NISTIR 5708

Electronic Access To Standards On The Information Highway

Sharon J. Laskowski
Venkata V. Ramayya

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Computer Systems Laboratory
Information Access and User Interface Division
Gaithersburg, MD 20899

NIST

NISTIR 5708

Electronic Access To Standards On The Information Highway

Sharon J. Laskowski
Venkata V. Ramayya

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Computer Systems Laboratory
Information Access and User Interface Division
Gaithersburg, MD 20899

August 1995

U.S. DEPARTMENT OF COMMERCE
Ronald H. Brown, Secretary

TECHNOLOGY ADMINISTRATION
Mary L. Good, Under Secretary for Technology

NATIONAL INSTITUTE OF STANDARDS
AND TECHNOLOGY
Arati Prabhakar, Director

Electronic Access to Standards on the Information Highway

Sharon J. Laskowski
The National Institute of Standards and Technology
Building 225, Room A216
Gaithersburg, MD 20899
sharon.laskowski@nist.gov

Venkata V. Ramayya
ANSI-National Standards Systems Network Project
The Boeing Company
1700 North Moore Street
Arlington, VA 22209
ramayya@bcstec.ca.boeing.com

1. Introduction

Industries across the U.S. reference tens of thousands of external standards during the course of the design and manufacture of products. These standards and related technical information must be identified, located, and ordered from a multitude of technical societies, associations, and companies around the world. For example, there are over 100,000 U.S. standards [2] as illustrated in Figure 1. To complicate matters, during the life cycle of a product, standards are often updated. As a result, industry spends large amounts of time and effort searching and maintaining standards.

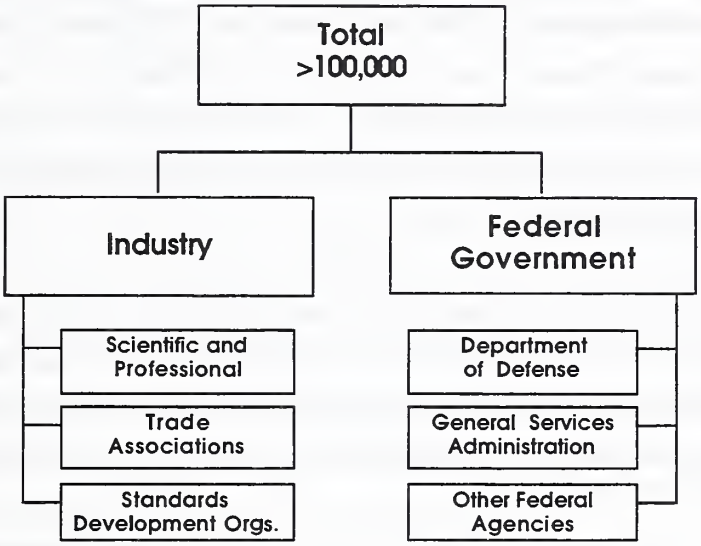

Figure 1 U.S. Standards

If businesses in the U.S. are to compete successfully in global markets, they need to have rapid access to both North American standards and other national and international standards. The timely development, publication, and implementation of standards is also a critical element in supporting effective competition in the global marketplace.

The analysis and investigation described in this paper is part of an ongoing National Institute of Standards and Technology (NIST) test bed effort that is providing technical assistance to the National Standards Systems Network (NSSN) Technology Reinvestment Program [3,19] directed by the American National Standards Institute (ANSI). First, we discuss how the information highway can be used to support electronic access to standards and standards information. We then show how a network of collections of standards and standards information based on an Internet architecture can be viewed as a "virtual library of standards" by the users, while allowing

standards providers to control and maintain the ownership of these collections on a set of heterogeneous databases. Next, we focus on how the needs of users and developers can be supported through searching for standards via a common user interface. We conclude with some thoughts on support for standards development on the information highway and other related issues.

