Outline for Digital Libraries Tutorial at SIGIR'96

TUTORIAL TITLE: Building and Applying Digital Libraries

INSTRUCTOR: Edward A. Fox

SCHEDULE: Sunday, August 18, 1996, 2:00 p.m. - 6:00 p.m.

PARTICIPANTS: Level: introductory to intermediate

Participants are encouraged from those working in fields like computing, library science, information science, and publishing. Those with little background on digital libraries are invited.

COURSE DESCRIPTION:

Objectives:

Attendees should be able to help in design, development, evaluation, and standardization efforts related to digital libraries. They should understand the key aspects of representative digital library projects, as well as their successes, failures, and implications for the future.

Planned Activities:

1. Participants will use Sun workstations (alone or in small groups), working with materials on the WWW, with the instructor coaching. Learning/reference modules developed by the instructor, in conjunction with Virginia Tech courses and an in-process book on digital libraries, will be the basis for this session.
2. Content will be drawn from many SOURCES. There will be in-depth coverage of the following PROJECTS:
    - TULIP (material science & engineering, with Elsevier, OCLC)
    - CS technical reports (WATERS, NCSTRL)
    - ETD - electronic theses and dissertations
    - Envision - CS literature
    - Interactive Learning with a Digital Library in CS - courseware
    - Monitoring, analyzing, visualizing and modeling WWW traffic
    - Applying IBM's digital library software
3. Participants will look at all of the relevant content from a number of PERSPECTIVES:
    - user and social needs;
    - interfaces and user interaction;
    - architectures, components, protocols;
    - content, publishing, and capture;
    - systems, engines, and operations.
4. Finally, participants will engage in discussion for the final part of the course regarding whichever of the following they find most interesting:
    - developing a world-wide digital library of theses and dissertations,
    - developing a world-wide digital library for computer science, or
    - another topic or topics selected by the class.

Introduction

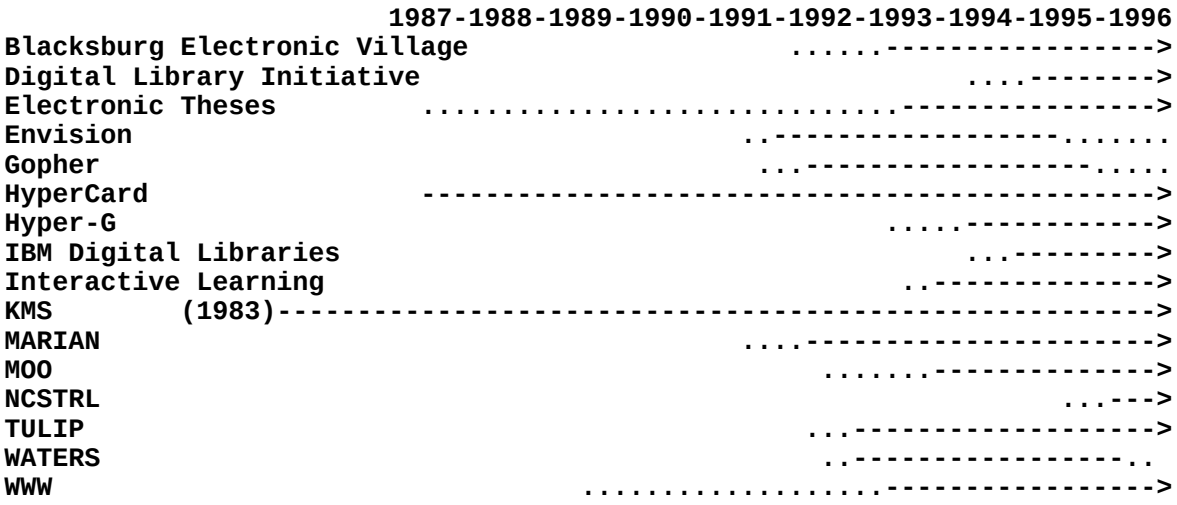


Figure 1: Timeline of Recent Information and Digital Library Systems

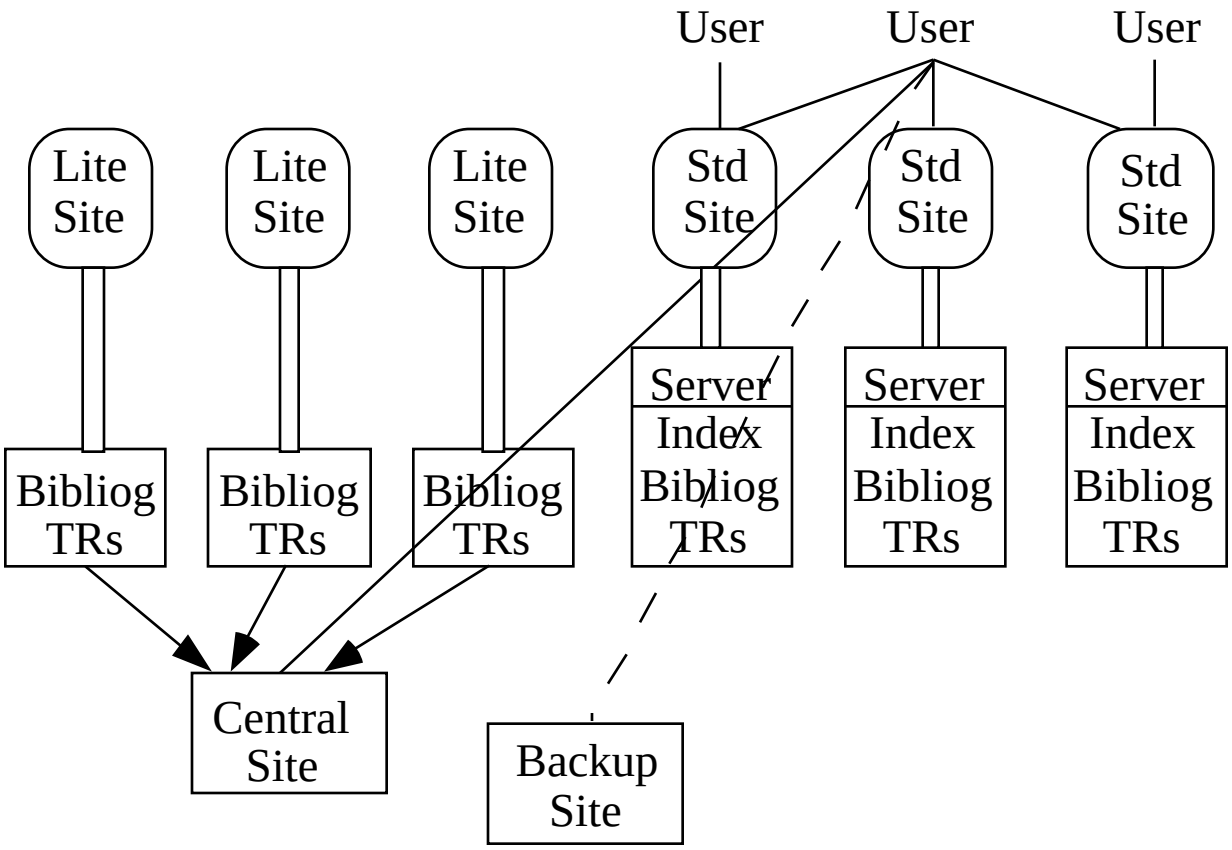


Figure 2: NCSTRL Architecture

TABLE OF CONTENTS

II. PROJECTS:

* Build upon existing electronic materials
  - Netlib [link] http://www.netlib.org/
    + Attribute/value search [link] http://www.netlib.org/utk/misc/netlib_query.html
* Build upon publishers' collections
  - CORE [link] http://ei.cs.vt.edu/~cs5604/DL/DL2.html
    + OCLC [link] http://www.oclc.org:5047/oclc/research/projects/core/
  - TULIP [link] http://www.elsevier.nl/info/projects/tulip.html
    + Elsevier + universities + OCLC (material science & engineering)
* Commercial services and systems
  - OCLC
    + SiteSearch [link] http://www.oclc.org/oclc/menu/site.htm
    + futures (part) ([link]) http://www.oclc.org/oclc/promo/9497site/9497.htm
  - IBM http://204.146.47.71:80/is/dig-lib/ [link]
    + case studies [link] http://204.146.47.71:80/is/dig-lib/dlis.htm
    + images - QBIC [link] http://wwwqbic.almaden.ibm.com/
    + rights management [link] http://204.146.47.71:80/is/dig-lib/dlfnct4.htm
* Enhance WWW (hypertext):
  - Hyper-G [link] http://ei.cs.vt.edu/~cs5604/Adv/Adv-Hyper-G.html
  - HyperWave [link] http://www.tu-graz.ac.at/0x811b0205_0x00071882;sk=D40F3456
  - HyperWave server features[link]
  - HyperWave authoring features[link]
  - HyperWave authoring specs [link]
  - Amadeus: collections with sizes, ranked text search ([link])
    http://ei.cs.vt.edu/~cs5604/Adv/Adv-Amadeus.html
  - Harmony orientation [link]
  - Harmony screens [link] http://ei.cs.vt.edu/~cs5604/Adv/Adv-Harmony.html
  - Harmony information structuring [link]
  - Harmony viewers [link]
* Community network multimedia history
  - BEV History [link] http://history.bev.net/bevhist/
    + Timeline [link] http://history.bev.net/bevhist/historyBase/mainTimeline.html
    + 1992 [link]
    + Article [link]
* Discipline - Computer Science
  - Technical reports
    + CSTR [link] http://WWW.CNRI.Reston.VA.US/home/cstr.html
    + WATERS
    + NCSTRL http://www.ncstrl.org/ [link]
      * Search results [link]
      * Search results abstract [link]
      * Doc. thumbnails ([link])
      * Doc. page 1 ([link])

* Discipline - Computer Science (continued)
  - Ptrs
    * DLs for CS http://fox.cs.vt.edu/DLCS.html [link]
  - Envision [link] http://ei.cs.vt.edu/~cs5604/Adv/Adv-Envision.html
    * Query screens ([link]) http://ei.cs.vt.edu/~cs5604/Adv/Env-q.html
    * Results screens [link] http://ei.cs.vt.edu/~cs5604/Adv/Env-r.html
    * Report [link] http://ei.cs.vt.edu/papers/ENVreport/final.html
    * Dienst http://researchsmp2.cc.vt.edu:8090/ ([link])
  - ILDLCS (Interactive Learning with a Digital Library in CS) [link]
    http://ei.cs.vt.edu/~cs5604/Adv/Adv-ILDLCS.html
    + Project management http://ei.cs.vt.edu
    + Tools: debates, QUIZIT, SWAN, Virtual Q&A
    + EI Courses [link] http://ei.cs.vt.edu/courses.html
      - CS5604 ([link]) http://ei.cs.vt.edu/~cs5604/
      - units ([link]) http://ei.cs.vt.edu/~cs5604/f95b/Units.html
        + DL unit ([link]) http://ei.cs.vt.edu/~cs5604/f95/U-DL/U-DL.html
      - CS4624 ([link]) http://ei.cs.vt.edu/~mm/
        + History ([link])
          http://ei.cs.vt.edu/CSNotes-cgi/infoBase/infoBase.
          pl?showPage+./001.sub/012.sub
        + Timeline ([link])
* Research - DLI [link] http://www.grainger.uiuc.edu/dli/national.htm
  - CMU (Carnegie Mellon U.)
    + Informedia [link] http://www.informedia.cs.cmu.edu/research/index.html
    + NetBill [link]
      http://www.ini.cmu.edu/NETBILL/publications/CompCon_TOC.html
  - Michigan
    + Agents [link] http://ai.eecs.umich.edu/people/jmvidal/papers/tpa/node2.html
    + Groups [link] http://http2.sils.umich.edu/UMDL/teams.html
  - Stanford [link] http://walrus.stanford.edu/~testbed/
    + Today's DLs [link]
      http://Walrus.Stanford.EDU/~testbed/testbed.slides/P002.htmls
    + Infobus [link] http://Walrus.Stanford.EDU/~testbed/testbed.slides/P003.htmls
    + COS - before[link] http://Walrus.Stanford.EDU/~testbed/cos/slides/P011.html
    + COS [link] http://Walrus.Stanford.EDU/~testbed/cos/slides/P002.html
    + COS - services list [link]
      http://Walrus.Stanford.EDU/~testbed/cos/slides/P015.html
    + COS event services ([link])
      http://Walrus.Stanford.EDU/~testbed/cos/slides/P023.html
  - UCSB (Santa Barbara)
    + tutorial [link] http://alexandria.sdc.ucsb.edu:3366/doc/tutorial/index.html
    + 1996 ([link]) http://alexandria.sdc.ucsb.edu/public-documents/annual-report/

- UIIC (Illinois)
    + interspace [link] http://www.grainger.uiuc.edu/dli/infrastr.htm
    + semantics [link] http://www.grainger.uiuc.edu/dli/semantic.htm

* National - Library of Congress
  - American Memory [link] http://lcweb2.loc.gov/ammem/ammemhome.html
  - Call [link] http://lcweb2.loc.gov/ammem/award/

* Genre - ETDs - electronic theses and dissertations
  - http://etd.vt.edu/etd/ [link]
  - Submission form [link] http://scholar.lib.vt.edu/cgi-bin/etd.cgi
  - Standards [link] http://etd.vt.edu/etd/faq/formats.html
  - Principles [link] http://etd.vt.edu/etd/principles.html

III. PERSPECTIVES:

* user and social needs
  - Allerton [see prior link]
  - D-Lib WG [see prior link]

* interfaces and user interaction
  - monitoring, analyzing, visualizing and modeling traffic [link]
    http://www.cs.vt.edu/~chitra/www.html
    + Sources of variation: daily, semester, local/remote
    + Excel graphs (18 months)
      (see http://ei.cs.vt.edu/~fox/EDMEDIA96)
        - Accesses [link]
        - Bytes [link]

* architectures, components, protocols
  - Z39.50
    + ptrs [link] http://ds.internic.net/z3950/z3950.html
    + Isite [link] http://vinca.cnidr.org/software/Isite/Isite.html
    + architecture [link] http://vinca.cnidr.org/resources/isite_overall.gif
  - CNRI
    + key issues [link] http://WWW.CNRI.Reston.VA.US/home/cstr/arch/slides.html
    + repositories of digital objects [link] http://www.cnri.reston.va.us/home/doa.html
  - agents
    + UMBC [link] http://www.cs.umbc.edu/agents/
    + Michigan [link]
  - bus: Stanford [see previous links]
  - systems, engines, and operations

* naming
  - PURLs [link] http://purl.oclc.org/
  - handles [link] http://www.cnri.reston.va.us/home/cstr/handle-intro.html

* library and information science
  - principles [link] http://ei.cs.vt.edu/~cs5604/DL/DL7.html
  - intellectual property rights
    + CMU NetBill [see prior link]
  - preservation [link] http://www-rlg.stanford.edu/ArchTF/
  - representations
    + Text Encoding Initiative [link] http://etext.virginia.edu/TEI.html
    + Amsterdam hypertext model ([link])
      http://ei.cs.vt.edu/~mm/gifs/Amsterdam-hm.html
    + Database: Berkeley
    + IR: Illinois [see prior links]
    + GIS: UCSB [see prior links]
    + Objects: Stanford [see prior links]
    + AI: Michigan [see prior links]
    + Multimedia: CMU [see prior links]
- information retrieval
    + CODER ([link]) http://ei.cs.vt.edu/~cs5604/Adv/Adv-CODER.html
    + MARIAN
      - overview ([link]) http://ei.cs.vt.edu/~cs5604/Adv/Adv-MARIAN.htm
      - page ([link]) http://opac3.cc.vt.edu/htbin/marian

IV. DISCUSSION:

* structuring principles
  - consider PERSPECTIVES
  - consider scalability, sustainability, usability
  - borrow from similar past/present projects
  - apply scenario-based design, claims analysis
  - summarize design, open problems
* developing a world-wide digital library of theses and dissertations
* developing a world-wide digital library for computer science, or
* topics selected by the class

V. RECOMMENDATIONS:

* Join projects: electronic theses and dissertations, CS technical reports
* Use our courseware and add to it
* Work toward “Open Digital Library”

ACKNOWLEDGEMENTS

Sponsors and Partners: Elsevier, IBM, NSF, OCLC, SURA
Co-PIs, Students, Staff
Projects
* NSF IRI-9116991, CDA 9312611, EID-9109853, CDA-9308259
* NCSTRL (ARPA, NSF)

Digital Libraries '95

The Second Annual Conference on the Theory and Practice of Digital Libraries

June 11-13, 1995 - Austin, Texas, USA

Full Papers

Interoperability, Scaling, and the Digital Libraries Research Agenda:

A Report on the May 18-19, 1995
IITA Digital Libraries Workshop
August 22, 1995

Clifford Lynch (clifford.lynch@ucop.edu)
Hector Garcia-Molina (hector@db.stanford.edu)

Converted to HTML using GradStudentWare 2.2
Contact Christian Mogensen with bug reports.

