project-specific elements) (W3C: World Wide Web Consortium, 1999; The Dublin Core Metadata Initiative, 1998). Metadata is heavily used in the testbed both to normalize searching and to maintain link information between objects (articles) in the testbed and related objects (articles, A & I service records, and so on) external to the testbed. The project-specific metadata semantics go well beyond the minimal metadata tagging semantics of the Dublin Core and similar schema designed for general use on the Web.

OpenText indexes are then built. The article metadata is indexed along with the full-text of the articles. The individual indexes can be searched separately or in parallel. Tag aliasing is applied to support normalized searching. CGI and ASP scripts are used to enhance search functionality; insert hyperlink information; perform transformations between SGML, XML, and HTML formats; and facilitate linking between testbed objects and related information both within, and external to, the testbed.

RETRIEVAL MODELS

To support effective retrieval in the testbed, the Illinois DLI testbed



SEARCH the full-text of 54 journals:

enter Name, Word or Phrase

gallium arsenide

select Parts of Article

Anywhere in Article
Article Title
Title, Heading, or Caption
Article Author Name
Article Author Affiliation
Any Author Name
Anywhere in Article Body
Abstract
Table Text
Figure Caption
Cited References
Journal Title

Examples:

Holonyak, N.	(Last name, First Initial or First Name)
nano*	(Word) (Use the (*) key for truncations and don't truncate fewer than 3 characters.
enhanced luminescence	(Phrase)

ADDITIONAL SEARCH TERM(s) optional

select option to refine search

also must contain
or may contain
but not contain
also must contain nearby

enter Name, Word or Phrase

select Parts of Article

Anywhere in Article

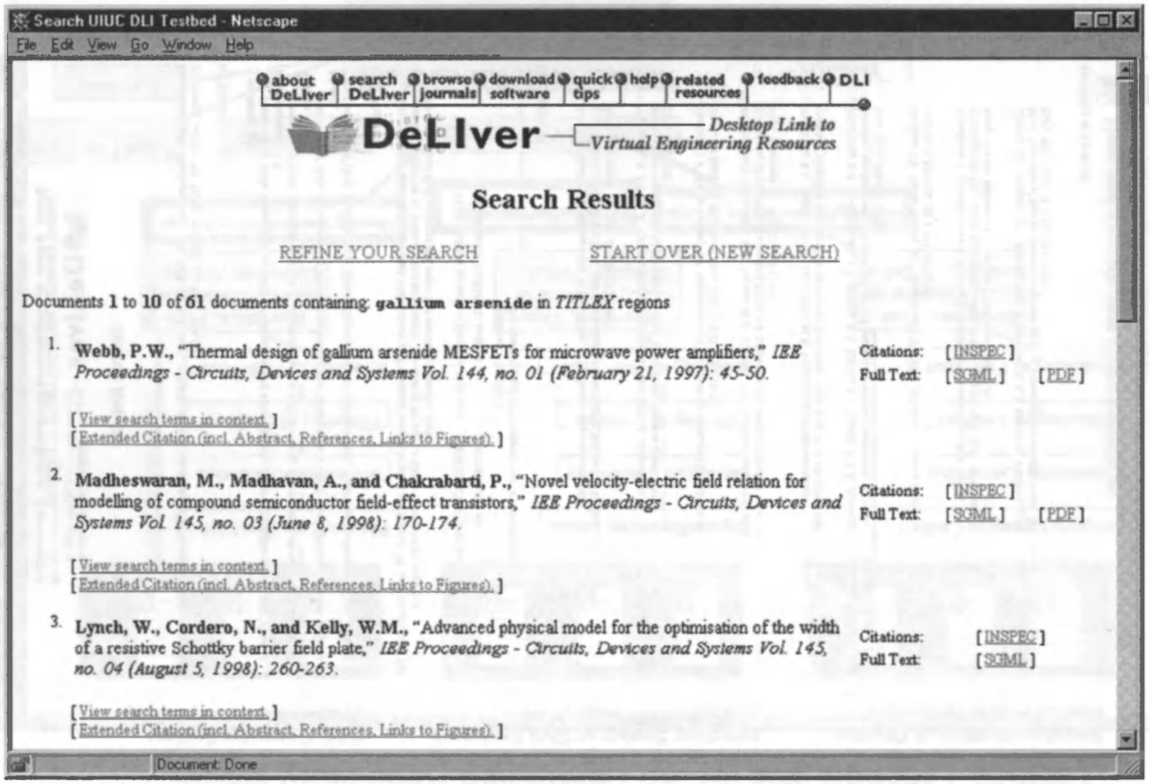
select option to refine search

also must contain
or may contain

enter Name, Word or Phrase

select Parts of Article

Figure 5. Illinois DLI Search Interface Screen Showing Variety of Article Elements that can be Searched