2. Background

The current U.S. voluntary standards system consists of a number of diverse, private sector organizations involved in the development, production, and delivery of standards. In 1992, recognizing that progress in readily identifying, accessing, and using these standards is of strategic significance to U.S. industry, ANSI's Standards Data & Services Committee (SDSC) developed a strategic plan for the electronic development, production, and delivery of standards, which became the basis for the current NSSN project. The NSSN Project [2] is chartered to develop an electronic link connecting all standards developers, information providers, and users of standards and standards information.

This two year project was funded in the fall of 1994, through NIST to ANSI, as a Technology Deployment Activity (Extension Enabling Services) of the Defense Dual-Use Assistance Extension Program of the Technology Reinvestment Project (TRP). The first year effort is focused on developing the underlying documentation for user requirements and functional specifications for the NSSN. The second year effort of the project develops a prototype demonstrating the technology necessary to accomplish the overall task of providing standards information across electronic links. The NIST test bed activity supports the ANSI NSSN project and its charter through the evaluation of promising technologies and the development of proof-of-concept implementations.

3. Electronic Access to Standards

In this section, we give an overview of current techniques for access to standards and then describe trends in electronic access that are relevant to the standards domain.

State of the Art for Access to Standards

Many standards collections are in electronic form, existing primarily as CD-ROM image databases with a search capability on title, author, and subject fields. Some collections, however, are stored in full-text, digitized form. Many are still available only in paper form or on microfilm with, perhaps, some electronic catalog listing. To locate a standard a typical user will go to a library facility, find a catalog (possibly in electronic form), consult some reference material, or request

the standard through an information service which compiles and maintains collections of standards. On-line information services and on-line catalogs often provide a search capability based on a database of pre-defined keywords or abstracts. Standards are produced with limited electronic support, such as electronic mail, teleconferencing, individual word processing, and fax transmissions.

Recently many standards organizations have begun converting their documents into electronic form. For example, the American Society for Testing and Materials (ASTM) has invested in an extensive re-engineering effort to place all their documents in electronic form and re-design their business process accordingly [9]. Some organizations, such as ANSI, the Institute of Electrical and Electronics Engineers (IEEE), and the International Organization for Standardization (ISO), have implemented World Wide Web (WWW) or gopher sites on the Internet making it possible to peruse their charters and catalogs on-line. The Defense Information Systems Agency Center for Standards has collected a set of defense related standards and has made them available through the WWW and an electronic bulletin board. Many standards, primarily dealing with the Internet, are freely available via file transfer, gopher, or WWW pages. The U.S. government has sponsored the Government Information Locator System (GILS) project to help the public access electronic information throughout the U.S. Government. Information delivery organizations, such as Information Handling Services [4] and the Document Center, maintain collections of standards in repositories and catalogs of these collections are available in electronic form on the WWW.

Trends in Electronic Access

Electronic access to information is rapidly gaining widespread acceptance [1]. The Internet has doubled in size every year since 1988, connecting more than five million computers. Digital storage cost is going down relative to the cost of library shelf space and electronic services are becoming more useful, affordable, available and usable [5]. With the advent of World Wide Web multimedia access, electronic "pages" of information are readily available through user-friendly browsers such as Mosaic and Netscape Navigator. Applications on local and isolated machines are moving into a distributed network environment based on the Internet paradigm and, as such, will support relatively transparent network access to information.

In addition, information retrieval technology is a rapidly advancing area of both research and commercial development [7]. The ANSI standard Z39.50 of information retrieval protocols for accessing full text databases has been released [13]. The Government Information Locator Service (GILS) [6] has been initiated. Filters and translators for various word processing formats are available. Public standards such as Standardized General Markup Language (SGML) and Hyper

Text Markup Language (HTML) hold much promise for the flexible construction, storage, and display of documents across diverse platforms.

The infrastructure to support electronic commerce is falling into place, with secure network server software and billing techniques becoming commercially available. Efforts such as CommerceNet, Digicash, and "virtual shopping malls" are springing up on the Internet.

In essence, what we have seen developing in the 90's is a change in focus away from large, centrally stored and controlled databases of information to decentralized networks of servers with browse and search capabilities, where the responsibility of maintaining the integrity and security of the data is left with the owners of the data.