Introduction

This report summarizes the results of a workshop on Digital Libraries held under the auspices of the U.S. Government's Information Infrastructure Technology and Applications (IITA) Working Group in Reston, Virginia on May 18-19, 1995. The objective of the workshop was to refine the research agenda for digital libraries with specific emphasis on issues of scaling and interoperability, and to identify the infrastructure developments needed to make progress on these issues.

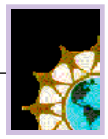
While there have been a number of workshops and other meetings examining the broader questions of support for applications in the National Information Infrastructure (NII), we believe this was the first workshop that focused specifically on Digital Libraries in this context. In the past year, Digital Libraries have emerged as one of the central and most compelling applications enabled by the NII; numerous digital library research projects are underway, including six large-scale pilot projects that have been funded jointly by ARPA, NASA, and NSF. While Digital Libraries are now a vibrant research area, and also a field in which considerable commercial development is taking place (presaging the future economic importance of Digital Library technology to the United States), many new questions are emerging as a result of this flowering of research

37th Allerton Institute 1995

Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

How We Do User-Centered Design and Evaluation of Digital Libraries: A Methodological Forum

This conference was sponsored by the National Science Foundation



The 1995 Allerton site is available via the EDFU Electronic Library

The Publications Office
Graduate School of Library and Information Science
University of Illinois, Urbana-Champaign

Last updated: 17 January 1996

Working Groups

One of D-Lib's principal activities is stimulating and supporting working groups that address aspects of Digital Library research. Some of these groups are created by D-Lib; some are affiliated with the Digital Library Initiative, or other federally funded projects; and some are independent groups.

The following working groups in Digital Library research are currently associated with D-Lib:

- Metadata to Describe Information in Digital Libraries
- User Needs Assessment and Evaluation
- Social Aspects of Digital Libraries
- Repository Interfaces
- Digitization and Conversion
- Naming Objects in the Digital Library
- Networked Computer Science Technical Report Library (NCSTRL)
- Task Force on Archiving Digital Information

D-Lib is also a sponsor of:

- The First ACM International Conference on Digital Libraries: Program and Proceedings

D-Lib is coordinated by CNRI and is sponsored by the Defense Advanced Research Projects Agency (DARPA) on behalf of the Information Infrastructure Technology and Applications (IITA) Working Group of the High Performance Computing and Communications (HPCC) program.

Home T-spotlight DL Research

wya/af/reb-a
Last revised: June 14, 1996

WORKING GROUPS

D-Lib Working Group on Metadata to Describe Information in Digital Libraries

Joint Chairs: Michael F. Goodchild, Terence R. Smith, University of California, Santa Barbara

If sense is to be made of the flood of information that will be available through digital libraries, it must be described effectively, so that it can be found, its value assessed, and its acquisition handled efficiently. Metadata is the term most often used to refer to the description of information objects to support these three functions of digital libraries. Digital library technology is capable of both supporting major augmentations to traditional metadata activities and providing a basis for catalog interoperability.

D-Lib is associated with two activities in this field. Both focus on the process by which creators of digital information can add metadata to their work at the time of creation. This metadata is then available for computer programs to use in building indexes and other access tools. It is also available as a basis for subsequent cataloguing or the creation of secondary information services.

The first of these activities comes out of the Alexandria Digital Library project at the University of California, Santa Barbara. This project concentrates on geospatial information, such as maps, but its studies of metadata are broad based and applicable to all types of on-line data. Alexandria is one of the projects in the ARPA/NSF/NASA Digital Library Initiative (DLI) and its metadata studies involve members of several of the other DLI projects.

The second activity is the Metadata I and Metadata II invited workshop series. The first of these was sponsored by OCLC and NCSA in March 1995, chaired by Stuart Weibel of OCLC. Its major contribution was the "Dublin Core" metadata elements. D-Lib has agreed to be a sponsor of subsequent workshops.

These two activities are inter-related. In particular, Alexandria is using the Dublin Core as a building block for its own developments.

wya/reb-a
Last revised: March 17, 1996

WORKING GROUPS

D-Lib Group on Naming in Digital Libraries

The D-Lib Group on Naming in Digital Libraries covers all aspects of naming of digital resources. This topic, which appears simple on the surface, proves to be remarkably subtle when applied to the complex world of digital libraries.

For several years, the Internet Engineering Task Force (IETF) was a focus for efforts to develop Uniform Resource Names (URNs). These are globally- unique, persistent, location-independent names that can be applied to any network resource. This work is being continued by an informal group of URN implementors. The focus of the D-Lib group is on the next stage, how to use names in large scale libraries.

User groups that wish to assign names to objects in a digital library are faced with a variety of issues. One type of question is the relationship of names to semantic concepts such as uniqueness, mutability, etc. Are these managed by the naming system or by an external system? In a large library, rules and conventions for assigning names can be very complicated. If users are to see the names, it is helpful if they have some structure to help them be remembered or recognized, but there are real dangers in attempting to embed semantic information into names.

There will be many naming schemes. Some, already exist and must be merged into the digital library. The integration of naming schemes is a technical challenge and an organizational one, requiring decisions about the registration of naming schemes, and the allocation of top-level names.

Few digital objects exist by themselves. They are parts of larger groups or made up of many components. Naming such complex and compound items proves to be intimately connected to questions of what metadata to keep for each component and how to represent the relationships between them. Proposed solutions include composite objects (which contain several separate objects) and meta-objects (which provide links to other digital objects).

Finally, all questions of naming must consider scale. Processes for naming and organizing small numbers of objects may be totally inadequate for large collections.

Home W Groups

wya
Last revised: February 5, 1996

WORKING GROUPS

D-Lib Working Group on Repository Interfaces

Chair: William L. Scherlis, Carnegie Mellon University

This working group focuses on technical issues associated with repository interoperation. As digital libraries proliferate, many approaches to managing digital assets and associated meta-data are emerging. There are important differences among these approaches, and these differences have technical, legal, social, economic, and political dimensions. How can multiple repositories coexist and interact effectively?

The working group is motivated by several important trends: The complexity and semantic richness of objects and meta-data managed by repositories is increasing. Information objects of greater value are now being managed more routinely, raising issues of security, access control, and support for commerce. Performance demands are increasing, as is the quantity and size of information objects, particularly in multimedia applications. Digital libraries are interacting more often with personal, group, and wide area information services. Finally, the distinction is blurring between digital libraries and other institutional information resources such as databases and corporate webs.

The starting points for the working group are technologies that support management of information objects, their names, and associated meta-data-databases, distributed file systems, object bases, and the Web. Several digital library research groups have started to develop concepts that could provide a basis for repository interoperation, including the CS-TR architectural work of Kahn and Wilensky, the Stanford Infobus project of Garcia-Molina and Winograd, and the agent architecture of the Michigan DLI project. In addition to the need to reconcile these various approaches, there is a broader need to put them in the context of standards efforts in the wider community, including Web-associated standards, CORBA, OLE, z39.50, and SQL and its successors. All of these deal with resolving names to objects, and all deal in some measure with meta-data.

The initial effort of the working group is (1) to identify the dimensions of the space of repository interaction and interoperability, and the issues associated with achieving some transparency for users of the digital libraries, and (2) to assess current research and development efforts to understand the differences among them.

Home W Groups

wya/reb-a
Last revised: February 5, 1996

WORKING GROUPS

D-Lib Group on Social Aspects of Digital Libraries

I. UCLA-NSF Workshop on Social Aspects of Digital Libraries

An invitational workshop was held at UCLA, February 15-17, 1996; 32 researchers, developers, and practitioners, 9 UCLA faculty facilitators, and 6 UCLA graduate research assistants participated. All materials from the workshop, including schedule and agenda, list of participants, participants' discussion papers and biographical statements, and summary reports presented at the meeting are available on the web site (http://www.gslis.ucla.edu/DL/).

We selected two research areas, each with three sub-topics, as focal points for a two-day workshop:

Information Needs: Identifying real information needs and developing digital libraries to meet those needs.

- Social context and culture
- Information needs and information seeking
- Linking user-learner needs and behavior to digital library design

End user searching and filtering: Designing digital libraries in which it is possible to find the right information in a glut of information.

- Organization, description and representation of information
- Search capabilities for users
- Interface design for information retrieval

II. Results of the workshop

While we bounded the scope of the workshop to provide a starting point for discussion and a set of criteria for selecting participants, our participants quickly expanded those boundaries.

The boundaries expanded in several directions:

- Level of analysis: Our scope, as stated in the background paper (see web site), focused on the needs and activities of the individual user. While important, we must recognize that individuals do not work with information resources in isolation from their communities. They perform individual tasks in the context of their work teams, classroom, and other social organizations. Many tasks are performed in group contexts; we must consider CSCW and collaboratory environments as well. Multiple levels of analysis are required.

- Scope of analysis: Our scope addressed information searching and retrieval processes. While important, we must set searching in the context of the cycle of information creation and utilization. People will create information in digitized form that becomes part of digital libraries and need tools and functional capabilities for doing so. They will search for information created by other people, and for purposes other than those intended by the creators, requiring a variety of searching functions. Once located, they will incorporate new information into other products and processes that become part of the life-cycle. We need consistent means to organize, describe, represent, and dispose of information throughout these activities and processes.

- Content vs. process: Our scope addressed digital libraries as a set of digitized resources and associated technical capabilities for searching for information, which is roughly the scope defined in the digital libraries initiative. This scope statement addresses the digitized content of digital libraries but does not recognize the social processes around digital libraries -- the "library" in digital libraries. We need to

d-lib magazine

An Agent-Based Architecture for Digital Libraries

William P. Birmingham
The University of Michigan
Electrical Engineering and Computer Science Department
School of Information Science and Library Studies
Ann Arbor, MI 48109
wpb@eecs.umich.edu

D-Lib Magazine, July 1995

d-Lib forum d-Lib magazine

Introduction

One of the most exciting promises of digital libraries is access to a great variety of information and *services* that transcend what is available today through on-line services, such as the World-Wide Web (WWW). A library is more than just stacks of materials on shelves; it is also highly trained people that provide valuable services. These services include such things as *organization and cataloging*, research, notification of new publications, and so forth. Indeed, one of the greatest assets of libraries are these high-valued services. The WWW, while it probably contains more information than any single traditional library, is arguably not as useful as a traditional library because it lacks these services (particularly organization and sophisticated search support). No one is dismantling their libraries because of the WWW yet. The University of Michigan Digital Library Project (UMDL) [1,2] believes that a successful digital library needs to provide both access to a wide variety of valuable content and services.

Because the range of both content and services that are possible for a digital library are potentially large (we cannot even imagine what will be available or needed in the future), there will be no single, complete digital-library solution. Rather, we expect that as editing tools become better and access to networks becomes easier and cheaper, there will be millions of content suppliers; "everyman" can become a vanity press on the information superhighway. We believe that the days of centralized suppliers of information (e.g., large publishing houses and traditional libraries) are numbered, and that the traditional notion of a "collection" will

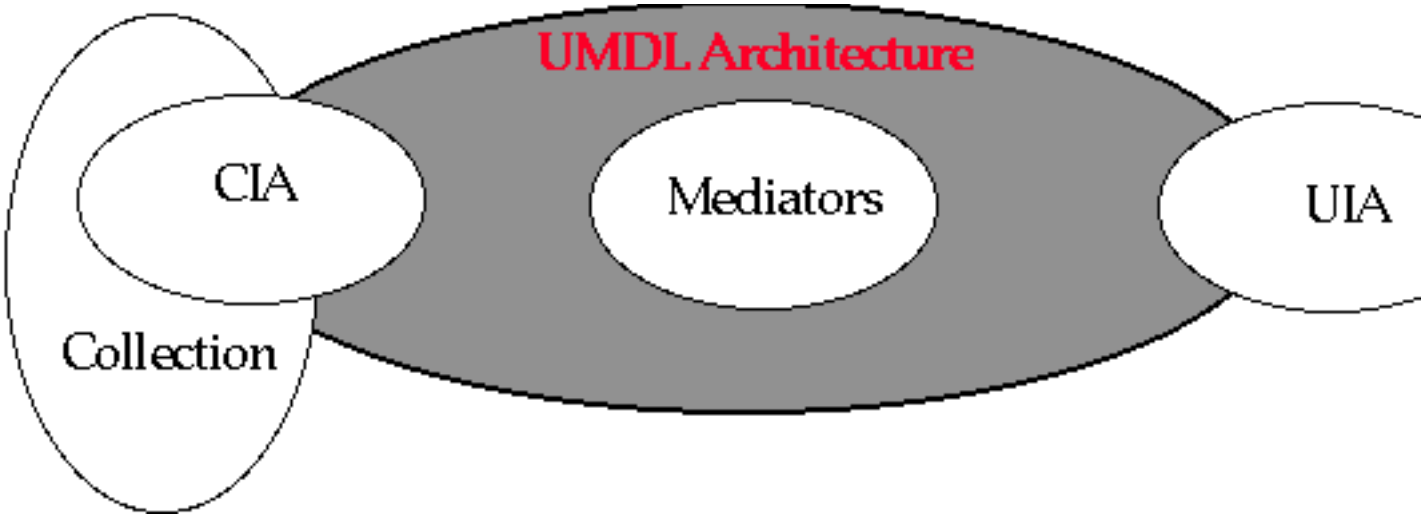


Figure 1: UMDL agent types

As the architecture is developed, the broad classes of agents depicted in Figure 1 will be continually refined; specialized agents will be added to the system as needed (the modularity property). For example, we can create user interfaces that are customized to a particular class of users, rather than to a particular collection or access mechanism (e.g., Boolean search over controlled vocabulary). In addition, the ability to *team* agents (as described in the next section, "What the architecture provides") dynamically creates new services with new agents, which is especially important since we anticipate the agent population will be constantly changing.

What the architecture provides

From a user's perspective, the types of high-level support that make a digital library worth using, such as searching, will be performed by a team of agents. For example, consider Figure 2, where a user (through the UIA) is searching for all articles by "Joan Q. Publique". Assuming that all agents have registered with the registry agent, the UIA contacts a query planner by first requesting the registry for a query planner that knows about author searching. The query planner then goes to the registry to get the addresses of a name authority (meta data that gives variations of Joan Q. Publique) and a name index (a partial listing of collections that contain works sorted by author). The planner then interrogates the authority, and then the index, finally determining the address of a particular collection. The collection is then accessed by the UIA using a protocol specific to the CIA.

It is easy to image how this process can be extended for different types of search by adding new types of agents (e.g., subject indexes and new kinds of query planners). The teaming methods gives the architecture a dynamic planning ability[5] that is critical for finding the best way to perform some service, as well as easily incorporate new types of search methods. There is, however, a cost.

d-lib magazine

Key Concepts in the Architecture of the Digital Library

William Y. Arms
Corporation for National Research Initiatives
Reston, Virginia
warms@cnri.reston.va.us

D-Lib Magazine, July 1995

Introduction

For the past two years, the Computer Science Technical Reports project (CS-TR) has been developing an architecture for a digital library with funding from the Department of Defense's Advanced Research Projects Agency (ARPA). This is a general purpose framework for a digital library in which very large numbers of objects, comprising all types of material, are accessible over national computer networks. It is described in a paper by Robert Kahn and Robert Wilensky (cnri.dlib/tn95-01).

This introduction describes the author's view of eight general concepts that emerged from the discussions. These concepts are key issues in the transition to a true digital library from the network services that we have today. The Kahn/Wilensky paper contains a comprehensive framework for resolving the issues.

General Principles

d-Lib forum d-Lib magazine

General Principles

1. The technical framework exists within a legal and social framework

d-lib magazine

Metadata: The Foundations of Resource Description

Stuart Weibel
Office of Research, OCLC Online Computer Library Center, Inc.
weibel@oclc.org

D-Lib Magazine, July 1995

This paper is an abbreviated version of the Summary Report of the OCLC/NCSA Metadata Workshop . It sets forth a proposal for the content of a simple resource description record (the Dublin Core Metadata Element Set) and outlines a series of further steps to advance the standards for the description of networked information resources.