Search Results

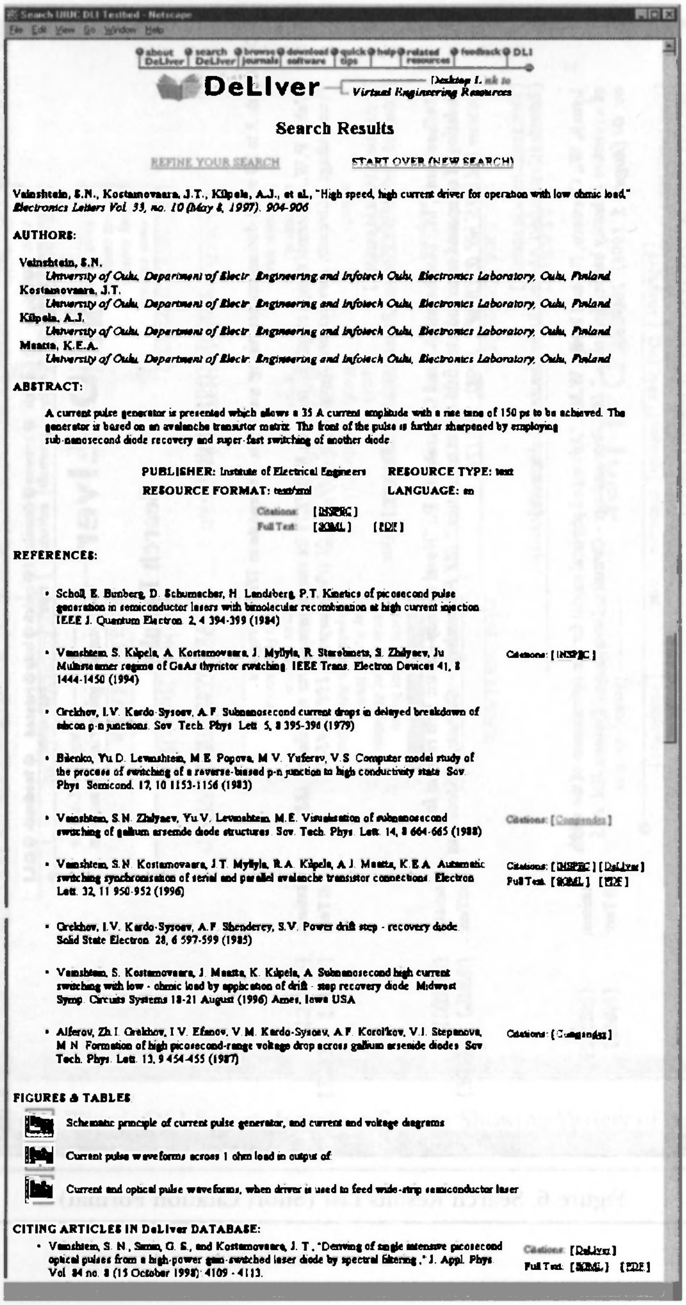
Documents 1 to 10 of 61 documents containing: **gallium arsenide** in *TITLEX* regions

1. **Webb, P.W.**, "Thermal design of gallium arsenide MESFETs for microwave power amplifiers," *IEE Proceedings - Circuits, Devices and Systems Vol. 144, no. 01 (February 21, 1997): 45-50.*
   Citations: [INSPEC]
   Full Text: [SGML] [PDF]
   [View search terms in context.]
   [Extended Citation (incl. Abstract, References, Links to Figures).]

2. **Madheswaran, M., Madhavan, A., and Chakrabarti, P.**, "Novel velocity-electric field relation for modelling of compound semiconductor field-effect transistors," *IEE Proceedings - Circuits, Devices and Systems Vol. 145, no. 03 (June 8, 1998): 170-174.*
   Citations: [INSPEC]
   Full Text: [SGML] [PDF]
   [View search terms in context.]
   [Extended Citation (incl. Abstract, References, Links to Figures).]