4. A Virtual Library of Standards

As recently as five years ago, there was much discussion of why scientific and technical libraries should maintain collections of standards and guidelines for maintaining such collections. Ricci [17] points out that standards collections "present a challenge because of the diversity of organizations which publish them, the variety of formats in which they are published, and a frequent lack of adequate description of the needed standard." We present the notion of an electronic, virtual library for standards as way to meet this challenge and achieve the NSSN objective of providing standards electronically on a timely and cost-effective basis.

Virtual Library Concept

A virtual library can be viewed as a distributed space of interlinking information sources or a collection of distributed information servers [5,14]. Locating information in a virtual library is easy if one knows where to look, how to ask, and the sources are well-structured and uniform. An "intermediary" [6] is often necessary to supply front-end access for a particular collection of information when it is stored in many locations and in many different forms. A well-designed intermediary interface can enhance the user's ability to browse and search, hide the underlying network structure, as well as provide other related services.

Virtual Library of Standards Architecture

Technical people are often burdened with the tasks of searching for and acquiring standards they can use. Traditionally, they must rely on a librarian, research specialist, or information service to

perform these tasks. The architecture shown in Figure 2 was designed to ease this burden by providing electronic access to collections of standards and standards information through a single access point. This distributed architecture of a virtual library illustrates how multiple collections