Introduction

The explosive growth of interest in the Internet in recent years has created a digital extension of the academic research library for certain kinds of materials. Valuable collections of texts, images and sounds from many scholarly communities -- collections that may even be the subject of state-of-the-art discussions in these communities--now exist only in electronic form and may be accessible from the Internet. Knowledge regarding the whereabouts and status of this material is often passed on by word of mouth among members of a given community. For outsiders, however, much of this material is so difficult to locate that it is effectively unavailable.

Why is it so difficult to find items of interest on the Internet or the World Wide Web? A number of well-designed locator services, such as Lycos [http://lycos.cs.cmu.edu/] , are now available that automatically index many of the resources available on the Web and maintain up-to-date databases of locations. But indexes are most useful in small collections within a given domain. As the scope of their coverage expands, indexes succumb to problems of large retrieval sets and problems of cross disciplinary semantic drift. Richer records, created by content experts, are necessary to improve search and retrieval. Formal standards such as the TEI Header and MARC cataloging) will provide the necessary richness, but such records are time consuming to create and maintain, and hence may be created for only the most important resources.

MAGAZINE

Uniform Resource Names

A Progress Report

The URN Implementors

D-Lib Magazine, February 1996

ISSN 1082-9873

Introduction

The development of networked information requires reliable ways to name resources on networks. The Internet community has adopted the term, "Uniform Resource Name (URN)", for a name that identifies a resource or unit of information independent of its location. URNs are globally unique, persistent, and accessible over the network.

The concept of universal names has been warmly embraced by the networking and library communities, but convergence on the details proved difficult until recently. During fall 1995, however, members of the principal groups that are actively working in the field reached outline agreement on most of the major topics. The main characteristics of this agreement are described in this paper.

The catalyst for the recent progress was a meeting in October 1995 hosted by Keith Moore at the University of Tennessee. Invitations were sent to every group that had a current Internet draft on this subject. The URN groups represented are listed at the end of this report. This meeting was followed by a series of discussions including informal sessions at the December meeting in Dallas, Texas, of the Internet Engineering Task Force (IETF).

Convergence is important because many people who manage large collections of on-line information have been reluctant to commit to using any form of URN during a period of flux. The present consensus has two major results:

- Users who wish to give permanent names to on-line resources can now plan to incorporate URNs from existing naming schemes in documents, indexes, and on-line systems. They can be reasonably confident that future developments of the URN framework will not force them to reformat or otherwise modify existing URNs.

- The implementation of this framework will remove the concern that using a particular name scheme might affect longevity or the future usefulness of assigned URNs. The framework allows continued support for existing URNs, through other resolution systems, if one name scheme ceases to be supported in its original form. Thus users who assign names within any of the agreed-upon schemes are assured against obsolescence.

This report summarizes the emerging consensus. A strength of the framework is that it allows different approaches to be pursued, and the framework has the ability to evolve over the long term. Naming is a complex issue and the groups are interested in URNs for a variety of different reasons. They bring different philosophies and different technical approaches. Their implementations range in scope and complexity. It is therefore encouraging for the community that they have reached general agreement and are working together to find technical solutions to the outstanding questions.

d-lib magazine

Working Towards an Understanding of Digital Library Use

A Report on the User Research Efforts of the NSF/ARPA/NASA DLI Projects

Ann Peterson Bishop
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
218 LIS Building, 501 East Daniel Street
Champaign, IL 61820
abishop@uiuc.edu

D-Lib Magazine, October 1995

Introduction

The Digital Library Initiative (DLI) projects, funded jointly by the National Science Foundation (NSF), the Advanced Research Projects Agency (ARPA), and the National Aeronautics and Space Administration (NASA) began about a year ago. Their user study teams have already produced some valuable findings and described some provocative theoretical and methodological challenges. From my vantage point as coordinator of the University of Illinois DLI Social Science team, I will highlight the efforts of the six projects to communicate with each other about user research. Links to the DLI project home pages and to some of the papers published by project members have been included for more in-depth coverage of some of the issues summarized. In this article, I will also discuss the upcoming Allerton Institute at the University of Illinois, a methodological forum on digital library use that will provide another means for researchers in a variety of disciplines and settings to share their ideas and concerns about the conduct of social science research related to digital library use.

The Growth of Digital Libraries and the Challenge of Understanding Their Use

Improvements in information technologies and increased support directed towards our national information infrastructure have led to the development of a wide range of digital library collections and services. Academic, special, and public libraries are implementing on-line systems that provide their patrons with electronic access to library catalogs and a variety of other information resources. NASA is developing on-line collections of images and data for scientists and engineers. Museums are digitizing their collections and making them available on the Internet. Members of scientific communities are building collaboratories to support their work and communication. Publishers are experimenting with the creation of digital archives of their journals and books. And individuals and groups from all walks of life are using community-based networks to provide local and global access to information resources they have created. In addition to this array of existing networked information tools and resources--all of which can be thought of as variations on theme of the digital library or as pieces and layers of the digital information infrastructure--research and development projects related to building the next generation of digital library systems are also flourishing.

Digital libraries pose fascinating socio-technical challenges for understanding their use. Those supporting the construction of digital libraries are naturally concerned that their investments pay off in terms of attracting users and making information services more effective and efficient. The design and evaluation of digital libraries,

- Meeting Schedule
- Content List
- Digital Libraries Research and Development Forum (D-lib)
- Florida Center for Library Automation's Digital Library Project
- IBM Digital Library
- Pointers to information about Digital Libraries
- References: Research Department Virginia Tech Computing Center
- Virginia Tech Tactical Plan (PDF) and Project Reports

Pieces of the Puzzle:

- **Agents**
  - UMBC Intelligent Software Agent Resources
  - Survey of Intelligent Software Agents
  - Full list of agent links
- **Metadata**
  - Metadata: the Foundations of Resource Description
  - OCLC/NCSA Metadata Workshop Report
  - RFC-1807
  - TEI
- **Naming**
  - Handles
  - PURL
- **Z39.50**
  - Isite Software
  - Library of Congress WWW/Z39.50 Gateway/Info
  - Prise 1.0 Software
  - Willow

Some of these documents are in Adobe's Portable Document Format (PDF). In order to view them, you will need a PDF viewer

---

University Libraries, Virginia Tech
Send Suggestions or Comments to webmaster@scholar.lib.vt.edu
Last updated: May 23, 1996

URL: *http://scholar.lib.vt.edu/digilib/*

Digital Library Source Book, 1993, ed. E. Fox

To order a paper copy, or find out background information please look at the README file. To use an Adobe Acrobat Reader or Exchange to work with the book, look at the PDF version. Otherwise, use the PostScript version that appears below in sections.

See also more information of interest:

- April 1995 *Communications of the ACM*
- *Gladney et al. report on DL requirements and architecture (PostScript)*
- *PowerPoint presentation by Fox for 1994 Digital Libraries Workshop at Rutgers (to be decoded by binhex)*
- *PowerPoint presentation by Fox for 1994 Digital Libraries Workshop at Texas A&M (printable, in black and white, to be decoded by binhex)*
- *PowerPoint presentation by Fox for DL Keynote at EG-MM'94 in Graz (to be decoded by binhex)*
- *PowerPoint presentation by Fox for DL Keynote at ISMIS'94 in Charlotte (to be decoded by binhex)*
- *WWW Pages for CS2984 Course Notes on Digital Libraries*

Netlib Repository at UTK and ORNL

Netlib is a collection of mathematical software, papers, and databases.

There have been 11,785,446 requests to this repository as of Sat Jul 13 02:23:01 EDT 1996 .

Software, papers, etc.

- Browse the Netlib repository
- Search the Netlib repository

Services provided at Netlib

- Conferences Database
- National High-Performance Software Exchange (NHSE)
- Numerical Analysis Net (NA-Net)
- Performance Database Server
- Top500 Supercomputer Sites

Information about Netlib

- Frequently Asked Questions about Netlib (FAQ)
- Netlib Editors
- Netlib Mirror Sites
- Netlib Server Statistics

Netlib Maintainers

Netlib Attribute-Value Database

This is an index of the Netlib Attribute-Value Database. Please type a query in the search dialog. You may use freeWAIS-sf query syntax.

This is a searchable index. Enter search expression:

Submit

Attribute names:
(global indicates that the field is included in the global index that will be searched if no attribute name is specified)

- **file** -- any portion of the pathname for a regular file
- **lib** -- any portion of the pathname for a directory
- **for (global)** -- problem solved or description
- **gams** -- GAMS class
- **prec** -- Fortran precision (single, double, complex, or doublecomplex)
- **title (global)**
- **alg** -- algorithm or method
- **by (global)** -- author (name <email>)
- **keywords (global)** -- terms as would be drawn from a subject thesaurus
- **lang** -- programming language

Search Examples:

1. To seach for single precision routines in the lapack directory that do Schur factorization:

   ```
   file=lapack and file=single and Schur
   ```

   (since the lapack single precision routines are in the lapack/single directory)

2. To search for curve fitting or gams class E1 and its subclasses:

   ```
   (curve and fitting) or gams=e1*
   ```

3. To do a literal search for 'cosine transform':

   ```
   'cosine transform'
   ```

Digital Libraries - Example: The CORE Project

Some digital libraries have been developed for a profession. The CORE Project is such an effort, for the field of chemistry. It involves the major US publisher and information provider of chemistry information, the American Chemical Society, and its subsidiary, Chemical Abstracts Service.

Statistics regarding CORE Digital Library:

- Pages: 430K (now 375K)
- Extracted Graphics: 387K
- Articles: 82K
- Gbytes Page Images: 50
- Gbytes Text: 4.4
- Gbytes Graphics: 6
- Gbytes Index: 11
- Scanning from: paper, microfilm
- High Resolution (to print): 300dpi (2560x3328) B&W
- Low Resolution (to display): 100dpi (856x1109) grey scale
- Conversion of Figures: extraction
- Conversion of Text: typesetter tapes to SGML
- search engine: OCLC's Newton
- Interfaces: OCLC's SCEPTER, Bellcore's Pixlook

The CORE Project: Overview

The CORE project is an electronic library prototype the provides networked access to the full text and graphics content of the American Chemical Society journals and associated Chemical Abstracts Services indexing since 1980 (some 250 journal years of data). The database is coded in SGML (Standard Generalized Markup Language) which was translated from the original typography codes, captures the structural richness of the original document and provides flexibility for indexing, searching and display. The prototype provides a full-scale laboratory environment in which to explore issues of database structure, user interface capabilities, and information retrieval questions on a large, real-world scholarly electronic journal database. The complete database, representing more than 600,000 pages of full text and graphics, will be available at Cornell University in late 1994. The major contributors of this electronic library project include:

- Cornell University (Mann Library)
- OCLC
- Bellcore
- American Chemical Society
- Chemical Abstracts Services

Relevant publications

- The CORE Project: Technical Shakedown Phase and Preliminary User Studies
- The Design and Implementation of XSCEPTER, an X-Windows Graphical User Interface to the CORE project

---

Some Images of XSCEPTER

XSCEPTER provides "on-the-fly" formating of SGML as defined by configurable style guides and a DTD, to provide rapid display of scholarly data. The XSCEPTER interface is coupled with NEWTON, OCLC's proprietary search engine, to provide navigational capabilities of the CORE collection.


XSCEPTER main window


Find Box to allow indexed full boolean searching of the CORE database.

TULIP The University Licensing Program

When you scroll further down this page you'll find

Introduction

TULIP is a cooperative research project testing system for networked delivery and use of journals, performed by Elsevier Science and nine Universities in the USA. The participants set three objectives at the outset:

Technical

To determine the technical feasibility of networked distribution to and across institutions with varying levels of sophistication in their technical infrastructure. "Networked distribution" means sending the information both across the national Internet and over campus networks to the desktops of students and faculty. Elsevier will deliver the journal information to participating universities in standard formats. The universities will incorporate the information in local prototype or operational systems. A wide variety of delivery alternatives, search and retrieval systems and print-on-demand options will be compared.

Organizational and economic

To understand, through the implementation of prototypes, alternative costing, pricing, subscription and market models that may be "viable" in electronic distribution scenarios; comparing such models with existing print-then- distribute models; and understanding the role of campus organizational units under such scenarios. The overall goal is to reduce the unit cost of information delivery and retrieval. "Viable" means economically and functionally acceptable to all parties.

User behaviour

To study reader usage patterns under different distribution (technical, organizational and economic) situations. Improvement in the functionality of the information, whether as to article structure or retrieval tools, will also be considered. Certain data will be collected uniformly at all sites for analysis in the aggregate and for comparison among different systems.

Click here to return to top of information

SiteSearch

Search for specific topics

Go to OCLC Home Page

- **Demonstration**
- **Descriptions of Products and Services**
- **News**
- **Publications**
- **User Documentation**

Demonstration

- The WebZ Demo Page lets you search OPAC data and several other reference databases (http://tikal.dev.oclc.org:2000)

Descriptions of Products and Services

- Introducing OCLC SiteSearch: To the Next Stage of the Electronic Library
- Elsevier Science/OCLC Electronic Publishing Pilot Program
- WebZ Server Questions & Answers
- Z39.50 Server System Questions and Answers

News

***OCLC Newsletter* Features**

- Georgia's GALILEO Project
- Interview: Merryll Penson, Ralph E. Russell, and William Gray Potter. The directors of three libraries involved in the GALILEO project discuss the creation of the statewide project, its current status and future plans

News Releases

- 'GALILEO' to Use OCLC SiteSearch Software to Deliver Information, FirstSearch to Georgia Libraries--December 1, 1995

See the complete news release list for earlier news releases.

Publications

Reference News

- Winter 1996, No. 29
- Fall 1995, No. 28
- Summer 1995, No. 27
- Spring 1995, No. 26
- December 1994, No. 25

Digital Library is the Information Age tool for extracting value from your information assets in new ways.

Digital Library Adapts To Your Needs

Every digital library presents unique challenges. That's why IBM Digital Library is not only comprehensive but scalable and flexible. To see digital libraries that exist today, go directly to the Case Histories. Quick and thorough overviews of IBM Digital Library solutions for organizations in government, higher education, media and cultural institutions are included. For fast access use the button below the Wheel.

On the other hand, get the essentials on building your own digital library by taking five steps through the main components of IBM Digital Library. The Wheel is your guide. Jump in anywhere, and explore the key functions: Create & Capture, Search & Access, Distribution, Rights Management, and Storage & Management.

To have an IBM representative with digital library expertise bring you detailed information for your own Digital Library project, use the More Info link at the bottom of every page. And help us serve you better by taking a few minutes to complete our survey while you're at it.

IBM Digital Library

Explore IBM Digital Library On A Case By Case Basis

Taking the initiative to build a solution using IBM Digital Library requires foresight, strategic acumen and trust in the technology that supports the enterprise. Arranged here by categories (Government, Higher Education, Media, and Cultural Institutions) are case histories of several IBM Digital Library solutions. No two are exactly alike, yet all share challenges that IBM Digital Library's end-to-end capabilities are uniquely qualified to handle.

Government

Government agencies collect and store massive amounts of information. Widespread bureaucracies need to share the information assets they command. Meanwhile, the public's desire to gain access to public information seems endless.

IBM Digital Library provides the foundation for linking hubs across a range of government sites. A user needn't travel to the capital to find governmental information. IBM Digital Library enables people to find what they need from wherever they are.

Derwent
With IBM Digital Library, Derwent offers network access to U.S. patent information. Users see a Lotus Notes interface while they access 20 years of patent data with weekly updates. A full patent document can be ordered from Derwent with a single button click.

Higher Education

The concept of a school without walls has been discussed for over a century. By setting the stage for distance learning and access to learning archives, both using multiple forms of media, IBM Digital Library finally makes the "open classroom" truly open. All the important documents can be available on-line 24 hours a day, 7 days a week.

Case Western Reserve University
In a joint effort with IBM, Case Western Reserve University's faculty, librarians and information service staff undertook a project to identify the critical needs and components for developing a digital library. A big part of the success story is IBM's rights management technology, which administers the terms and conditions for use of copyrighted materials.

Indiana University School Of Music
Indiana University's Variations Music Information System, created with IBM Digital Library, puts an entire music library on-line. It allows the music students to immerse themselves in their studies wherever they are, whenever they want.

Marist College
With 100,000 objects currently digitized and stored, Marist College anticipates adding an average of 10,000 more per month. Within the next four years, Marist will have several million objects in their digital library.

This supports the new paradigm for learning, providing access to learning archives, at higher educational institutions.