3. **Lynch, W., Cordero, N., and Kelly, W.M.**, "Advanced physical model for the optimisation of the width of a resistive Schottky barrier field plate," *IEE Proceedings - Circuits, Devices and Systems Vol. 145, no. 04 (August 5, 1998): 260-263.*
   Citations: [INSPEC]
   Full Text: [SGML]
   [View search terms in context.]
   [Extended Citation (incl. Abstract, References, Links to Figures).]

Figure 6. Search Results List (Short Citation Format)


Search UIUC DLI Testbed - Netscape
File Edit View Go Window Help
about DeLiver
search DeLiver
browse journals
download software
quick tips
help
related resources
feedback
DLI
DeLIver
Desktop Link to Virtual Engineering Resources

Search Results

REFINE YOUR SEARCH START OVER (NEW SEARCH)

Vainshtein, S.N., Kostamovaara, J.T., Kilpela, A.J., et al., "High speed, high current driver for operation with low ohmic load," *Electronics Letters Vol. 33, no. 10 (May 8, 1997). 904-906*

AUTHORS:

Vainshtein, S.N.
University of Oulu, Department of Electr. Engineering and Infotech Oulu, Electronics Laboratory, Oulu, Finland
Kostamovaara, J.T.
University of Oulu, Department of Electr. Engineering and Infotech Oulu, Electronics Laboratory, Oulu, Finland
Kilpela, A.J.
University of Oulu, Department of Electr. Engineering and Infotech Oulu, Electronics Laboratory, Oulu, Finland
Maatta, K.E.A.
University of Oulu, Department of Electr. Engineering and Infotech Oulu, Electronics Laboratory, Oulu, Finland

ABSTRACT:

A current pulse generator is presented which allows a 35 A current amplitude with a rise time of 150 ps to be achieved. The generator is based on an avalanche transistor matrix. The front of the pulse is further sharpened by employing sub-nanosecond diode recovery and super-fast switching of another diode.

PUBLISHER: Institute of Electrical Engineers **RESOURCE TYPE:** text
RESOURCE FORMAT: text/xml **LANGUAGE:** en

Citations: [INSPEC]
Full Text: [SGML] [PDF]

REFERENCES:

- Scholl, E. Bimberg, D. Schumacher, H. Landsberg, P.T. Kinetics of picosecond pulse generation in semiconductor lasers with bimolecular recombination at high current injection. IEEE J. Quantum Electron. 2, 4 394-399 (1984)
- Vainshtein, S. Kilpela, A. Kostamovaara, J. Myllyla, R. Starobinets, S. Zhilyaev, Ju Multistreamer regime of GaAs thyristor switching. IEEE Trans. Electron Devices 41, 8 1444-1450 (1994) Citations: [INSPEC]
- Grekhov, I.V. Kardo-Sysoev, A.F. Subnanosecond current drops in delayed breakdown of silicon p-n junctions. Sov. Tech. Phys. Lett. 5, 8 395-396 (1979)
- Bilenko, Yu.D. Levinshtein, M.E. Popova, M.V. Yuferev, V.S. Computer model study of the process of switching of a reverse-biased p-n junction to high conductivity state. Sov. Phys. Semicond. 17, 10 1153-1156 (1983)
- Vainshtein, S.N. Zhilyaev, Yu.V. Levinshtein, M.E. Visualisation of subnanosecond switching of gallium arsenide diode structures. Sov. Tech. Phys. Lett. 14, 8 664-665 (1988) Citations: [Compendex]
- Vainshtein, S.N. Kostamovaara, J.T. Myllyla, R.A. Kilpela, A.J. Maatta, K.E.A. Automatic switching synchronisation of serial and parallel avalanche transistor connections. Electron Lett. 32, 11 950-952 (1996) Citations: [INSPEC] [DeLiver] Full Text: [SGML] [PDF]
- Grekhov, I.V. Kardo-Sysoev, A.F. Shenderey, S.V. Power drift step - recovery diode. Solid State Electron. 28, 6 597-599 (1985)
- Vainshtein, S. Kostamovaara, J. Maatta, K. Kilpela, A. Subnanosecond high current switching with low - ohmic load by application of drift - step recovery diode. Midwest Symp. Circuits Systems 18-21 August (1996) Ames, Iowa USA
- Alferov, Zh.I. Grekhov, I.V. Efanov, V.M. Kardo-Sysoev, A.F. Korol'kov, V.I. Stepanova, M.N. Formation of high picosecond-range voltage drop across gallium arsenide diodes. Sov. Tech. Phys. Lett. 13, 9 454-455 (1987) Citations: [Compendex]