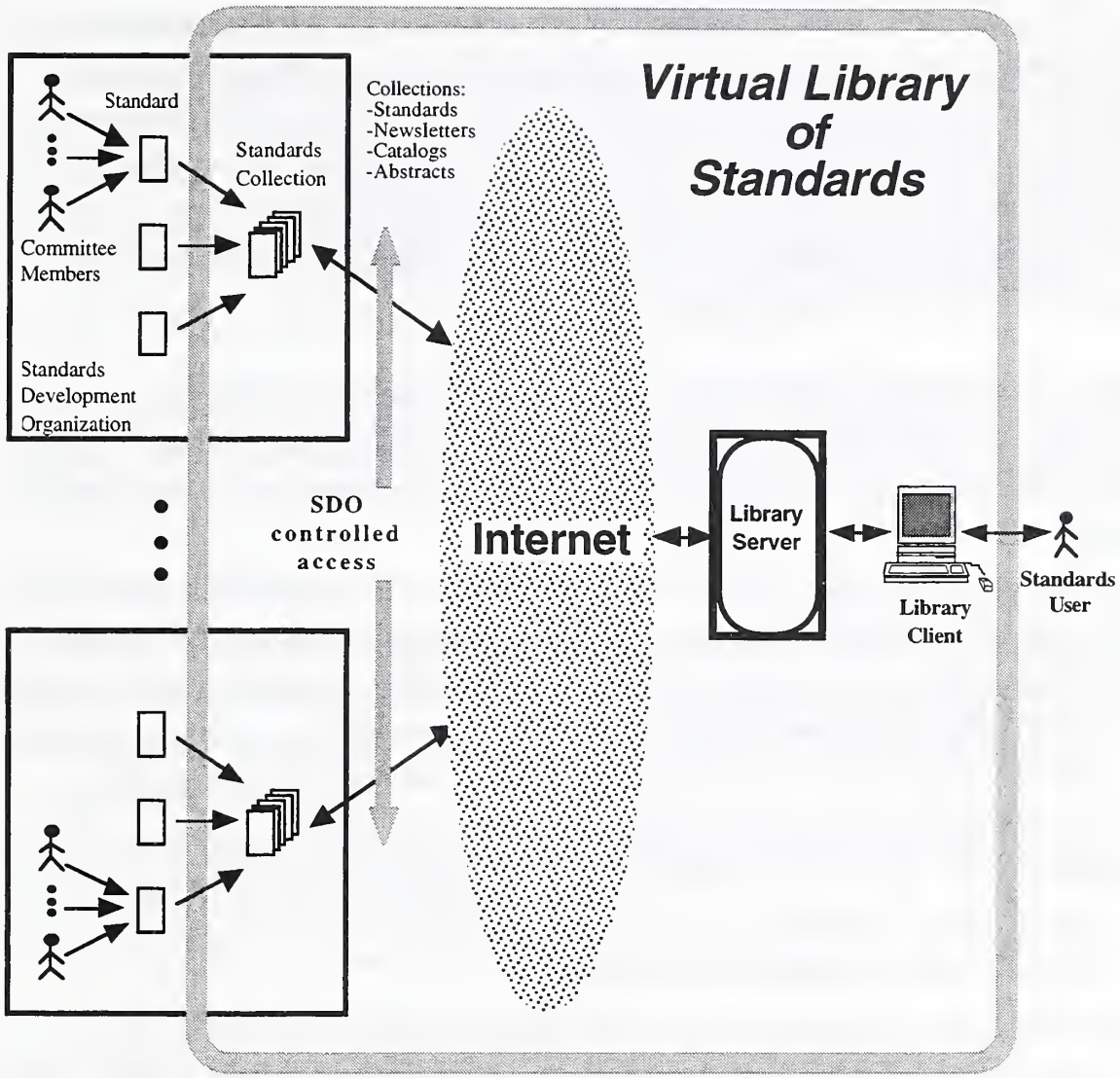

Figure 2 Virtual Library of Standards Architecture

of standards and standards information can be presented to the user as one virtual collection of searchable information. The library server can locate, retrieve, and deliver a standard, then charge for it electronically through the user interface. A well-designed interface can guide a novice user through the collection to confirm that a standard is the relevant one and also provide information about the status of the standard. An experienced user can still view a particular collection of

standards, standards information, catalogs, or abstracts by using the library server as an intermediary or by accessing a standards development organization (SDO) server directly.

While improving user access to their information, this architecture also allows SDOs to retain control of their standards. An SDO creates documents electronically which then reside in one of many independent standards collections. The standards development organization determines what level of access is allowable for that collection and then makes the associated documents and their access characterization available to the library server via the Internet. This architecture can also support functions for user notification of progress on standards in development, commercial transactions, and electronic authoring of documents, which are all important elements of the standards development and delivery process.

Mapping the Standards Library Concept onto the Information Highway

In this section, we list the six facets Ricci [17] used to describe a "Standards Information Service" and follow each with a short re-interpretation in the context of an Internet-based virtual library of standards:

- Acquiring and building a usable collection of standards -- In a distributed, Internet domain this implies that there must be links to searchable collections of standards and catalogs. Searching for standards should be supported by a common user interface, rather than many different interfaces and search engines on different servers. However, this common entry point should not prohibit a user from going directly to where the information resides.
- Obtaining standards documents as requested -- Since many standards must be purchased, there must be a secure way to browse for a standard, order, and pay for it electronically, or minimally to get information on how to order by phone, mail or fax. Response time and behavior should be predictable and consistent.
- Providing reference services -- Tools that access all information via a full text search on abstracts or complete documents combined with on-line thesauri, catalogs, subject indexes, and standards organization newsletters can provide a complete reference service.
- Maintaining the collection -- In a distributed WWW domain, all links to sites must be kept current and each owner of a collection of standards must continue to maintain it on-line. Software should automatically check links to documents and document servers and track obsolete standards and draft standards. Information brokers or intermediaries can maintain servers for standards organizations which do not have the desire to maintain their individual collections electronically.

- Documenting/cataloging -- From an implementation standpoint, it is crucial to have a unique identifier associated with each standard, so if a standard is moved or renamed at its location, its virtual location remains stable. All information sources should have descriptors containing their structure and access characteristics available to the common user interface [20].
- Encouraging use of standards in all appropriate applications -- Easy access via computer implies better, more extensive awareness and use of standards. Critical to the success of access via a virtual library is a user interface and search engine to support providing information in a concise, coherent fashion to the user. This virtual library view also enables individuals and organizations to create (and modify) their own virtual sub-collections for specific applications or projects as required.

As we have shown, each of these characteristics easily maps to the electronic domain. Next, we look at the specific user needs that have been identified for the NIST test bed analysis and evaluation.

5. User Needs

The NIST test bed design decisions were based on the user needs [15] collected under the auspices of the NSSN project for electronic access and development of standards. We have consolidated these needs into two categories: standards users and standards developers.