Media
The convergence of entertainment, information and technology is dynamically changing the media and entertainment industries. IBM Digital Library brings intelligence and value to content creation and hosting, asset banking, workgroup access, electronic commerce, rights management and protection, royalties payments and licensing, distribution and archiving. What's more, IBM Digital Library can be "under the covers," whereby content owners can establish their own branded image for the services and archives they create.

EMI's KPM Music Library
Among the world's largest suppliers of music (and other media) for producers of movies, TV shows, advertising and presentations, EMI Music Publishing has opened the vaults of KPM's Music Library through IBM Digital Library and IBM's partner Multimedia Archive and Retrieval Systems plc.

Institute for Scientific Information
After searching 6 months for a technology partner to build an "electronic library system," the Institute for Scientific Information chose IBM Digital Library for good reasons. ISI's Electronic Library Pilot project contains over 1,350 journals including tables of contents, bibliographic data and abstracts.

Cultural Institutions
Imagine reading the Dead Sea Scrolls. Or the Rosetta Stone. Until very recently, only the privileged few among the world's scholars could view the great artifacts of human history. With IBM Digital Library, the great works of mankind can be opened up to the world. The opportunity to preserve for future generations the knowledge of antiquity has arrived -- IBM Digital Library.

The Vatican Library
An IBM Digital Library project of monumental proportions, the Vatican Library holds over 150,000 manuscripts, including the oldest known manuscript of the Bible from 350 A.D., and 1.5 million books including 8,000 published during the first 50 years of the printing press.

Archivo General de Indias
To better serve researchers and to preserve its archives of 90 million pages of historical materials documenting the Spanish conquest of the new world, Archivo General de Indias joined IBM Spain and the Ramon Areces Foundation in creating a digital library. Currently, more than 9 million pages may be searched and accessed on 40 IBM workstations.

The QBIC Project

This site received a 4 star rating from McKinley Group's editorial team.

Query By Image Content(QBIC)

On-line collections of images are growing larger and more common, and tools are needed to efficiently manage, organize, and navigate through them. We have developed a system called QBIC which allows complex queries of large image databases. The queries are based on image content -- color percentages, color layout, textures, and shapes of image and their objects. Some of this technology is currently available in IBM's Ultimedia Manager product.

To try the World Wide Web QBIC search engine on a database of approximately 1,900 images, click on the button below:

Try our latest browser: Try Demo

You must have a HTML3.0 capable browser to run the above demo. If you don't, try downloading the new IBM WebExplorer for OS/2 or Netscape.

Otherwise, you can use our older browser: Try Demo

Did You Enjoy Our Demo?

If you did, we are looking for early-adopters in selected application areas. If you have a web site with images and would like to use this technology, please contact The QBIC Group

If you didn't, we would like to know if we can improve it in some way. If you have any suggestions or comments please mail to qbicwww@almaden.ibm.com

If you would like to be on our mailing list, please enter your name in the following box and press Enter:

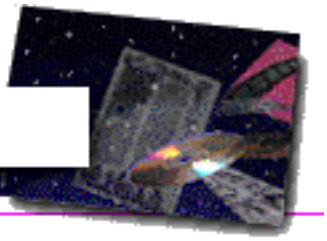
Your e-mail address:

Other links related to QBIC:

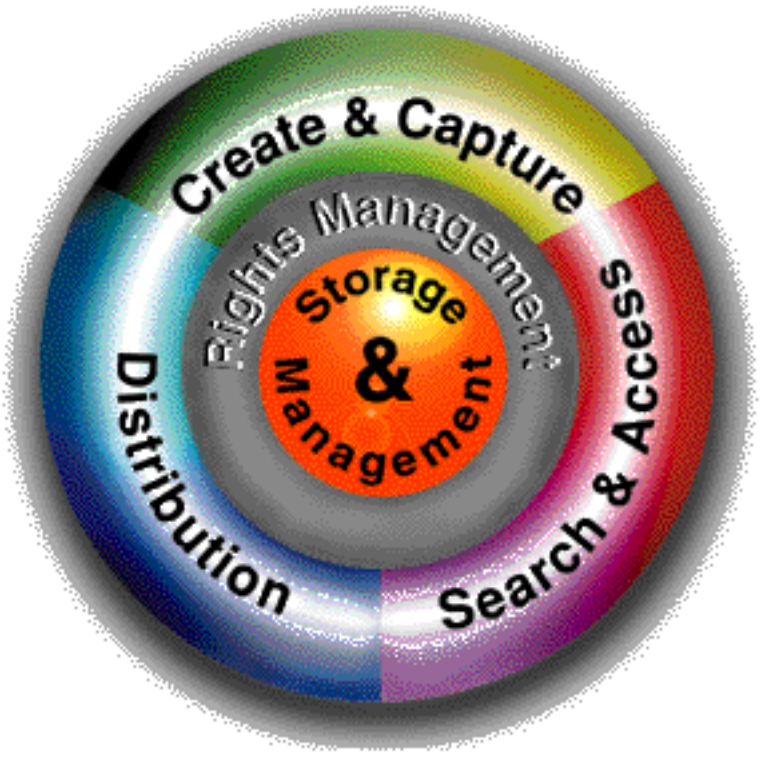
- A show and tell of the QBIC technology.
- Technical paper requests on QBIC.
- Ultimedia Manager 1.1 - A product that incorporates some of the QBIC technology.
- DB2 Extenders, which will soon include QBIC technology.

IBM

Software



IBM Digital Library



Rights Management: New technologies bring new opportunities, but not without risk

The definition of content ownership is not universal. Copyrights may be protected in one country and practically ignored in the next. So how do you protect your intellectual property rights in a digital library that anyone with a PC can learn to use in a few minutes? Rights Management is the answer. IBM has concentrated on Rights Management as a key factor in creating IBM Digital Library and allowing you to develop a full digital library solution to meet your business needs.

The challenge is to provide ease of use, privacy, content integrity and cost utility for users while establishing bulletproof Rights Management solutions for content owners. To meet that challenge, IBM Digital Library incorporates the rich legacy of IBM's decades-long innovations in networking security and transactions technology.

The Legacy At Work

Compare using a Digital Library to making a withdrawal from an automated banking machine. A user is identified, enters a password, requests information or an object, and the system checks the request against the user's eligibility. Upon approval the requested objects can be watermarked to deter illegal duplication. And the entire process is conducted in a secure environment without the threat of intrusion. Meanwhile, compensation for the value added to the information or object is duly processed.

Did you know that the vast majority of the world's automated banking systems are built on IBM software and technology? Trust and reliability are what make legacies, and the Rights Management functions of IBM Digital Library have inherited both.

Signed, Sealed, Delivered

IBM Digital Library can authenticate original media -- photos, manuscripts, audio, video, film and pictures -- by using electronic signatures. Digital content can be recognized as authentic with these signatures. Watermarks, a form of electronic signature currently in use throughout several IBM Digital Library solutions, are encoded onto photos, films, videos and manuscripts. Visible watermarks can be graphically representative of a content owner's identity, like a logo or crest. Watermarks are sophisticated identifiers that inhibit the

misappropriation of content owners' assets while assuring users' confidence in the authenticity of the content.

IBM infoMarket Search service represents a giant leap forward in Network-centric computing for both content owners and users. The infoMarket service enables users to search simultaneously available network databases (private, public or both). For content owners, infoMarket provides its Plug-N-Publish® toolkits and the Cryptolope®, an encryption-protected "envelope" that can travel on public networks.

Anybody who wants to open a Cryptolope to read its contents must use a key to unlock it. Users can preview a Cryptolope's contents, then decide whether to pay for the key. For sensitive content needing increased security, a Cryptolope may require several keys. And a Cryptolope can travel on networks with only the intended recipient being aware of its presence. All the while, the infoMarket service keeps impeccable records of rights payments transacted.

IBM Digital Library follows through for all content owners. With the rise of multimedia, the work of several content authors is contained within a single media object. For instance, a digitzed document might contain a photo, an illustration, a page of text and some music---each authored by a different person who should be compensated. Furthermore, a customer accessing this document might not be required to purchase the entire work, but rather just the individual section that is accessed. The Rights Management capabilities of IBM Digital Library offer discreet recognition for each content author, keeping track of who gets paid for what.

IBM Digital Library provides trusted means for protecting and managing the rights of content owners. Rights Management issues impacts every aspect of IBM Digital Library -- Create & Capture, Storage & Management, Search & Access and Distribution. Use The Wheel at the top of this page to continue exploring IBM Digital Library technology.

Hyper-G

The Hyper-G system has been developed in Graz, Austria. It can function as a standard server for WWW, accessible through Mosaic at any Hyper-G server, such as the Hyper-G root or the Hyper-G W3 Gateway at Virginia Tech . Note that *J.UCS* is an electronic journal, coordinated from Graz, accessible using Hyper-G, with links from the 2 above-mentioned servers.

Hyper-G also has many innovations relative to Mosaic, such as:

- Links are stored in a distributed object-oriented database.
- Users communicate with a local server which caches data from other servers.
- Multimedia information can be presented as it is received instead of having to wait for the entire file to be transmitted.
- Viewers are available for each type of information, so a link can be to a portion of an image, or part of a movie, or string of words in a text file.
- Local maps of links to and from the current node are automatically drawn.
- Users can organize documents using collections, that can be expanded or collapsed with a click of the interface.
- Multilingual support is present in the interface and in selecting documents.
- Automatic indexing of new nodes and built in search support.

Hyper-G has a command-line interface, a Mac interface under development, and:

- Powerful PC client (Amadeus)
- Powerful UNIX client (Harmony)

User: www-anonymous

What do you want to serve today...?

HYPER-G NOW HYPERWAVE

NEXT GENERATION WEB SOLUTION

HyperWave Server is advanced web server (WWW) technology, based on an **object-oriented database** which was developed especially for hypermedia document management. HyperWave guarantees **automatic hyperlink consistency** and supports hyperlinks to and from multimedia documents, **full text retrieval** and a sophisticated **access control** system with user management and user groups.

HYPERMEDIA AUTHORING UTILITIES

HyperWave Author - code-named "Harmony" for the UNIX version and "Amadeus" for the Windows version - accesses HyperWave servers across the Internet or any internal TCP/IP network, allowing users to view and **manipulate information** in multiple ways. Advanced **navigational tools** help users orient themselves and avoid becoming "lost in hyperspace".

[About] [Products] [Documentation] [FAQ] [Software] [News]

More attribute information.
Author: iicm
Parent(s):

IFABO 95
About IICM, HMS & their projects
IICM's Public Services
About HyperG
Working with HyperG in Paderborn

User: www-anonymous

KEY FEATURES

WORLD WIDE WEB

HyperWave Server software represents one of the most powerful WWW technologies currently available. Because it supports common network protocols and document formats, HyperWave Servers can be browsed and administrated with widely-used WWW browers such as Netscape and Mosaic. Easy navigation is granted by hyperlinks and HyperWave's additional folder type: the collection.

- Integrated search engines
- Multilinguality
- Hyperlink consistency
- Meta-information
- Object-oriented database

INDUSTRY STANDARDS

Compatibility with industry standards is very important for WWW server technologies. By supporting standard network protocols such as HTTP/1.0, HyperWave Server provides unlimited connectivity to all kinds of HTTP clients and servers. The server provides support for Multi-purpose Internet Mail Extension (MIME) types and standard document formats such as HTML, GIF and JPEG. HyperWave Server also interacts easily with business applications using the Common Gateway Interface (CGI), the major standard gateway in the WWW.

ACCESS CONTROL

HyperWave Server stores all documents in an object-oriented database, it does not use the UNIX or Windows file systems. Providing its own access authorization system, HyperWave Server is much more secure than any other filesystem-based WWW server. It provides sophisticated access control to individual documents and collections using usernames, passwords, named groups and read/write/unlink rights. The server can be administrated remotely by all members of a special group "system". Billing and cashing - extremely important for Internet commerce - can be achieved using integrated accounting facilities.

REPLICATION

The HyperWave Interchange Format (HIF) lets users interchange interlinked multimedia webs between HyperWave Servers. Tools for importing and exporting collection trees to and from this format make it easy to replicate information on other servers. Applications reach from the mirroring of documentation or electronic journals to firewalls: HyperWave Servers let you create and modify public information on your side of the barrier and then replicate it to the "outside" server.

CONNECTIVITY

HyperWave Server provides connectivity to other applications and software modules via the Common Gateway Interface (CGI). This ensures interoperability with business applications, client/server databases, expert systems and special search engines. HyperWave Tools as a suite of command-line utilities provide a fast and robust interface to the server for batch programming and scripting. Finally the HyperWave SQL gateway ensures connectivity to common SQL databases: e.g. Oracle.

More attribute information.

User: www-anonymous

KEY FEATURES

HYPERMEDIA AUTHORING

HyperWave Author is the ultimate interactive authoring tool for HyperWave Servers. Users can author remotely over network boundaries: the Internet or any other TCP/IP based network can be used. HyperWave author provides full support for the HG-CSP network protocol, special HyperWave Server features such as database and search facilities are seamlessly integrated into the interface.

OBJECT DATABASE

Object orientation is one of the key concepts of HyperWave. HyperWave Author provides full support for HyperWave Server's object-oriented database system, allowing easy insertion and editing of server-side objects. HyperWave Author for Windows additionally provides a local version of the database, letting users author web applications offline which they can later easily upload to a HyperWave Server.

VRML AND POSTSCRIPT

HyperWave Author software includes IICM's free VRML scene viewer and a viewer for PostScript documents. VRML is the standard 3D data format in the WWW. PostScript is the industry standard for electronic publishing. HyperWave Author provides integrated PostScript viewer software, including the facility for inserting hyperlinks in PostScript documents: annotations to non-HTML documents are possible because of HyperWave's link database approach.

ADVANCED NAVIGATION

Critics of the WWW often mention the so-called "lost in hyperspace" syndrome. HyperWave Author provides advanced navigation concepts and demonstrates that there are solutions to this problem: tree-view collection browsers let you navigate easily through big web servers and dynamically generated hyperlink maps help you keep masses of interlinked information up to date.

DISTRIBUTED INFORMATION MANAGEMENT

HyperWave's authoring software provides the facility of distributed information management: every logical part of a company can have a virtual web server without having the overhead of setting up its own real web server. A company can have one corporate identity on the web, running a WWW service where every department of the corporation is responsible for its own part.

MULTILINGUAL DOCUMENTS

HyperWave Author supports easy creation and editing of multilingual web applications. HyperWave's support for multilingual document clusters is especially interesting if your company is located for example in Europe or Asia, or any other part of the world where more than one language is common. HyperWave Author's advanced navigational tools help you to get an overview of complicated multilingual webs.

More attribute information.
Author: gmesaric
created: 96/04/16 09:24:32
modified: 96/04/17 10:28:14

User: www-anonymous

TECHNICAL SPECIFICATIONS

HYPERWAVE AUTHOR FOR WINDOWS (AMADEUS)

- Compatible with industry standards
  - Supports HTML
  - Views common image formats such as GIF and JPEG
  - Integrated MPEG movie player
  - Comes with VRML (VRweb) and PostScript viewers
- 32-bit application (runs under Windows 3.1x using WIN32s)
- Efficient interactive connection to HyperWave Server (HG-CSP)
- Local object database for offline hypermedia authoring
- Sophisticated interface for HyperWave's integrated search engines
- Supports multilingual documents
- Full support for access control (identification, rights)
- Windows95 compliant TreeView for collection browsing

HYPERWAVE AUTHOR FOR UNIX (HARMONY)

- Compatible with industry standards
  - HTML edit API tool
  - Views standard image formats (GIF, JPEG, TIFF)
  - Integrated movie (MPEG) and audio (AVI, AU) players
  - Comes with VRML (VRweb) and PostScript viewers
- Advanced navigation tools (local map, 3D landscape)
- Interactive client-server protocol to HyperWave Servers
- Sophisticated interface to HyperWave's search engines
- Multilingual document management
- Support for access control (identification, rights)
- Point-and-click hyperlink creation
- Integrated communication facilities (talk/conference)

SUPPORTED PLATFORMS

HyperWave Author for Windows (Amadeus)

Vendor	Architecture	Operating System	Memory Requirements
Intel	486, Pentium	Windows95/NT	8 MB

HyperWave Author for UNIX (Harmony)

User: www-anonymous

Harmony's orientational aids

Harmony has many built-in features to discourage the phenomenon of "getting lost in hyperspace" while browsing large information spaces.

Local Map

Harmony's Local Map presents a dynamically generated graphical overview of the link relationships of a chosen document. Both incoming and outgoing hyperlinks are represented. Selecting an object toward the edge of the map and generating a new display offers a new means of associative browsing.