FIGURES & TABLES

Schematic principle of current pulse generator, and current and voltage diagrams

Current pulse waveforms across 1 ohm load in output of:

Current and optical pulse waveforms, when driver is used to feed wide-strip semiconductor laser

CITING ARTICLES IN DeLIver DATABASE:

- Vainshtein, S. N., Simin, G. S., and Kostamovaara, J. T., "Deriving of single intensive picosecond optical pulses from a high-power gain-switched laser diode by spectral filtering," J. Appl. Phys. Vol. 84 no. 8 (15 October 1998): 4109 - 4113. Citations: [DeLiver] Full Text: [SGML] [PDF]

Figure 7. Extended Citation for Testbed Article; Item Metadata is Used to Create this View

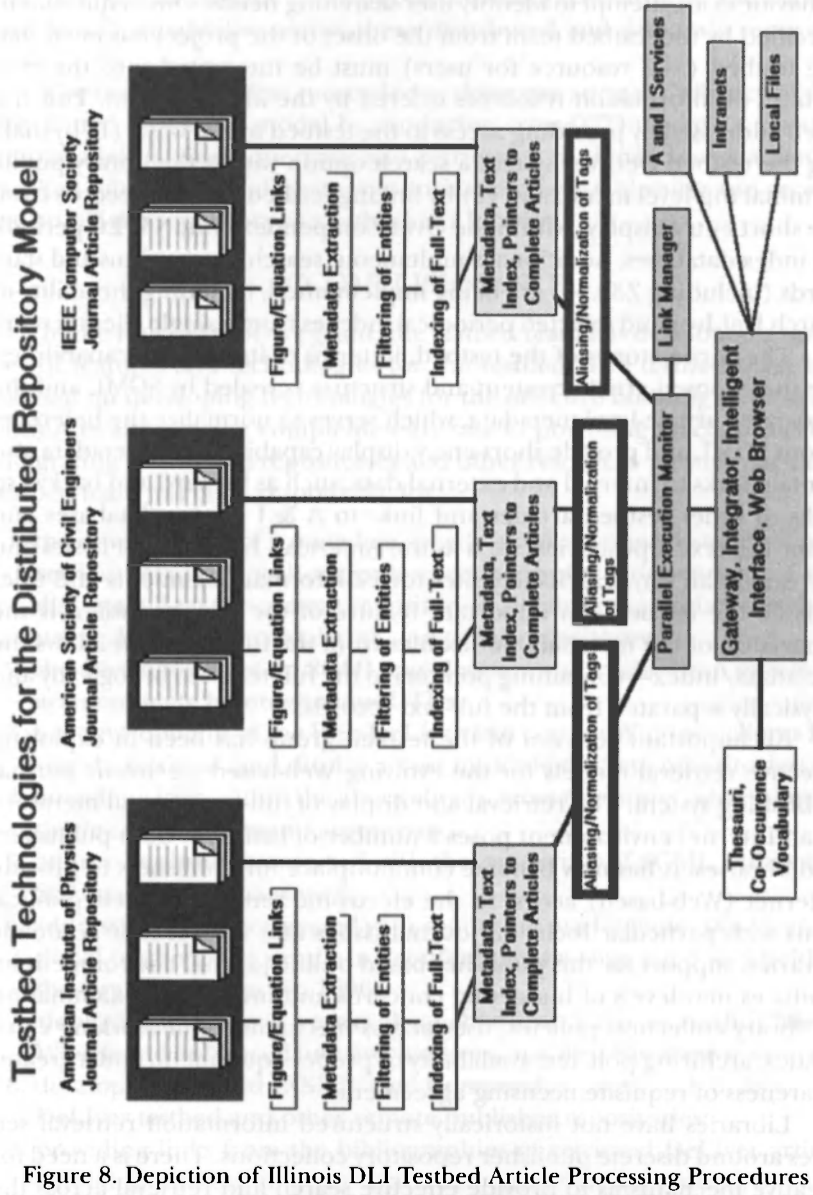

Testbed Technologies for the Distributed Repository Model
American Institute of Physics
Journal Article Repository
American Society of Civil Engineers
Journal Article Repository
IEEE Computer Society
Journal Article Repository
Figure/Equation Links
Metadata Extraction
Filtering of Entities
Indexing of Full-Text
Figure/Equation Links
Metadata Extraction
Filtering of Entities
Indexing of Full-Text
Figure/Equation Links
Metadata Extraction
Filtering of Entities
Indexing of Full-Text
Metadata, Text Index, Pointers to Complete Articles
Metadata, Text Index, Pointers to Complete Articles
Metadata, Text Index, Pointers to Complete Articles
Aliasing/Normalization of Tags
Aliasing/Normalization of Tags
Aliasing/Normalization of Tags
Thesauri, Co-Occurence Vocabulary
Parallel Execution Monitor
Link Manager
Gateway, Integrator, Intelligent Interface, Web Browser
A and I Services
Intranets
Local Files

Figure 8. Depiction of Illinois DLI Testbed Article Processing Procedures

and evaluation teams have also carried out studies of end-user searching behavior in an attempt to identify user-searching needs. One requirement specified by the testbed team from the onset of the project has been that the testbed (as a resource for users) must be integrated into the continuum of information resources offered by the library system. This has been addressed by providing access to the testbed in two ways: (1) by making the testbed DeLIver system a search option within the library public terminal top-level menu; and (2) by linking testbed full-text records from the short entry displays within the Ovid Compendex and INSPEC periodical index databases. Additional simultaneous search mechanisms and standards (including Z39.50) are being implemented, including the ability to search DeLIver and selected periodical indexes from a single client screen.

