

Building the Interspace: The Illinois Digital Library Project

To appear in *CACM (Communication of the Association for Computing Machinery)*, April 1995 special issue of Digital Libraries

The University of Illinois is building a large-scale digital library testbed, planned to grow to thousands of users and thousands of documents, with the goal of bringing professional quality search and display to Internet information services. Concurrently, research is designing and implementing a prototype of the Interspace, a vision of what the Internet will evolve into, where the distributed network of interconnected machines is replaced by a distributed space of interlinked information.

The testbed collection consists of articles from engineering and science journals and magazines, obtained in SGML format directly from major partners in the publishing industry. This collection will be managed by the University Library on a production basis, growing into a standard service of the new Grainger Engineering Library Information Center.

The testbed software will support comprehensive search and display of complete contents of articles, including text, figures, equations, and tables. The software is based on NCSA Mosaic as a multi-platform World-Wide Web connection to commercial software, currently SoftQuad Panorama for SGML display and Dataware BRS for fulltext search.

The National Center for Supercomputing Applications is developing a custom version of their Mosaic for this testbed with sufficient client and server interfaces and gateways (e.g. CCI and Z39.50) to bring professional display and search to widely deployed Internet information services.

The testbed users will be faculty and students at the University of Illinois initially, then spread to the CIC consortium (Big Ten universities).

The user evaluation will interview hundreds of users in focus groups to provide detailed cognitive descriptions of needs and uses, plus survey thousands for a grosser statistical picture. The software will also be instrumented to learn how to deduce more detailed ethnographic information from large-scale network usage.

The technology research efforts are centered around scale and functionality, and will migrate into the testbed as they prove effective. Providing semantic retrieval at a deeper level than commercial search is necessary to support wide ranges of users across wide ranges of collections.

The utility of physical library classifications for networked digital collections is being investigated by experimenting with interfaces using major classification schemes (e.g. Dewey Decimal, Library Congress, INSPEC thesaurus). A complementary effort is to generate classifications automatically. This will be tested using a concept space approach based on co-occurrence matrices, which has proven effective in specific domains such as molecular biology.

The research efforts will be assembled into a next generation system based on a new architecture for the Interspace. This will provide an environment for an information space of structured objects across the network. Such an environment will enable information sources and services (data and programs) to be plugged into the space, while supporting interactive functionality. For example, a user could execute an equation from an article while browsing or record a navigation path through the collection to share with others. Support of analysis and communication for programs and people will provide a new level of functionality for network information systems. For more information see: <http://www.grainger.uiuc.edu/dli>

Building the Interspace: The Illinois Digital Library Project (University of Illinois at Urbana-Champaign)

The University of Illinois is building a large-scale digital library testbed, which will make research-oriented information collections available via robust, state-of-the-art full-text search and retrieval database technologies accessed through intelligent multimedia interfaces. This testbed of online journals, obtained from professional society and commercial publishers, will be accessible online via TCP/IP networks to an academic community of thousands of potential users. Over the course of the project, the usability, accessibility, and value of the testbed documents will be enhanced through the use of such techniques as interactive linking between documents, semantic retrieval algorithms, dynamic document annotations, etc. Throughout the project, ongoing research into the technology and sociology of the testbed and how it is used will provide a better understanding of the dynamics, economics, and potential benefits of future Digital Libraries and a clearer picture of how such systems may evolve and scale to become major components of the National Information Infrastructure.

The testbed collection will consist of articles from engineering and science journals and magazines, obtained in SGML format directly from major partners in the publishing industry. Current committed partners include: IEEE Computer Society, APS (American Physical Society), AIAA (American Institute of Aeronautics and Astronautics), ASCE (American Society of Civil Engineering), IOP (Institute of Physics), and John Wiley & Sons. This collection will be managed by the University Library on a production basis, growing into a standard service of the new Grainger Engineering Library Information Center. A prototype client-server database and interface will be developed and tested for the Windows environment. From this prototype, a customized version of the NCSA

Mosaic software will be developed at the National Center for Supercomputing Applications, building on and utilizing related state-of-the-art software development work being done by corporate partners including SoftQuad, Spyglass, and Microsoft.

Research, based in the University of Illinois Graduate School of Library and Information Science and involving researchers from the MIS department of the University of Arizona and elsewhere, will encompass sociological evaluation of the testbed and prototype design of future scaleable information systems. The end-result of this collaboration will be a blueprint for advanced interactive library and information retrieval systems which will help future researchers exploit the 'Interspace', a distributed space of interlinked information. The Interspace, an environment of structured objects across the network, will enable disparate component data and programs to be dynamically plugged into the space while maintaining interactive functionality.

For more information see: <http://www.grainger.uiuc.edu/dli>