Location Feedback

When you select a document or collection in the Local Map, in the search result list, or follow a hyperlink, the location of the corresponding object in the collection hierarchy is **automatically** displayed in the collection browser, providing a powerful aid to orientation.

History

The History Browser offers a timeline of past interactive waypoints, including previous search panels.

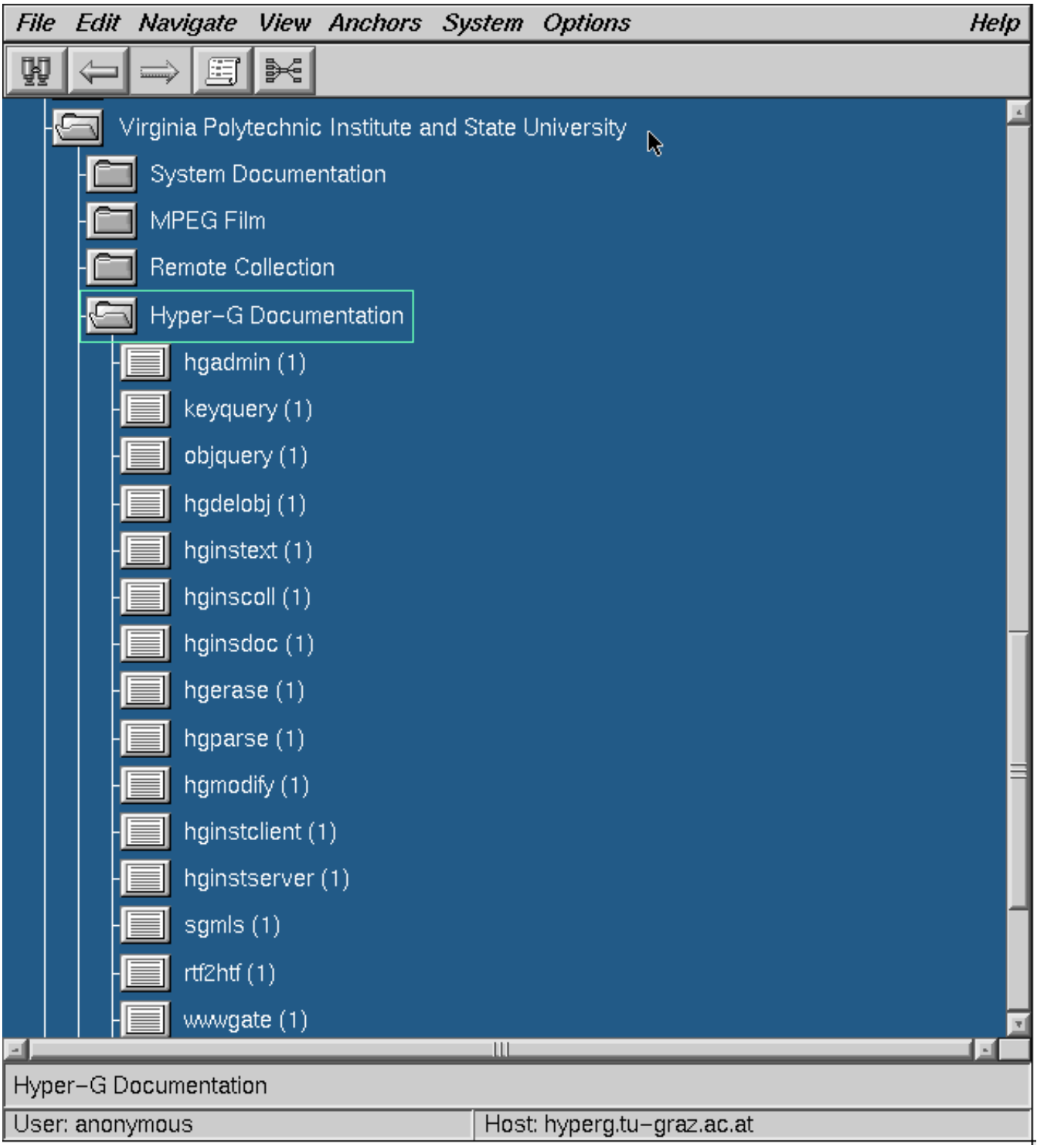
3D Information Landscape

The Information Landscape is a three-dimensional graphical overview map of the collection structure. Users can "fly" over the hyperspace landscape looking for salient features, select interesting documents, etc. This feature requires platform support for IrisGL, OpenGL or Mesa and is currently available for SGI, DEC Alpha, Solaris, Linux and HPUX machines.

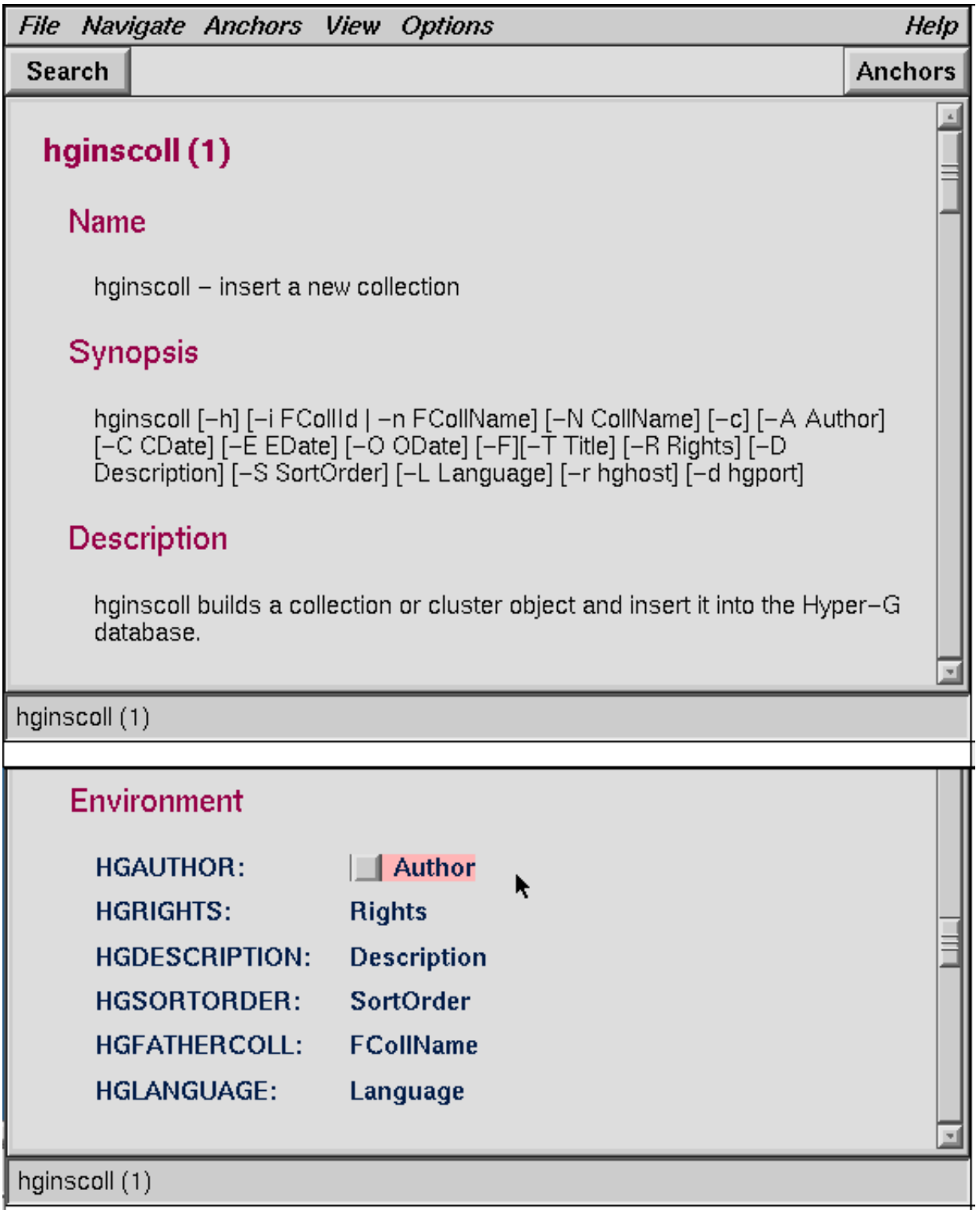
More attribute information.
Author: iicm
created: 95/12/18 08:19:00
modified: 95/12/20 08:06:04

Parent(s):

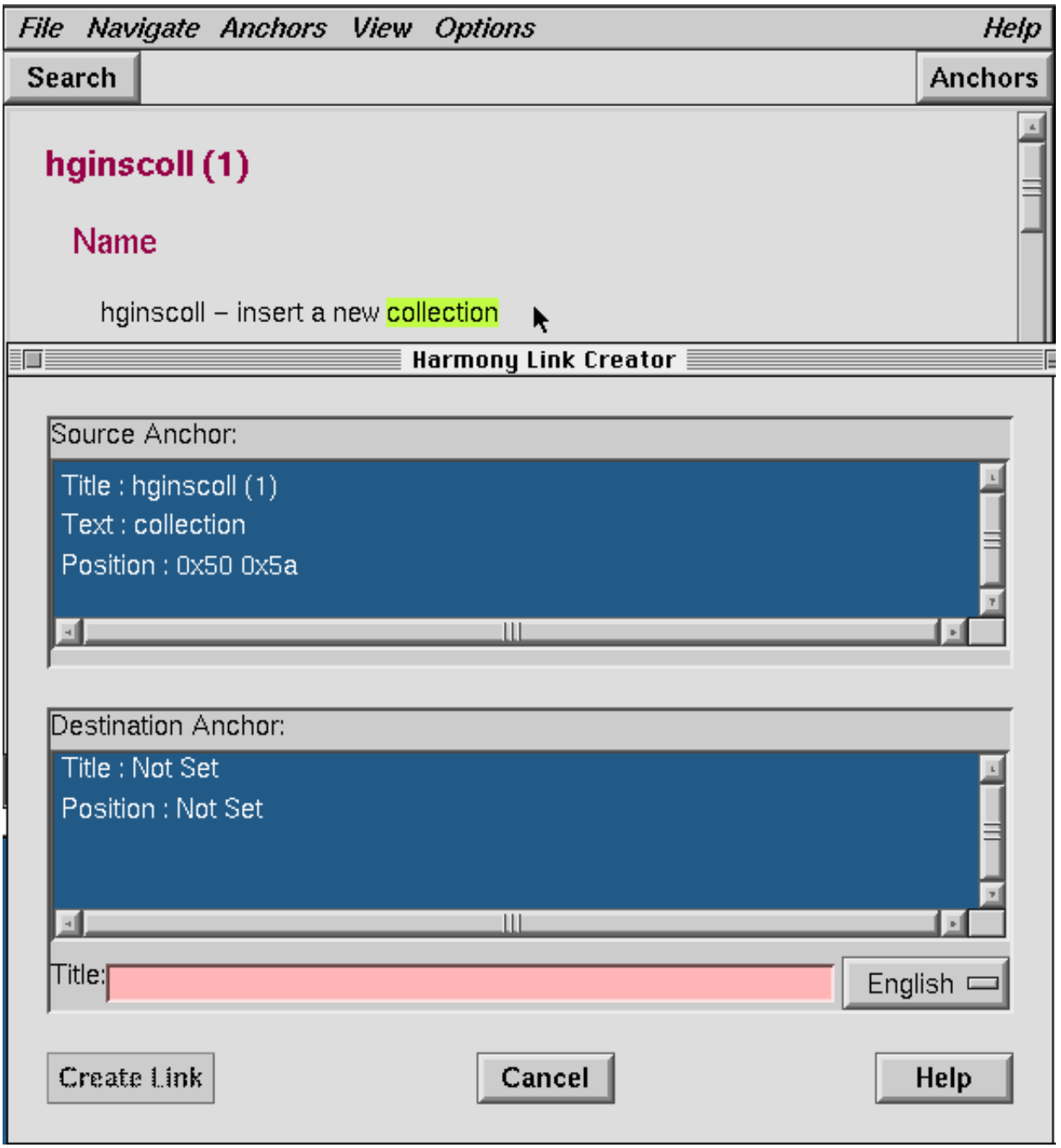
Harmony's orientational aids



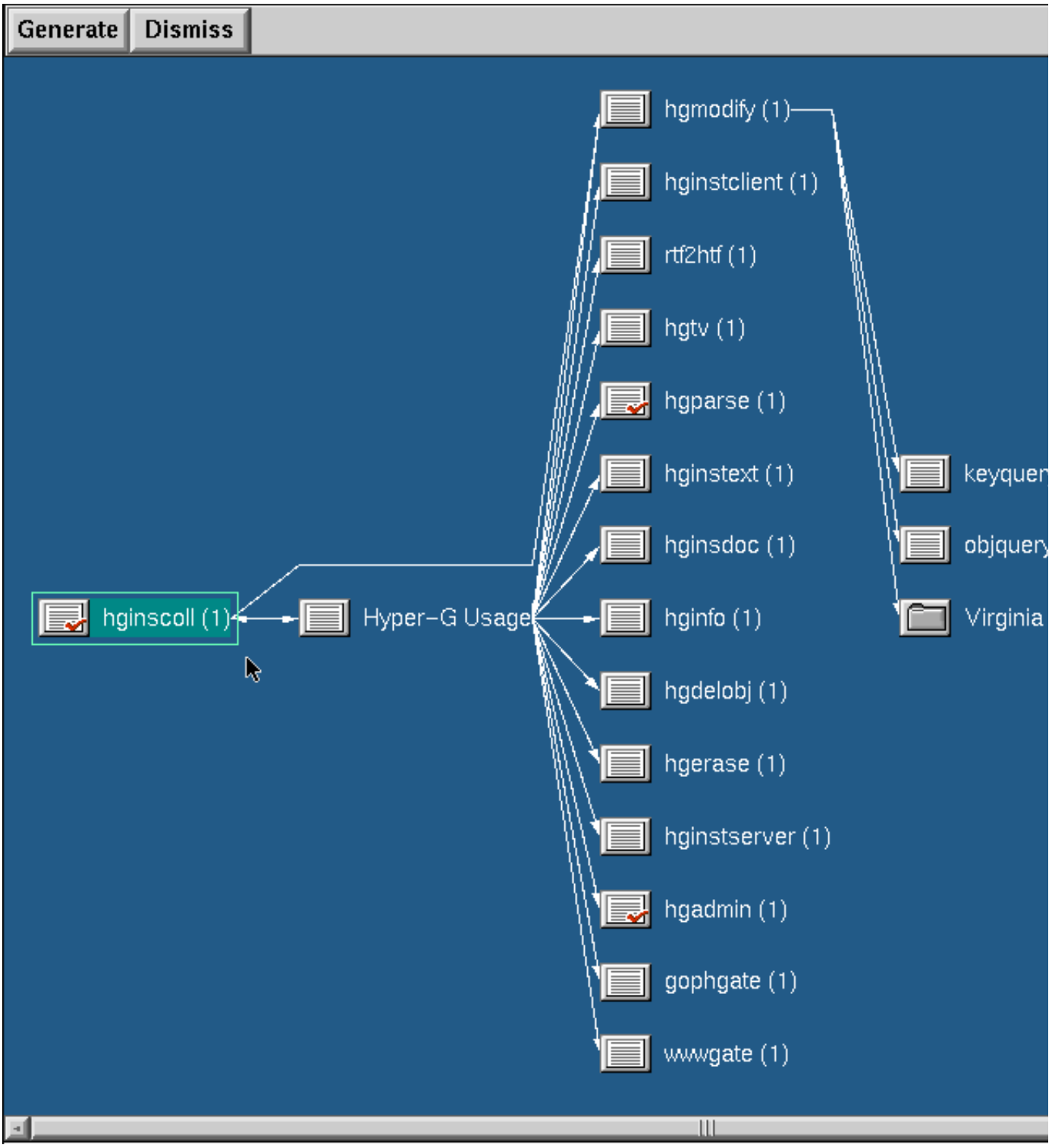
- viewing two text nodes and marking an anchor



- making a link



- viewing a local map


Generate
Dismiss
hgmodify (1)
hginstclient (1)
rtf2htf (1)
hgtv (1)
hgparse (1)
hginstext (1)
hginsdoc (1)
hginfo (1)
hgdelobj (1)
hgerase (1)
hginstserver (1)
hgadmin (1)
gophgate (1)
wwwgate (1)
keyquer
objquery
Virginia
hginscoll (1)
Hyper-G Usage

User: www-anonymous

Information structuring facilities in Harmony

Hierarchical Browsing

Hyper-G servers use not only hyperlinks as a means of structuring information content but organize information into so-called **collections**, which are similar to directories, as well. All information on a Hyper-G server must be part of at least one collection, making it possible to access every document without the need for hyperlinks. Harmony's Collection Browser displays the hierarchical membership structure of Hyper-G data, like a graphical file browser and allows you to select and access objects that interest you.