The cornerstones of the testbed, in terms of its retrieval capabilities, are the exposed article content and structure revealed by SGML and the associated article-level metadata, which serves to normalize the heterogeneous SGML and provide short-entry display capability. The metadata also contain links to internal and external data, such as forward and backward links to other testbed articles and links to A & I service databases and other full-text repositories, such as the American Institute of Physics and the American Physical Society sites for PDF format documents and titles outside the testbed. An important feature of the testbed design is the separation of the metadata/index files from the full-text. This allows the metadata/index—containing pointers to the full-text—to be logically and physically separated from the full-text records.

An important concern of the testbed group has been in exploring effective retrieval models for the evolving Web-based electronic journal publishing system. The retrieval and display of full-text journal literature in an Internet environment poses a number of issues for both publishers and libraries. It has now become commonplace for publishers to provide Internet (Web-based) access to the electronic versions of their publications with particular focus on journal issues and articles. For academic libraries, support for this publisher-based online journal environment introduces new levels of budgeting concerns and involves an examination of library collection policies, user access mechanisms, networking capabilities, archiving policies, availability of proper equipment, and a greater awareness of requisite licensing agreements.

Libraries have not historically structured information retrieval services around discrete publisher repository collections. There is a need for creative mechanisms to provide effective search and retrieval across the burgeoning number of distributed heterogeneous publisher repositories. To support this, the testbed team has proposed a distributed repository model that "federates" or connects the individual publisher repositories of full-text documents. In the DLI testbed model, these distributed repositories are federated by the extraction of normalized metadata, index, and link data from the heterogeneous full text of the different publishers. This model addresses the challenge of providing standardized and consistent search capabilities across these distributed and disparate repositories.

The testbed team has succeeded in demonstrating the efficacy of the distributed repository model by producing cross-DTD metadata, providing parallel database querying and retrieval techniques across a distinguished subset of the full-text repositories, and by establishing and accessing an off-site repository at a publisher's location.