- Standards Users Needs
  - User interface with intuitive, on-line help, on multiple platforms, maintaining original page structure of the documents;
  - Single point, seamless access to all standards with transparent access to content;
  - Full-text, parametric, and data-type searching with help from user-profiling and thesaurus tools, all independent of the location of the standards;
  - National and international standards, status and history of the standards, newsletters, technical documents and other documents;
  - A common naming convention, links between related information, and searchable, manipulable content coding of standards; and
  - Support services such as a help desk, documentation, training, proactive alerts and "single-point" billing.

- Standards Developers Needs
  - Electronic support for the standards development process, such as communication, authoring, and workflow management software;
  - Easy transfer of information with common, consistent content tagging; and
  - Protection against unauthorized modification and preservation of intellectual property rights and revenues [21].

Keeping these priorities in mind, the NIST test bed effort chose to begin its investigations based on the user needs with a focus on the user interface and search components. A separate test bed effort is being initiated to address the user needs related to the standards development process.

6. Test Bed Investigations

The main role of the NIST test bed [10] is to test and evaluate concepts and related technologies and products for standards information and document development, storage, retrieval, and delivery. The test bed performs these functions by conducting proof-of-concept experiments, typically in the form of rapid prototyping of software integrated with state-of-the-art hardware.

Test Bed Architecture

The test bed architecture is based on the assumption that any virtual library of standards will be supported by a group of distributed systems that implement a set of "agreed to" protocols to facilitate information interchange among the participating systems. A "client-server" approach is typical of applications that require distributed network access to information and many new software tools are based on this paradigm.

Our design decisions are also based on the assumption that as little an overhead burden as possible be placed on the individual servers containing the document collections. This implies that the recommended software and hardware be inexpensive, the document collections be easily convertible to searchable collections, and the maintenance of these collections, such as the checking of valid hyperlinks and updating of search indexes, be automated to a large extent.

The Internet satisfies the above criteria with its basic functions: electronic mail, remote login, file transfer, WAIS, WWW, and others. Information servers are accessible directly or through the test bed server via any client WWW browser (or other access methods permitted by the

information server), on a variety of platforms. The test bed, depicted in Figure 3, is an Internet-based client-server architecture accessible by WWW graphical browsers, such as NCSA Mosaic, Netscape Navigator, and Lynx for text-only browsing in UNIX. The initial platforms for both the clients and servers are assumed to be computer workstations with graphics capability and include UNIX, Macintosh, and PC-based machines.