Search

Harmony's Search Dialog supports both attribute (keyword, title, author, creation time, etc.) and content (full text) searches. The scope of searches is user-definable, ranging from individual collections to all collections on all Hyper-G servers worldwide. Search results are presented as a ranked list and can be used as the scope for further searches.

Hyperlinks

Harmony supports hyperlinks between arbitrary document types, including text, image, film, PostScript, and 3D scenes. Both source and destination anchors can be defined interactively.

More attribute information.
Author: iicm
created: 95/12/18 07:32:25
modified: 96/01/17 14:06:58

Parent(s):

Information structuring in Harmony

User: www-anonymous

Harmony's document viewers

Documents in Harmony are displayed by separate viewer processes in windows of their own:

Text Viewer

A generic SGML parser is used to display Hyper-G (HTF) and WWW (HTML) texts. Inline images in XBM, XPM, GIF, TIFF, and JPEG formats are supported.

Image Viewer

GIF, JPEG, TIFF, and PNG images are supported and may be zoomed, panned, etc. A special feature is live display -- when turned on, images are built up progressively on-screen as they are loaded. The autofit option automatically scales images to fit the current image viewer window.

Film Viewer

MPEG-1 video streams are supported. Options include live display while loading, double size display, alternative dithering methods, and gamma correction. After loading, selective portions of the film may be replayed, the frame rate altered, etc.

Audio Player

The Audio player is a graphical interface shell around whatever native audio command the system provides. The Audio Player supports both the Network Audio System (NAS) and local audio commands provided on your system. A full-featured, native Harmony Audio Player is under development.

PostScript Viewer

PostScript files can be displayed page by page, zoomed, printed, etc.

VRweb 3D Scene Viewer

3D model descriptions are displayed and can be manipulated or traversed in three dimensions. Hyperlinks are attached to objects in the model. The scene viewer is the Harmony version of VRweb, which supports models in VRML and SDF formats.

Welcome to the **BEV HistoryBase**, a WWW History Page for the Blacksburg Electronic Village! Try out the BEV History Timeline to learn more about the history of our electronic community. For a non-graphical alternative, check out the Textual BEV History Timeline. Both contain the same information so feel free to browse either.

[Main Timeline | Contribute | What's New? | Search]

Message of the Day Listings

Blacksburg Telecommunications Advisory Committee Meeting Minutes

BEV Media Coverage Archive

BEV Group Home Pages

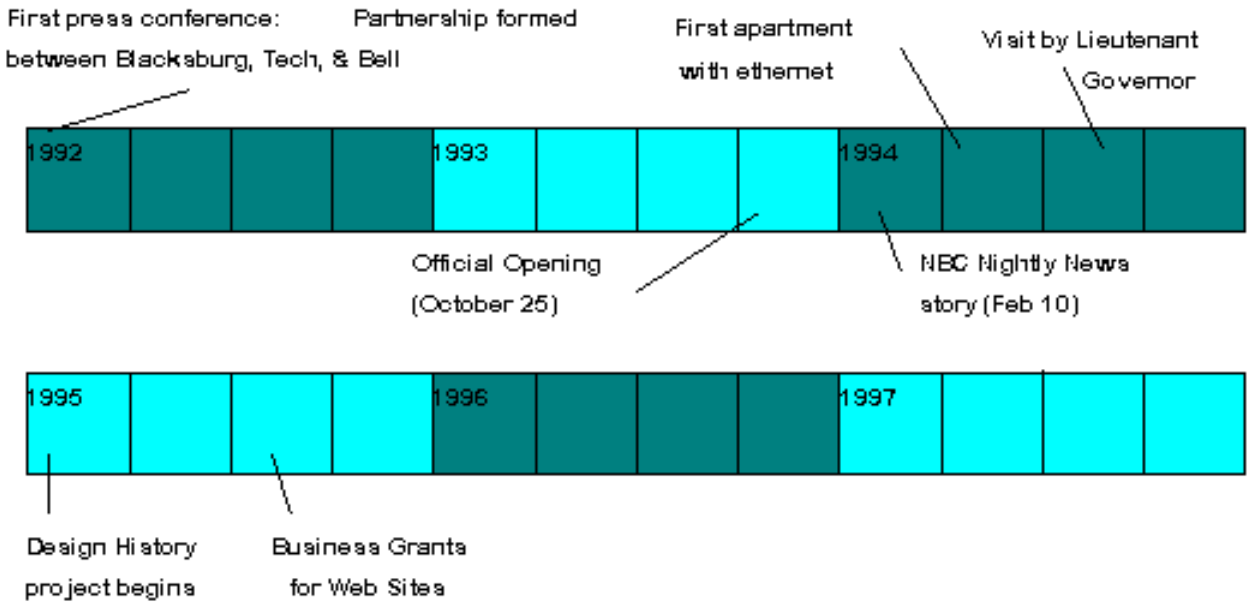
This project is supported by NSF Grant CDA-9424506. A copy of the grant proposal is online.

Last updated 27 October 1995 / schmidt@cs.vt.edu

HistoryBase Main Page | Contribute | What's New? | Search | BEV Homepage

BEV HistoryBase: Main Timeline

Click in a box to see a more detailed history for that quarter



Click in a box to see a more detailed history for that quarter

HistoryBase Main Page | Contribute | What's New? | Search | BEV Homepage

HistoryBase Main Page | Main Timeline | Prior Quarter | Next Quarter | Contribute | What's New? | Search

Quick-Click Timeline: 1992 1993 1994 1995 1996 1997

BEV HistoryBase: Jan-Mar, 1992

- January 18 1992 Plan would change fiber of Blacksburg
- January 21 1992 Blacksburg: Model of computer future
- January 21 1992 Blacksburg may become 'electronic village'
- January 21 1992 Hi-tech may be the norm
- January 21 1992 Fiber optics may link Blacksburg
- January 21 1992 A look into the future
- January 21 1992 Project envisions Blacksburg as an 'electronic village'
- January 27 1992 Virginia Tech Launches Study for Fiber Optic Community Network
- January 30 1992 Electronic village proposed
- January 31 1992 Gut (Comic Strip)
- February 1992 Electronic Village: Technology showcase
- February 06 1992 Blacksburg Telecommunications Advisory Committee Minutes
- February 11 1992 Electronic village could make us lazy
- February 13 1992 In a Small Mountain Town, The 21st Century Is Calling
- February 25 1992 The Blacksburg Experiment
- March 03 1992 Blacksburg Telecommunications Advisory Committee Minutes

HistoryBase Main Page | Main Timeline | Prior Quarter | Next Quarter | Contribute | What's New? | Search

HistoryBase Main Page | Main Timeline | Quarter Year View | Contribute | What's New? | Search | BEV Homepage

TITLE: Blacksburg may become 'electronic village'

Contributor: KENNETH WILLIAM SCHMIDT JR (wschmidt@bev.net)
Submit Date: Sep 05 1995
Document Date: January 21 1992
Document Categories: Media Coverage: Newspaper

Author: Linda F. Jilk
Publication: *The News Messenger* (Montgomery County, VA)

Summary: This article contains the same information as all the other project announcement articles that appeared around this time. It describes the project at a very high level, gives a little technical information, describes some scenarios of how the BEV could be used, and presents a vision for the future with the project.

Add an annotation to this document

HistoryBase Main Page | Main Timeline | Quarter Year View | Contribute | What's New? | Search | BEV Homepage

Corporation for National Research Initiatives

CS-TR
Computer Science Technical Reports

- An Introduction to the CS-TR Project, Robert E. Kahn, December 11, 1995
- Participants
- Architecture of the Digital Library
- Implementations
- Contributed technology

Participants

Each participant has provided on-line information about their work.

- Carnegie Mellon University
- Cornell University
- University of California at Berkeley
- Stanford University
- Massachusetts Institute of Technology
- CNRI

Architecture of the Digital Library

Members of the CSTR project have been developing the basic architecture that must underlie a world wide digital library, where valuable information is stored. This work includes:

- An architecture for the digital library.
- A handle system to maintain unique identifiers for objects in the Digital Library.

Implementations

Several public systems have been implemented with support from CSTR and are available for public use. (Some of these services are under development and subject to change at short notice.)

- Dienst, a distributed search system for technical reports (Cornell)
- Mercury, a centralized search system for technical reports (Carnegie Mellon)

Networked Computer Science Technical Reports Library

NCSTRL *(pronounced "ancestral") is an international collection of computer science technical reports from CS departments and industrial and government research laboratories, made available for non-commercial and educational use. The NCSTRL collection is distributed among a set of interoperating servers operated by participating institutions. Read the official NCSTRL press package for a description of the background, goals, and organization of NCSTRL.*

Search the NCSTRL collection

- The **Fielded Search Form** allows you to perform a search on several fields of the bibliographic data, and/or to limit the search to specific institutions,
- **Or** enter one or several words into the box below to list all documents in our collection whose author, title, or abstract contain any search word:
- **Or** browse reports at any of the participating institutions.

I want to join NCSTRL, tell me more

Read the faq for institutions interested in participating in the NCSTRL collection.

More Information

Find out what's new with NCSTRL or browse a list of documents related to NCSTRL.

NCSTRL at Cornell Computer Science. Send email to tech-reports@cs.cornell.edu.

[Search | Home page]

Simple Search Results

Search text:

hyperbase

Search Summary:

Organizations you selected are listed below by number of titles found.

- *(1)Virginia Polytechnic Inst. and State University*
- *(1)Boston University*

Search Results:

Virginia Polytechnic Inst. and State University
- *A Query Language for Information Graphs.* Sangita C. Betrabet, Edward A. Fox and Qi-Fan Chen. (TR-93-03)

Boston University
- *Proceedings of the Workshop on Versioning in Hypertext Systems.* David Durand, Anja Haake, David Hicks and Fabio Vitali. (95-001)

[Search | Home page]

NCSTRL
This server operates at Cornell University.
Send email to tech-reports@cs.cornell.edu

Text Linking and Retrieval Experiments for Textbook Components

Gerard Salton, Chris Buckley and Zhongnan Zhao
TR90-1125
May 1990

Experiments are described designed to retrieve individual paragraphs of textbook material in answer to user-submitted queries. The retrieval strategies are based on the global comparison of paragraph texts, as well as on the local processing of text sentences. Furthermore, the retrieved items may be freely chosen, or may alternatively be restricted to certain areas in a clustered arrangement of book paragraphs. The retrieval results indicate that high retrieval values are obtainable for the more refined retrieval strategies, ranging between 0.70 and 0.80 in search precision.

How to view this document

- Display an **overview** of the document in one of the following formats.
  - Overview of thumbnail pages
  - Structural overview
- Display a **selected page** in one of the following formats (document has 14 pages).

raw OCR output
hi-resolution tiff image
inline gif image

1 Display page

- Display the **whole** document in one of the following formats.
  - OCR text (produced by OCR, may have errors) 26029 bytes.
- Print or download all or selected pages.

Digital Libraries for CS

Here are some pointers to Digital Libraries / bibliography servers related to CS.

ACM Digital Library Collection at Virginia Tech
Small test collection of CACM articles from those scanned in as part of the NSF-supported Envision project.
ACM Graphics Bib. DB
SIGGRAPH Online Bibliography Database
ACM HCI Bib. DB
interactions Bibliographies on Human-Computer Interaction
BibNet Project and TeX Users Group FTP bibliographies
bibliography collections from Nelson Beebe including HTML with extensive internal and external hypertext links. See examples: IBM Systems Journal, DEC Technical Journal. See program to build these from BibTeX.
CACM Collection (1959-1979) using Inquery
U. Mass. CIIR demo of Inquery with CACM test collection
Collection of Computer Science Bibliographies
from Alf-Christian Achilles; updated monthly; 790 locally stored bibliographies; more than 530,000 references; 20,000 references contain URLs to an online version of the paper; more than 1600 links to other sites carrying bibliographic information; uses Glimpse
Databases and Logic Programming (mirror)
bibliography server by Michael Ley
NCSTRL
Networked Computer Science Technical Report Library
Univ. of Wales Cardiff CS Courseware
Courseware on Algorithms, AI, C, Graphics, Image Processing, Parallel Processing, Vision, X

Envision

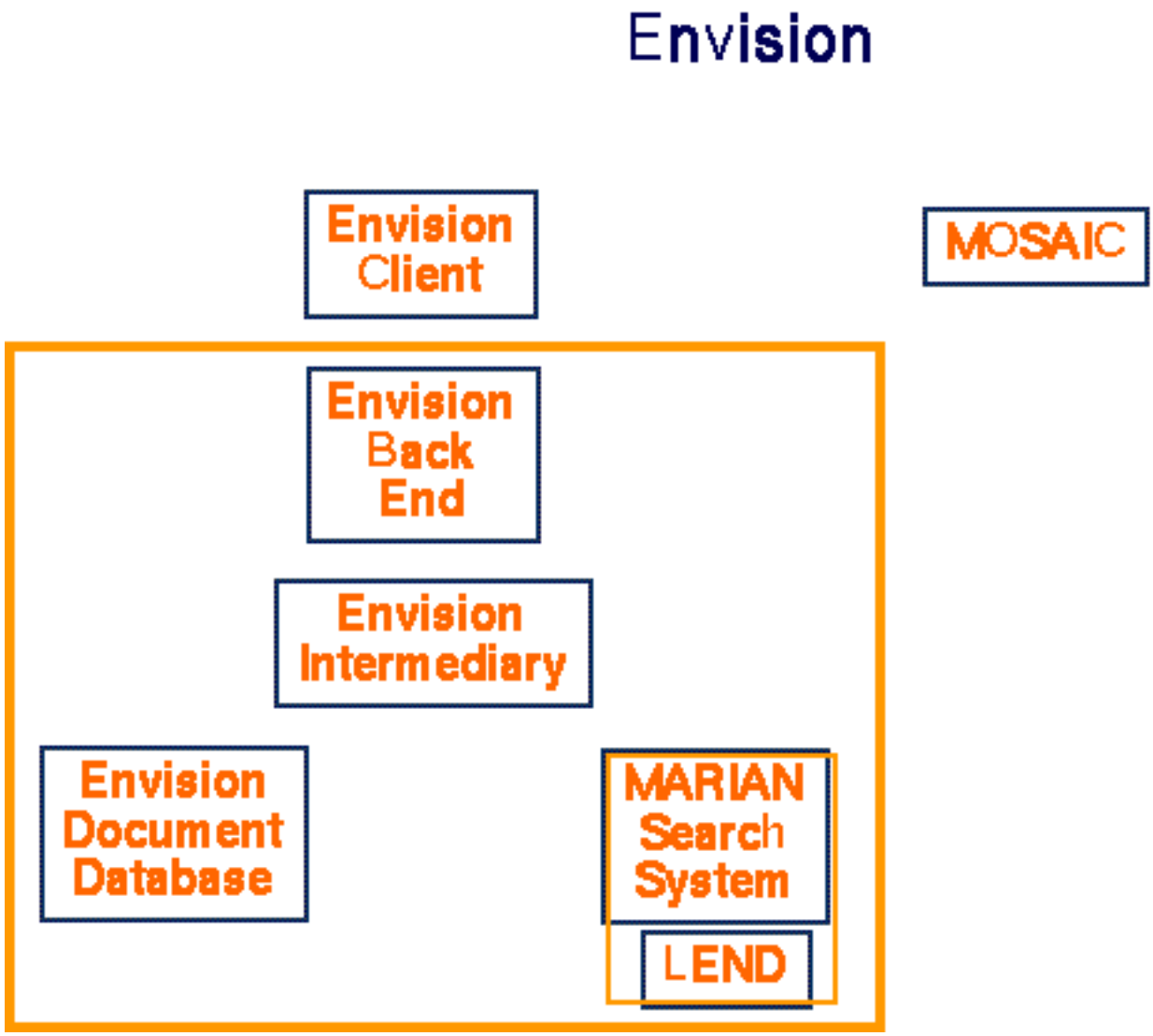
The Envision Project was funded as **A User Centered Database from the Computer Science Literature** by NSF for 1991-95. ACM has provided free access to their publications.

Efforts have concentrated on building an archive based upon SGML, developing an object-oriented database, applying the MARIAN retrieval system and WWW, and constructing a special search interface based upon user wishes.