ACCOMPLISHMENTS

In the four years of the grant, the testbed team has developed a number of features and technologies for the testbed. The testbed team has focused on developing technologies for the effective building of local repositories and also the complementary task of providing mechanisms for integrating distributed repositories and other resources. In summary, the testbed team has been responsible for:

1. the development of a metadata specification to support standardized retrieval across repositories; short-entry display independent of the discrete full-text document repositories; and links to associated testbed items, A & I service databases, and other repositories;
2. the development of an SGML tag aliasing or normalization system to accommodate heterogeneous DTDs;
3. the development of the Web DeLIver and custom Windows clients for search, retrieval, and display across multiple discrete repositories;
4. providing, from within the above clients, cross-repository retrieval from single search command arguments;
5. addressing issues connected with the rendering of SGML within the Softquad Panorama viewer;
6. addressing issues connected with rendering mathematics (an international mathematics rendering conference was organized and held at the Grainger Library in 1996);
7. deploying enhanced retrieval mechanisms, such as Author Word Wheels and enhanced link mechanisms, in a Web-based environment;
8. developing an Ovid INSPEC and Compendex proxy with links to the DeLIver testbed and other remote publisher repositories;
9. providing links from the bibliographies of retrieved DeLIver article extended citations to other articles contained in the testbed;
10. providing forward citation links within testbed article extended citations to subsequently published articles that cite the retrieved testbed article;
11. providing links from the retrieved DeLIver articles and references in the bibliographies of retrieved DeLIver articles to INSPEC,

Compendex, SPIN, and other records in periodical index and repositories systems;

12. employing Web-based user questionnaires and surveys;
13. generating detailed user transaction logs, gathered at the search argument level, with the automatic identification and storage of characteristics of each user search sessions;
14. providing in-depth analysis of user search behavior, including statistics on the frequency of use of each DeLIver search feature;
15. providing simultaneous searching of a user entered search argument in DeLIver and periodical index databases;
16. employing a Web-Kerberos based user authentication via the UIUC Bluestem Web-based user authentication system (Cole, 1997); and
17. testing the capability of digital signing of documents.

This work has been accomplished with the cooperation and support of our publisher partners and through the use of commercial software from OpenText, Hewlett-Packard, SoftQuad, and Microsoft. The testbed team has made available the results of the project to our publisher partners and sponsors in annual workshops and through regular communications.

LESSONS LEARNED

The potential of SGML (and now XML) has been borne out by experience. The full-text indexes are extremely rich, supporting a measure of search precision unavailable in previous full-text search systems. Figure 5 shows the search fields available to end-users in the current interface. SGML has greatly facilitated extraction of metadata and insertion of hyperlinks to related resources within and external to the testbed.

Rendering of complex mathematical mark-up continues to be problematic. Until recently the testbed relied solely on the Panorama SGML viewer originally marketed by SoftQuad. In spite of promises to improve the rendering engine, development has lagged (Panorama was recently sold to Interleaf) and there still isn't a version of Panorama for the Macintosh. Rapid development of XML, advances in the latest version of HTML, and development of Cascading Style Sheets (CSS) are improving prospects for better rendering. Nonetheless, our experience with Panorama demonstrates the degree to which libraries and information providers are dependent on the commercial sector for essential technology.

A detailed transaction log analysis of 4,158 end-user search sessions has been conducted. Several interesting results have been gleaned from the transaction logs. These include: there is very little use being made of either "Help" or "Quicktips" functions; browsing of tables of contents is being performed in 39 percent of the search sessions; full-text searching is the predominant search mode, but in 24 percent of the sessions users performed a search within a specific field; full-text is displayed in more sessions (69 percent) than extended citations (19 percent); in 25 percent of the sessions, users did multi-concept searching; and an average of four full-text documents are viewed per session.

Overall development of the testbed has taken longer than anticipated. With some notable exceptions—e.g., the lack of a robust SGML viewer—the technology needed has been available by the time needed. The development of processing procedures, the normalization of DTDs, and the development and implementation of metadata semantics have taken longer than anticipated. Technology infrastructure changes happen much more quickly than process changes that involve changing how libraries and information providers do their jobs.

Implementation of a digital information resource requires tighter integration of the parties involved. Small changes by a publisher in tagging semantics can require corresponding changes in indexing scripts, metadata extraction procedures and, further downstream, style sheet design. Conversely, changes in browser software or rendering client can necessitate changes in tagging and indexing. Because each of these tasks may be performed by a different agency, close efficient working relationships are essential.