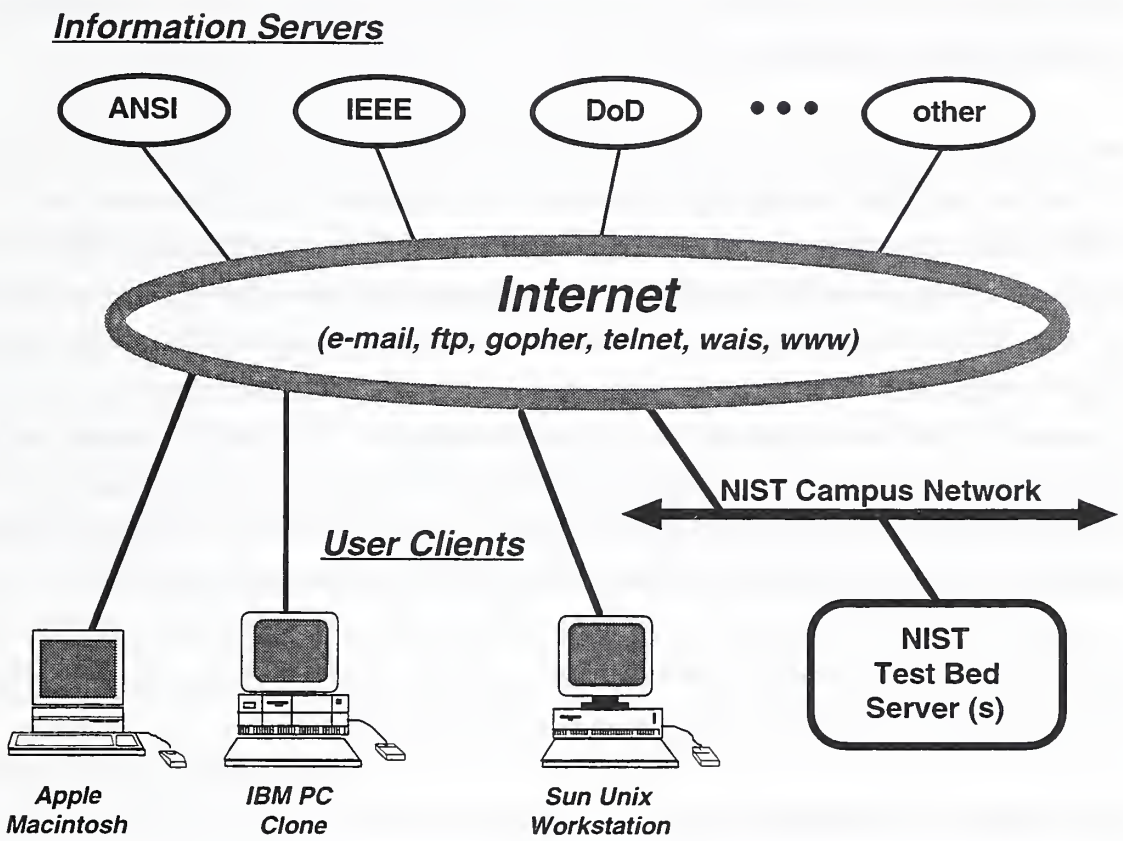

Figure 3 NIST Test Bed Architecture

Initial Investigations

The test bed effort began in earnest in March of 1995. Our choice of issues to tackle first was based on the view that easy-to-use searching for information is a critical function in the virtual library paradigm. Thus, the first areas of investigation are searching and user interface design. We are also concentrating on text-only databases, since the widely available information retrieval techniques are based on keyword search on text databases, though the browsers can handle multimedia information.

User Interface Design and Usability Considerations

For virtual, single-point user-client access, the test bed server user interface[1] should provide:

- a unified help facility with consistent information about the system and standards;
- a consolidated set of pointers, with search and retrieval capability to other standards-related information; and
- a focused search and retrieval capability to databases of standards.

This functionality permits the user to see the same interface regardless of the standard location, format, or server platform.

One important aspect of the test bed activity is the determination of the usability [11] of various interface designs. This is done in conjunction with a set of users, user requirements and recommended design practices and guidelines, such as the HTML guidelines from National Center for Supercomputing Applications (NCSA) and Nielsen's discussion of how SUN's new home page was designed [12]. These practices include: consistency throughout the interface, Web (WWW) pages that fit on one screen, and icons that intuitively match their function. In addition to the look and feel of the user interface, it is important to evaluate other usability criteria such as performance. Usability testing with a select set of users, throughout the development cycle (formative evaluation), will provide an opportunity to collect user feedback about the organization of the information and the ease-of-use early in the specification development process.

Some of these usability design considerations are currently being implemented on a set of test bed Web pages that form a prototype single-point access user interface.

Test Document Collections

The test bed is primarily concerned with documents that are either electronically accessible, or, at least, are referenced by a document in electronic form. They contain information about a standards provider, about standards, or are standards themselves.

The test bed effort has acquired a set of non-proprietary test documents in different formats. These include:

---

[1] At the other end of the spectrum, the server would simply be a "launching pad" of hyperlinks to other servers. This design imposes considerable burden on the user, requiring the user to have a knowledge of the location and contents of the servers. The user may also have to adjust to the differences in the feel and structure of the user interfaces associated with the standards servers.

- Several Federal Information Processing Standards (FIPS) in ASCII and Word Perfect format which were converted to HyperText Markup Language (HTML) -- These were previously available on-line for browsing or downloading with no keyword search capability.
- American Petroleum Institute catalog of standards -- These were extracted from a PC-based standalone application, split into title and abstract segments and then placed in HTML format.
- A set of General Motors standards in Standardized General Markup Language (SGML) -- A simple conversion put these in HTML format.