The interface includes:

- a query screen
- a results list screen
- a results visualization screen
- Mosaic display of retrieved documents

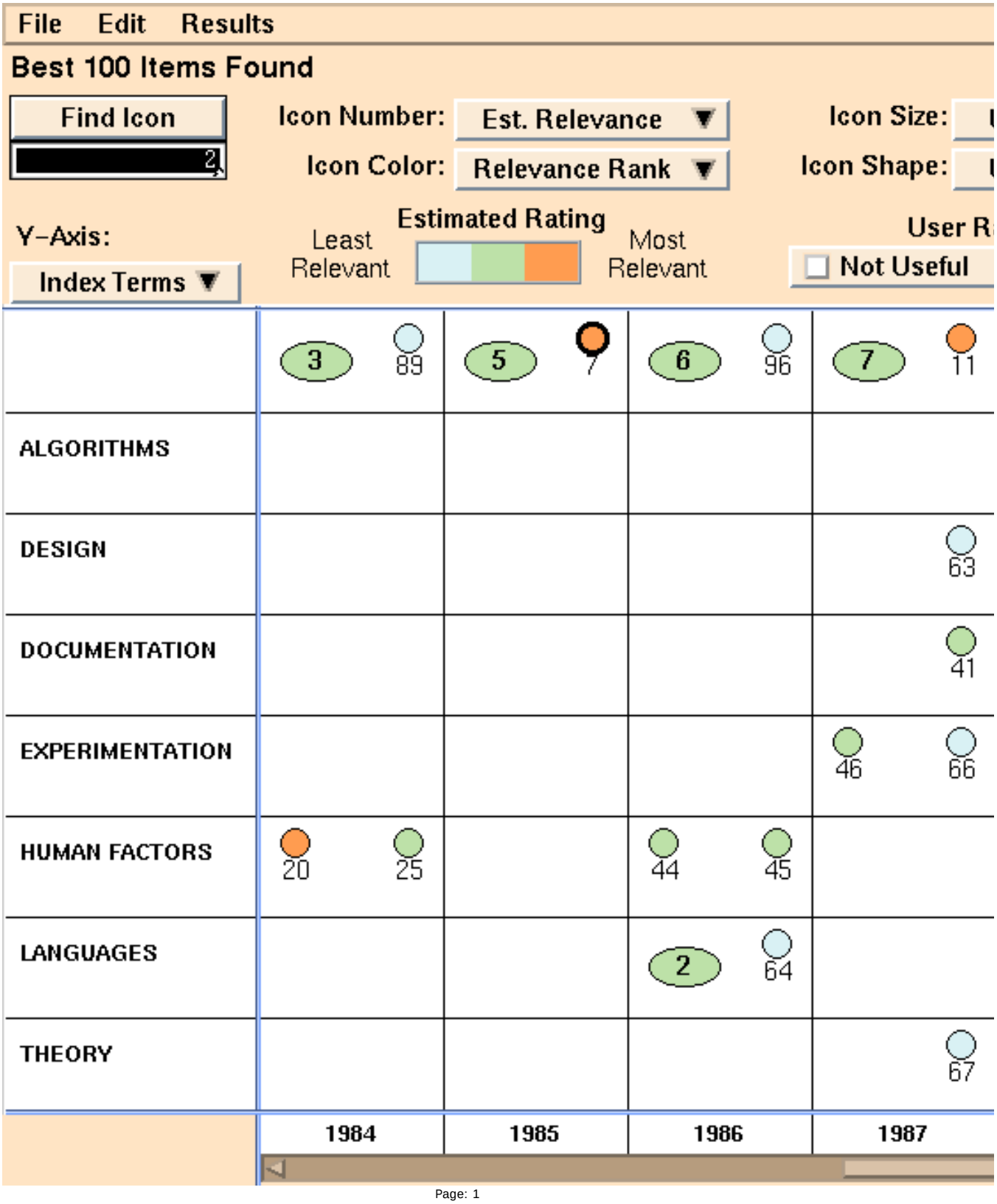
The system architecture is a combination of various elements:



Envision - Results Screens

The interface includes:

- Graphic View 1:



PROJECT ENVISION FINAL REPORT
A User-Centered Database from the Computer Science Literature
NSF Grant IRI-9116991

Edward A. Fox, Lenwood S. Heath, Deborah Hix
Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061-0106

Converted to HTML Wed Jul 5 17:41:14 EDT 1995

Summary of Completed Project

With the support of the National Science Foundation and the Association for Computing Machinery (ACM), the Envision project has developed a prototype digital library of computer science literature that is highly usable (from user-centered design), highly structured (from SGML and an object database), and highly integrated (from hypertext links among objects). The result is a representation of part of the computer science literature as a cohesive body of knowledge that can be searched and viewed in innovative ways. The user interface was designed with careful attention to user needs and desires (through interviews with potential users), to graphic detail (through involvement of an artist and attention to the research literature on graphical perception and psychophysics), and to usability (through an iterative process of usability evaluation). Recognizing the need to translate enormous quantities of documents in an unlimited variety of input formats into a single standard format, the project developed a flexible system for analyzing the structures (e.g., titles, authors, paragraphs, and references) within a document and translating that structure into any standard markup scheme. The Envision distributed server supports simultaneous access to the library by a number of users and in a variety of ways. The Envision software is soon to be installed at ACM headquarters and made available to ACM members. The Envision system will continue in use at Virginia Tech and Norfolk State University to support the work of a related NSF Educational Infrastructure grant.

Technical Information

The list of publications resulting from Envision research appears in the References section. The data collected during this project include electronic versions of computer science literature (Section 2.1). A great deal of software was created or adapted during this project (Section 2.2). A number of people have contributed to the success of the Envision project. These are listed in Appendix A. We are particularly proud of the number of undergraduate students who were able to obtain research experience on the Envision project.

Computer Science Literature

The library contains bibliographic records, full-text articles, and scanned page images. The bulk of the approximately 100,000 bibliographic records are from ACM's *Computing Archive.* We have also incorporated publicly available bibliographies from Ohio State University, the University of Arizona, and the University of Melbourne. We have approximately 700 full-text articles from *Communications of the ACM* and several of the

ACM *Transactions.* Finally, we have about 13,000 scanned page images, from various ACM publications and the technical report series of the Virginia Tech Department of Computer Science.

Envision Software

The major software components of the Envision system are the following.

1. **The Envision Client.** This component interacts with a user to accomplish the tasks of querying the Envision library and visualizing result sets in the Envision graphical display. This client interface is a major innovation of the Envision project and required the greatest amount of effort in interaction design and evaluation, in software design, and in software development.
2. **A WWW Viewer.** Envision employs a WWW browser as its presentation front end. Currently we use Mosaic running on a UNIX workstation.
3. **The Envision Intermediary.** This component communicates with the Envision client over the network to maintain session information, packages queries for the MARIAN search system, and packages result sets to pass back to the Envision client.
4. **The MARIAN Search System.** This component, developed in a separate research effort to access a library catalog, searches the Envision library for documents relevant to the user's query. The search can be based on a combination of title, author, and content words. Result sets are ranked by estimated relevance.
5. **Enhanced WWW Server.** Envision documents are viewed via a WWW interface that accesses a WWW server enhanced by CGI scripts that retrieve Envision objects from the object database and package them into HTML for presentation.
6. **The Object Database.** The Envision object database maintains our view of the structure of the library in terms of classes such as document, person (author), institution, publication, and keywords. Objects in this database refer to related objects, providing a rich hypermedia structure.
7. **The DELTO System.** The DELTO (Document Analysis and Translation) system addresses the need to convert documents in many ill-defined input formats that are received for inclusion in the Envision library into the standard SGML structural representation needed by the Envision object database and MARIAN searchers. This system emphasizes flexibility and automation. DELTO is a major innovation of the Envision project.

Components 1 and 2 run under the X Window System; these have been tested on Sun, DECstation, and DEC Alpha workstations. Components 3 and 4 run on a NextStation. Components 5, 6, and 7 run on a DEC Alpha and should port easily to other UNIX systems.

A public release of the Envision software is due during the summer of 1995. The Envision client will be freely available over the Internet by anonymous ftp from Virginia Tech. Initially, the server components (3, 4, 5, 6, and 7) and the actual library of electronic documents will be released to the ACM, as well as used in a related NSF Educational Infrastructure project at Virginia Tech and Norfolk State University.

References

1 G. A. Averboch. A system for document analysis, translation, and automatic hypertext linking. Master's thesis, Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1995.

2 S. Betrabet, E. A. Fox, and Q. Chen. A query language for information graphs. Technical Report TR 93-03, Department of Computer Science, Virginia Polytechnic Institute and State University, 1993.

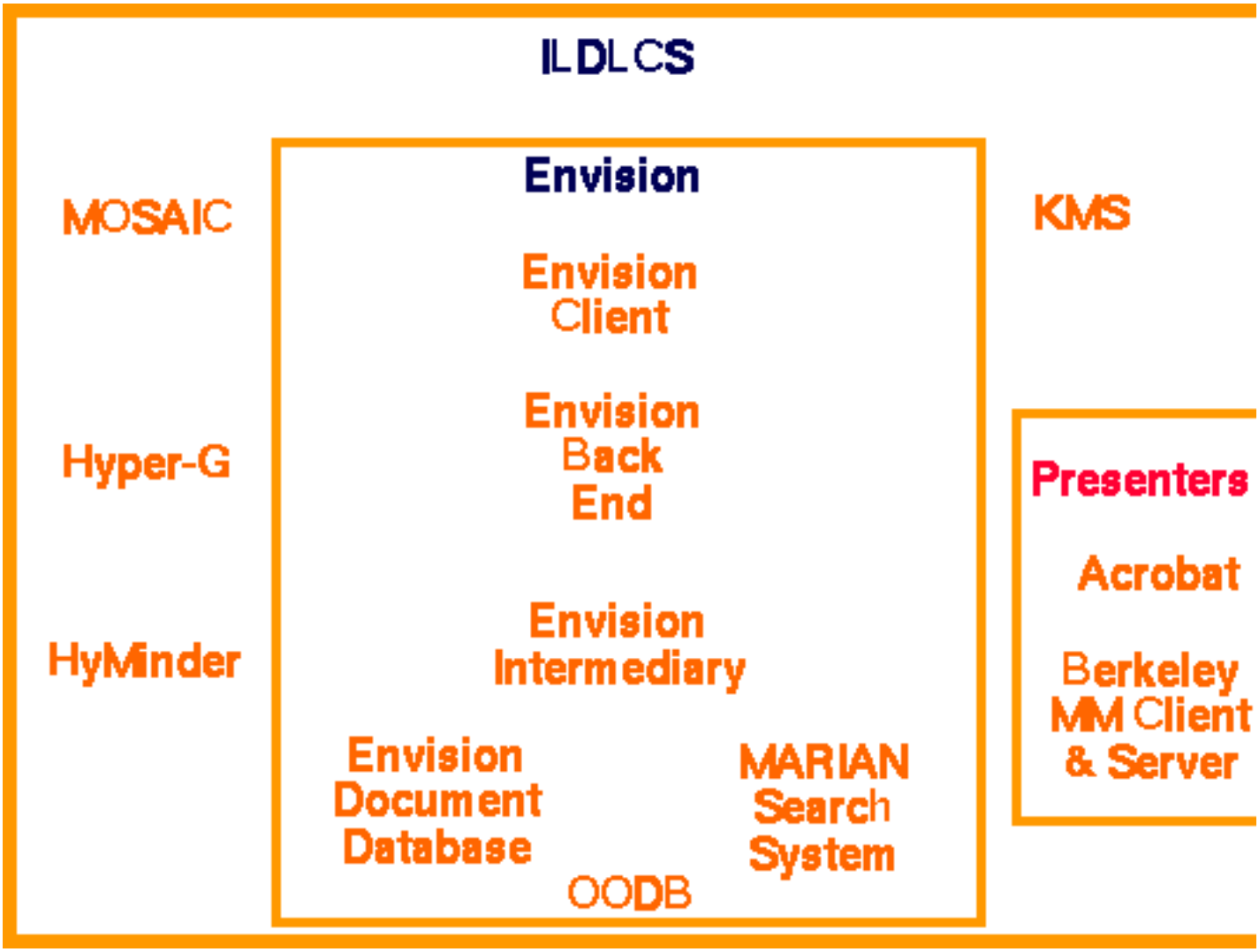
3 D. J. Brueni, B. Cross, E. A. Fox, L. S. Heath, D. Hix, L. T. Nowell, and W. C. Wake. What if there were desktop access to the computer science literature? In *Proceedings of the 21st Annual ACM Computer Science Conference*, pages 15-22, 1993. Also available as Tech. Report TR 92-42,

ILDLCS

The ILDLCS Project was funded as **Interactive Learning with a Digital Library in Computer Science** by NSF for 1993-96. ACM has provided free access to their publications, as have several other publishers. Norfolk State University is a partner in this effort, which building upon the Envision Project. More details are given online.

Efforts have concentrated on developing courseware for 4 courses that have been redone in paperless manner, constructing tools to help with algorithm visualization, and extending the Envision efforts to help with as many CS courses as possible.

The system architecture is a combination of various elements:



CS Courses

Welcome to one of the largest (over 25 courses, over 4500 nodes) repositories of Computer Science courseware! I hope you benefit and send me comments and suggestions!
Regards, Prof. E. A. Fox for
Virginia Tech CS Dept.'s NSF Education Infrastructure Project

- MaSc1044: Computer Science: A Liberal Arts Approach
- CS1206: Operating System Tools
- CS1604: Computers and Networked Information
- CS1704: Introduction to Data Structures & Software Engineering
- CS2304: Self Study Programming in C
- CS2504: Introduction to Computer Organization
- CS2604: Data Structures and File Processing
- CS2704: Object-Oriented Software Design and Construction
- UH3004: High-Performance Scientific Computing
- CS3204: Operating Systems
- CS/Math 3414: Numerical Methods
- CS3604: Professionalism in Computing
- CS4104: Data and Algorithm Analysis
- CS4114: Formal Languages
- CS4124: Theory of Computation
- CS4204: Computer Graphics
- CS4214: Simulation and Modeling
- CS4624: Multimedia, Hypertext and Information Access
- CS4984: Introduction to Human-Computer Interaction
- CS5014: Research Methods in Computer Science
- CS5024: Models and Analysis
- CS5034: Models of Computation
- CS5114: Theory of Algorithms
- CS5204: Operating Systems
- CS/EE5515: Computer Architecture
- CS5604: Information Storage and Retrieval
- CS6104: Symbolic Computation
- CS6204: The World-Wide Web: Beyond the Basics

- CS6404: Advanced Topics in Mathematical Software
- CS6604: Interactive Accessibility

Catalog Pages

- Ugrad
- Grad

Class Data Archives

Searching All Courses and other Pages on ei.cs.vt.edu

Summary about Harvest collection from ei.cs.vt.edu

Usage Statistics

See also

- CS listing for World Lecture Hall
- NSF Computer Science Courseware Repository (NSFCSCR)
- Computational Science Education Project

DIGITAL LIBRARY INITIATIVE

The Initiative's focus is to dramatically advance the means to collect, store, and organize information in digital forms, and make it available for searching, retrieval, and processing via communication networks -- all in user-friendly ways.

Funded through a joint initiative of the NSF/ARPA/NASA Digital Library Initiative

IEEE *Computer's* May 1996 Special Issue on Digital Libraries

Carnegie-Mellon University

Full-content search and retrieval of video

Principal Investigator: Howard Wactlar

Contact: Colleen Everet, (412)268-7674

Stanford University

Interoperation mechanisms among heterogeneous services

Principal Investigator: Hector Garcia-Molina

Contact: Maryanne Siroker, (415)723-0872

University of California at Berkeley

Work-centered digital information services

Principal Investigator: Robert Wilensky

Contact: Crystal Williams, (510)642-0930

University of California at Santa Barbara

Spatially-referenced map information

Principal Investigator: Terrence R. Smith

Contact: Patty Towne, (805)893-7665

University of Illinois at Urbana-Champaign

Federating repositories of scientific literature

Principal Investigator: Bruce Schatz

Contact: Susan Harum, (217)244-8984

University of Michigan

Intelligent agents for information location

Principal Investigator: Daniel Atkins

Contact: Laurie Crum, (313)763-6035

Outline
Sponsors
Links
Search
Home

A variety of research supports the ability of the Informedia Digital Library to index and retrieve video, audio, text and image materials:

Informedia project publications

Research Area Descriptions

- Image Understanding
- Natural Language Processing
- Speech Recognition
- Human Computer Interfaces
- Networked Data Transport
- Network Billing and Security

The NetBill Overview

NetBill: An Internet Commerce System Optimized for Network Delivered Services

Marvin Sirbu, Engineering and Public Policy Department
J. D. Tygar, Computer Science Department

Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

As the explosive growth of the Internet continues, more people rely on networks for timely information. However, since most information on the Internet today is free, intellectual property owners have little incentive to make valuable information accessible through the network. There are many potential providers who could sell information on the Internet and many potential customers for that information. What is missing is an electronic commerce mechanism that links the merchants and the customers.