In the electronic journal environment, roles and responsibilities are more fluid. While documents may reside on a publisher's server, metadata may reside elsewhere—e.g., on an abstracting and indexing service's hardware. Different agencies may create different metadata for the same objects—e.g., using different controlled vocabularies. Libraries may implement their own gateways and portals or may contract for such services with consortia or other third parties. A single article may be found through different gateways, using different index and metadata providers, even if full content of the article itself still comes from a single publisher's server. Archival responsibilities may be distributed among libraries, consortia, and publishers.

In the rapidly evolving electronic journals environment, academic libraries will need to re-examine their collection development policies in terms of ownership versus access, become more actively involved in institutional and consortial licensing agreements, and become more actively involved in campus networking, server, and workstation policies and technologies.

FUTURE FOCI

The testbed team expects to continue work on the issues addressed in the DLI grant through a Corporation for National Research Initiatives (CNRI) grant and the establishment of a Collaborating Partners program. CNRI has established a collaborative Digital Library (D-Lib) Test Suite program encompassing five operational digital library testbeds. The D-Lib Test Suite program is expected to provide a fertile research environment for the information science community. The testbed team members and associated researchers will explore a number of evolving information technologies.

The entire testbed was recently converted to XML. Testbed articles are now retrievable in XML and HTML as well as in PDF and SGML. This has already improved rendering options and overall quality. The potential of the Math ML standard to support more accurate rendering of testbed content will be investigated (W3C: World Wide Web Consortium, 1999).

In addition, further testing of distributed architecture models will be done to test scalability and performance of the options. The use of Document Object Identifiers (DOIs) ("Technology Update: Digital Object Identifiers," 1998) and other emerging standards to enhance and facilitate link management will be investigated.

Also, additional simultaneous search features—e.g., allowing simultaneous searching of non-testbed information resources—will be further refined. It is expected that search agent technologies, including Knowbot software, will play an important role in the evolving distributed repository model being promoted by the Illinois testbed.

REFERENCES

Cole, T. (1997). Using Bluestem for Web user authentification and access control of library resources. *Library Hi-Tech, 15*(1-2), 58-71.

Dublin Core Metadata Initiative. (1998). *The Dublin Core: A simple content description model for electronic resources* [online]. Retrieved January 24, 2000 from the World Wide Web: http://purl.oclc.org/dc/.

Schatz, B.; Mischo, W. H.; Cole, T. W.; Bishop, A.; Harum, S.; Johnson, E.; Neumann, L.; & Chen, H. (1999). Federated search of scientific literature: A retrospective on the Illinois Digital Library Project. *IEEE Computer, 32*(2), 51-59.

Schatz, B.; Mischo, W. H.; Cole, T. W.; Hardin, J.; Bishop, A.; & Chen, H. (1996). Federating diverse collections of scientific literature. *IEEE Computer, 29*(5), 28-37.

Sperberg-McQueen, C. M. (1994). The Text Encoding Initiative: Electronic text markup for research. In B. Sutton (Ed.), *Literary texts in an electronic age: Scholarly implications and library services* (Papers presented at the 1994 Clinic on Library Applications of Data Processing, April 10-12, 1994) (pp. 35-56). Urbana-Champaign: Graduate School of Library and Information Science, University of Illinois.

Technology Update: Digital Object Identifiers. (1998). *Online & CD-ROM Review, 22*(2), 115-118. See also: International DOI Foundation (1998). The Digital Object Identifier System [online]. Retrieved January 24, 2000 from the World Wide Web: http://www.doi.org/.

Tenopir, C., & Ro, J. S. (1990). *Full text databases.* New York: Greenwood Press.

Terry, D. (1991). Sidebar 4: Open Text Corporation. *Library Hi Tech, 9*(3), 7-44.

Weibel, S. (1994). The CORE Project: Technical shakedown phase and preliminary user studies. *OCLC Systems and Services, 10*(2 & 3), 99-102.

W3C: World Wide Web Consortium. (1999). *Resource Description Framework (RDF) Model and Syntax Specification: W3C Recommendation, 22 February 1999* [online]. Retrieved January 24, 2000 from the World Wide Web: http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/.

W3C: World Wide Web Consortium. (1998). *Mathematical Markup Language (Math ML) 1.0 Specification: W3C Recommendation,07 April 1998* [online]. Retrieved January 24, 2000 from the World Wide Web: http://www.w3.org/TR/1998/REC-MathML-19980407/ [28 February 1999].