In addition, Universal Resources Locators (URLs) for WWW pages of numerous standards organizations and standards themselves, currently accessible on the Internet, were also collected. Links to these collections have been placed on a test bed Web page with full text, key word search capability.

Longer term experimentation with document collections on the test bed is planned in the areas of: optical character recognition, representation and searching of images and figures, and hyperlinked text. This would give some insight into the level of effort needed to bring older standards databases into newer and more accessible formats, as well as provide an opportunity to identify what type of user interface and data management techniques are needed to deal with complex documents.

Standards Information Servers

The standards servers, where the actual documents reside, must use a common set of protocols, for example ANSI standard Z39.50, in order for the test bed library server to query these databases. The concept is that a standards server, upon receiving a query from the library server, would translate it into the local query language, do the searches using the local search engine, convert the retrieved information into the protocol format, and send the information back to the library server. This approach requires that the standards server possess a basic search mechanism and understand the protocol. Because the servers will be providing information at various levels of detail, the data also needs to be tagged with indicators such as access restrictions, format, and revision status [20]. We are implementing this approach with the document collections described earlier.

Search Capabilities

Efficient search strategies and intelligent search engines are needed to cope with the distributed nature of the document servers and the complex architectures of compound documents, such as standards documents with graphics, tables, and mathematical equations. We are experimenting with various search strategies and also with different search engines, such as WAIS and NIST's PRISE[2] .

We are implementing these searches in several modes. The first mode is direct retrieval from the document collections described above. Here the user specifies the database and the keywords. Other collections can be searched, but the user must jump to those links and use that search interface. This is a baseline that allows us to compare several search engines and interfaces.

For the second mode, indexes for all searchable databases reside in the test bed library server (for efficiency). This allows the test bed to automatically determine which collections to search and retrieve from. We envision that this access method, if successful, would become the preferred access path. In the long term, it will enable the test bed server to provide, for example, a uniform, consistent interface, help for a novice user, a thesaurus, and user updates via email about changes in standards. The user need only construct keyword queries in relatively "natural" language with the help of the server to locate a standard. We are experimenting with several implementations to test for performance and usability.

A third mode is that of searching for information about standards development organizations, newsletters, and other reference material. In our initial implementation, we have a software "spider" or "webcrawler" tool that automatically follows hypertext links in documents several levels down, checks for inactive links, and indexes the documents for searching. For example, the user could ask about "IEEE membership" and be referred to IEEE's Web site where IEEE maintains information on how to become an IEEE member.

Use of Standards in Test Bed

Wherever possible we have applied the appropriate standards to the test bed itself. These include the ANSI standard Z39.50 information retrieval protocols, SGML, HTML, and other standards. We are reviewing other standards from ANSI, ISO, and NISO for content coding and for designing abstracts, indexes, and other elements needed to improve search capabilities.

---

[2] The PRISE system is typical of the advanced, statistically-based full text search engines currently available, while WAIS uses a simple word occurrence count to determine if a set of keywords matches a document.

7. Future Investigations

We have described how the NIST test bed effort is contributing to the evolution of an architecture to support access to standards on the information highway by viewing the domain of standards as a virtual library. Follow-on test bed investigations will include:

- more complex documents;
- advanced document content and structure searching;
- automatic maintenance of hyperlinks and indexes;
- collaborative authoring tools; and
- security and billing.

We view the investigation of groupware for authoring as the next critical issue that the test bed should address. Standards developers critically need support for on-line development of standards and other documents, and for the management of the document development activity [18]. The IEEE SPA System [8,16] is being constructed to support the on-line authoring of standards and provides an array of support tools for the authoring of standards that include email, bulletin boards, and editors. Software tools for collaborative authoring over a network and for managing the work process are moving from the research environment into commercial applications. Tools exist for many of these functions, but not in the unified context of the Internet, standards collections, standards committees across organizations, and group (as opposed to single) authoring tools. The test bed will be exploring software architectures that permit the integration of these functions into a cohesive unit.

We recognize that billing and security are critical to an operational system running on the Internet. After the initial test bed objectives are met, we will explore the options that are available to satisfy these and other production system requirements. Tools being developed for electronic commerce, such as those from CommerceNet and the NETBILL project, will be evaluated for inclusion in this later exploration.

8. References

[1] Anderson, C., "The Accidental Superhighway: A Survey of the Internet", *The* [*Economist*, vol. 336, no. 7921, July, 1995, pp. 1-18.

[2] "National Standards Systems Network", TRP Proposal, American National Standards Institute, New York, NY, July 23, 1993.

[3] Anthes, G.H., "Proposed National Standards Network to Ease Development", *Computer World*, vol. 28, no. 35, August 29, 1994, p. 28.

[4] Belfrage, R. M., "Standards Information Services--The IHS Solution", *Science and Technology Libraries*, vol. 10, no. 3, 1990, pp. 83-90.

[5] Fox, E. A, et. al., "Digital Libraries, Introduction", *Communications of the ACM*, Vol. 38, No. 4, April, 1995, pp. 22-28.

[6] "Government Information Locator System Report to the Information Infrastructure Task Force", May, 1994. Available in electronic form at http://www.usgs.gov/public/gils/gilscopy.html.

[7] Harman, D. K., "Overview of the Third Text REtrieval Conference (TREC-3)", pp. 1-19, in Overview of the Third Text Retrieval Conference (TREC-3), D. K. Harman, ed., NIST Special Publication 500-225, National Institute of Standards and Technology, Gaithersburg, MD, April, 1995.

[8] *IEEE SPASystem Handbook for Standards Developers*, V. 2.0, available in electronic form at http://stdsbbs.ieee.org:70/0/pub/ieeestds.htm.

[9] Larned, B. E., "ASTM's ASPIRE Project Moves Ahead", *ASTM Standardization News*, vol. 22, August 1994, pp. 50-53.

[10] Laskowski, S., Schanzle, C., White, D., and Ramayya, V. "National Standards Systems Network (NSSN) Test Bed Plan for Year I", NSSN Project, May 23, 1995.

[11] Nielsen, J., Usability Engineering, Boston: Academic Press, 1993.

[12] Nielsen, J., "Interface Design for SUN's WWW Site", *SUN on the Net*, July, 1995. Available in electronic form at http://www.sun.com/cgi-bin/show?sun-on-net/uidesign/designstory.html.

[13] ANSI/NISO Z39.50-1995 - Information Retrieval Application Service Definition and Protocol Specification for Open Systems, National Information Standards Organization (NISO), Bethesda, MD, 1995

[14] "Framework for National Information Infrastructure Services", NIST Internal Report 5478, National Institute of Standards and Technology, Gaithersburg, MD, July, 1994.

[15] User Needs from SDSC - Informal NSSN Project Document.

[16] Read, W. S., "Streamlining the Standards Development Process", *IEEE Computer*, vol. 27, no. 11, November, 1994, pp. 93-94.

[17] Ricci, P. L., "Which Will It Be? A Quality Standards Information Service or a Collection of Standards?, Science and Technology Libraries, vol. 10, no. 3, 1990, pp. 3-35.

[18] Spring, M. B., Grisham, G., O'Donnell, J., Skogseid, I., Snow, A., Tarr. G., and Wang, P., "Improving the Standardization Process —From Courtship Dance to Lawyering: Working with Bulldogs and Turtles" to appear in *Standard Policy and Information Infrastructure*, B. Kahin and J. Abbate, eds, MIT Press, 1996.

[19] Taninecz, G., "Standard-Bearer for Locating Standards", *Industry Week*, May 1, 1995, pp. 53-56.

[20] Thomas, R. and Sicker, A., "Standards Information Tag Data Dictionary", Nov. 1994, unpublished manuscript.

[21] Weiss, M. B. H. and Spring, M.B., "Selected Intellectual Issues in Standardization" Paper presented at the Twentieth Annual Telecommunications Policy Research Conference, Solomons, MD, September 12-14, 1992.