NetBill is a business model, set of protocols, and software implementation allowing customers to pay owners and retailers of information. While NetBill will enable a market economy in information, we still expect that there will be an active exchange of free information.

The highlights of the NetBill model include:

- Has a very low transaction costs for micropayments (around 1 cent for a 10 cents item)
- Protects the privacy of the transaction
- highly scalable
- certified delivery mechanism which delivers information goods if and only if the customer has payed for them.

This paper discusses the design of the NetBill protocol and our World Wide Web (WWW) prototype implementation

Note: the paper is contained in a single file (49Kbytes). You can also get a postscript version. The following links can be used to go to a specific section of the paper.

Next Up Previous
Next: Task Planner Agent **Up:** Task Planning Agents in **Previous:** Introduction

Task Planning in the UMDL Architecture

A fundamental activity in the University of Michigan Digital Library (UMDL) is connecting people/agents that need help accomplishing their tasks with people/agents that are capable of doing those tasks. A canonical example is that of a library user in search of information who needs to contact the appropriate collections which contain the information. The job of a task planning agent (TPA) is to forge these connections between agents -- to help agents team up with the right agents in a large, open, and constantly evolving network of agents.

We envision that there will be many TPAs within the UMDL. They will have in common the general role of finding resources (agents with capabilities and/or content of interest), but each will possess specific knowledge and procedures for doing so, depending both on the characteristics of the tasks that are in need of resources, and on the resources available for doing the search (e.g. monetary funds, user patience, etc.). The class of tasks that we are initially focusing on are, not surprisingly, query answering tasks. Thus, in this paper, we will consider the subclass of TPAs that are specialized for query tasks, and will focus specifically on how we have designed and built an instance of a TPA for query planning.

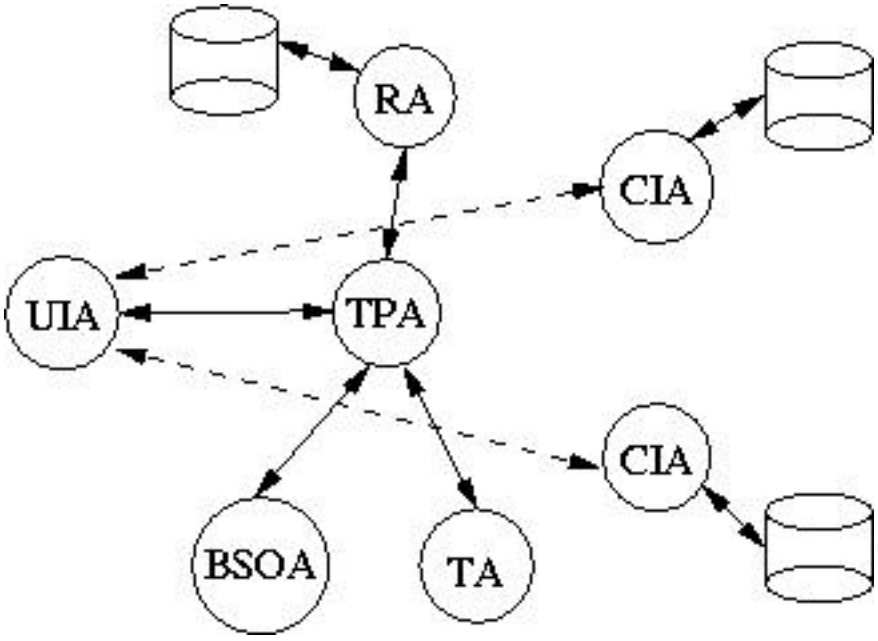


Figure 1: Schematic view of the UMDL architecture showing the agents in it and the communications between them. The solid lines represent the first part of the search, where the UIA looks for applicable collections. The dashed lines are the second part, when the UIA queries the collections that were found. Future versions will have several instantiations of each agent.

A query planning TPA is ultimately responsible for finding one or more collections based on the needs of a user; in order to do so, it communicates with many agents in the UMDL, as seen in Figure 1. The TPA receives the specification of the user's needs, along with parameters concerning task planning such as constraints on the number of collections to find or the effort to expend in finding them, from the User Interface Agent (UIA) which interacts directly with the user. The TPA needs to examine the query task to identify the characteristics of collections that might satisfy it, and from this formulate a query to the Registry Agent (RA) to see whether such collections exist. The TPA can inspect the responses from the RA and might simply forward

UMDL Organization

Picture of UMDL Research Project Categories of Activities (DEA) HERE!

The above graph illustrates the the interaction among several activities of our digital library project. More in-depth activities of many of these teams are represented below. Some of the small teams do not have their own web page.

The UMDL project has a continually evolving set of teams which interact to create the production system and do relevant research. Members of all of these teams meet on the first Friday of each month to report progress, discuss important issues, and get updates on events and activities of the project.

Below is a list of current teams, as well as a "master" list of active project members (does not include people who are loosely affiliated with the project). By selecting one of the teams, you will be able to see a list of members, regular meeting times, and meeting notes and other documentation.

Advanced User Interface Group
This team is primarily concerned with a state-of-the-art user centered design digital library. Their research are long term and results will be folded into the production system.

Architecture
This team develops the agent architecture and conducts research in related areas.

ColSR
ColSR, or Collection Search and Retrieval, is focused on search and retrieval functions related to the overall digital library collection.

Conspectus Definition and Registry
This team works on the development of the conspectus.

ConSR
ConSR, or Conspectus Search and Retrieval, is focused on search and retrieval functions related to the conspectus.

Intellectual Property and Economic Issues

This team is working on commerce mechanisms, intellectual property license management, agent negotiation protocols, and resource allocation issues for the UMDL.

 Operating Committee
This team is comprised of team leaders from all areas.

 Use and Evaluation in Education
This team is responsible for deployment and evaluation of the production system in high schools and public libraries.

 User Interface Design and Evaluation
This team is primarily concerned with the design of the current production system.

 Master List of all active members of the project
This alphabetical master lists provides personal web sites and affiliations for UMDL project members.

 Project Partners
This lists partners, from industry to educational, who are involved in the project.

Return to the Main Page

Comments or questions may be sent to: UMDL.INFO@umich.edu

Stanford Digital Library Testbed Development

Department of Computer Science
Stanford University
Stanford, CA

Development team:

- Scott Hassan
- Andy Kacsmar
- Andreas Paepcke
- Tom Schirmer

What's new!

A major function of the Stanford Digital Library test bed is to allow experimentation with 'glue' for interactions with online services. In view of this requirement we chose a distributed object approach as our base technology. This decision was explained in the slides of our presentation to the first advisory board meeting on January 9, 1994.

We selected Xerox PARC's ILU as our implementation of distributed objects. ILU is roughly an implementation of the Common Object Request Broker (CORBA) standard, providing language bindings for C++, C, CommonLisp, Python and Modula-3. This means that we can do our implementations in any of these languages.

The best way to start is to look through the CORBA/ILU documentation, and then to look at some examples for the language you are interested in (see below). After you think you understand the basics (feel free to consult with us), ILU bugs can be reported to ilu-bugs@parc.xerox.com. General ILU questions can be directed to ilu@parc.xerox.com. To get on the ILU mailing list, send mail to ilu-requests@parc.xerox.com.

Note that for C++ users we have acquired a site license for the products ObjectCenter/ViewCenter/TestCenter. These are a debugger, motif interface builder and memory-leak-debugger/performance tuning tool respectively. They are available on HP platforms under /local/CenterLine/bin. For documentation, see Andreas.

- DL Object Interchange service and examples

We are working on a technical performance evaluation of ILU, HTTP, and basic TCP. We will be comparing ILU's performance to IBM's DSOM and maybe Microsoft's COM.

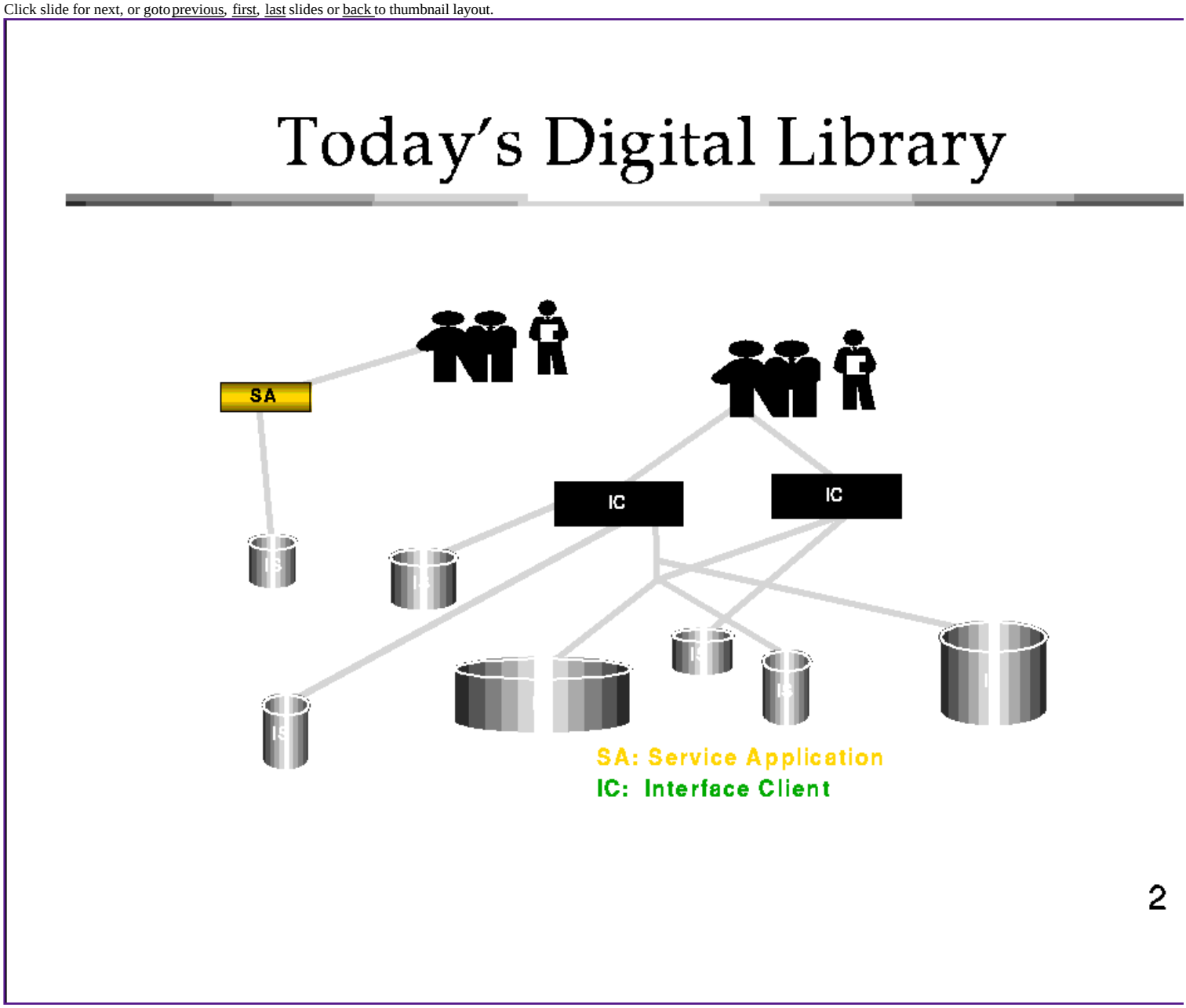
- ILU -- Our installation and examples.
- Python -- Our installation of Python programming language
- ObjectCenter -- Discussion of ObjectCenter
- COS -- Digital Library Testbed Common Object Services
- CVS -- Our use of CVS in the testbed.
- Various Manuals (CVS, Python, ILU)

- Testbed Activity Slides

Digital Libraries Webmaster
Webmaster@diglib.stanford.edu

Testbed: Slide 2 of 12.

Click slide for next, or goto previous, first, last slides or back to thumbnail layout.

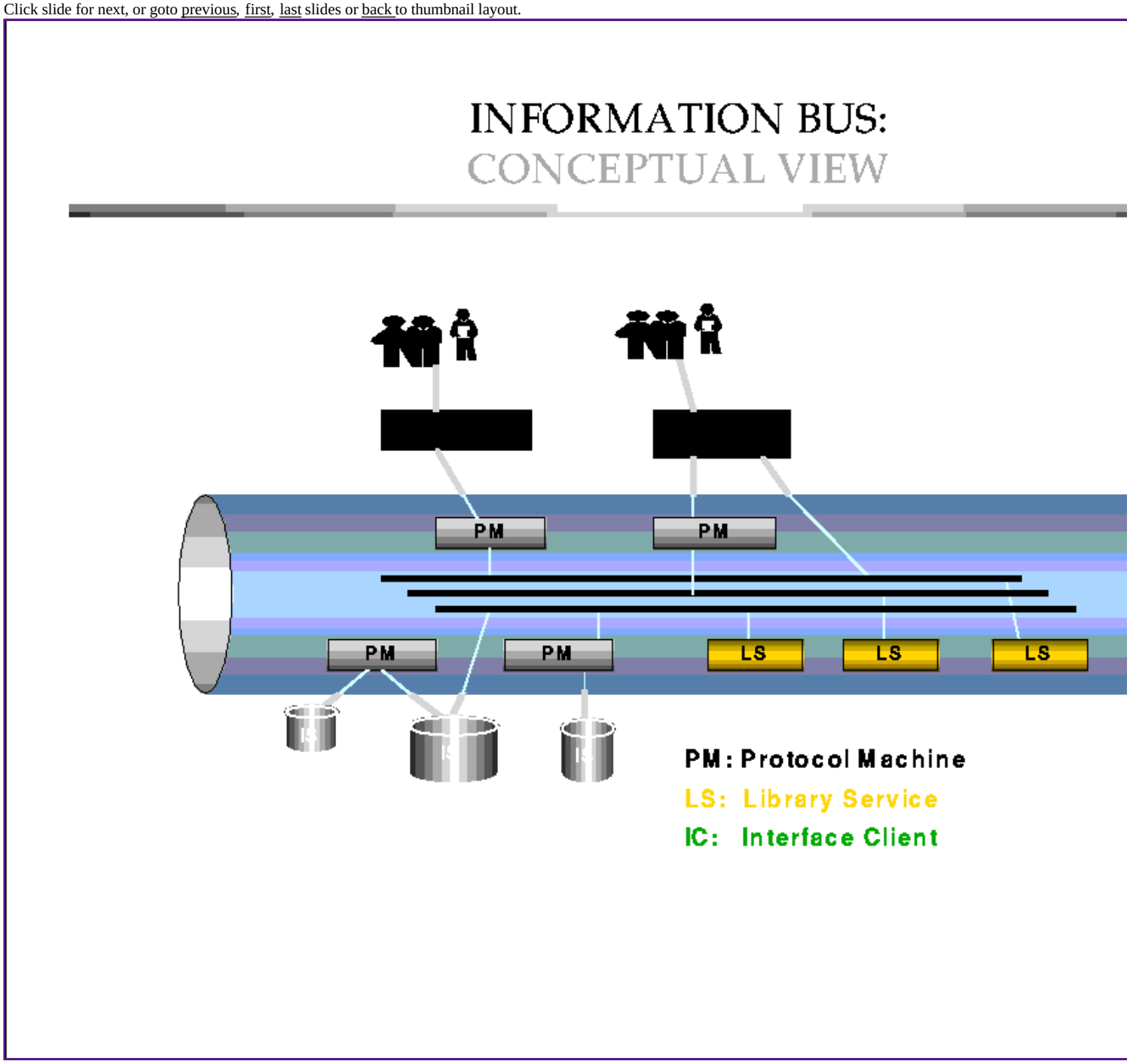


Click slide for next, or goto previous, or back to thumbnail layout.

Digital Libraries Webmaster
Webmaster@diglib.stanford.edu

Testbed: Slide 3 of 12.

Click slide for next, or goto previous, first, last slides or back to thumbnail layout.

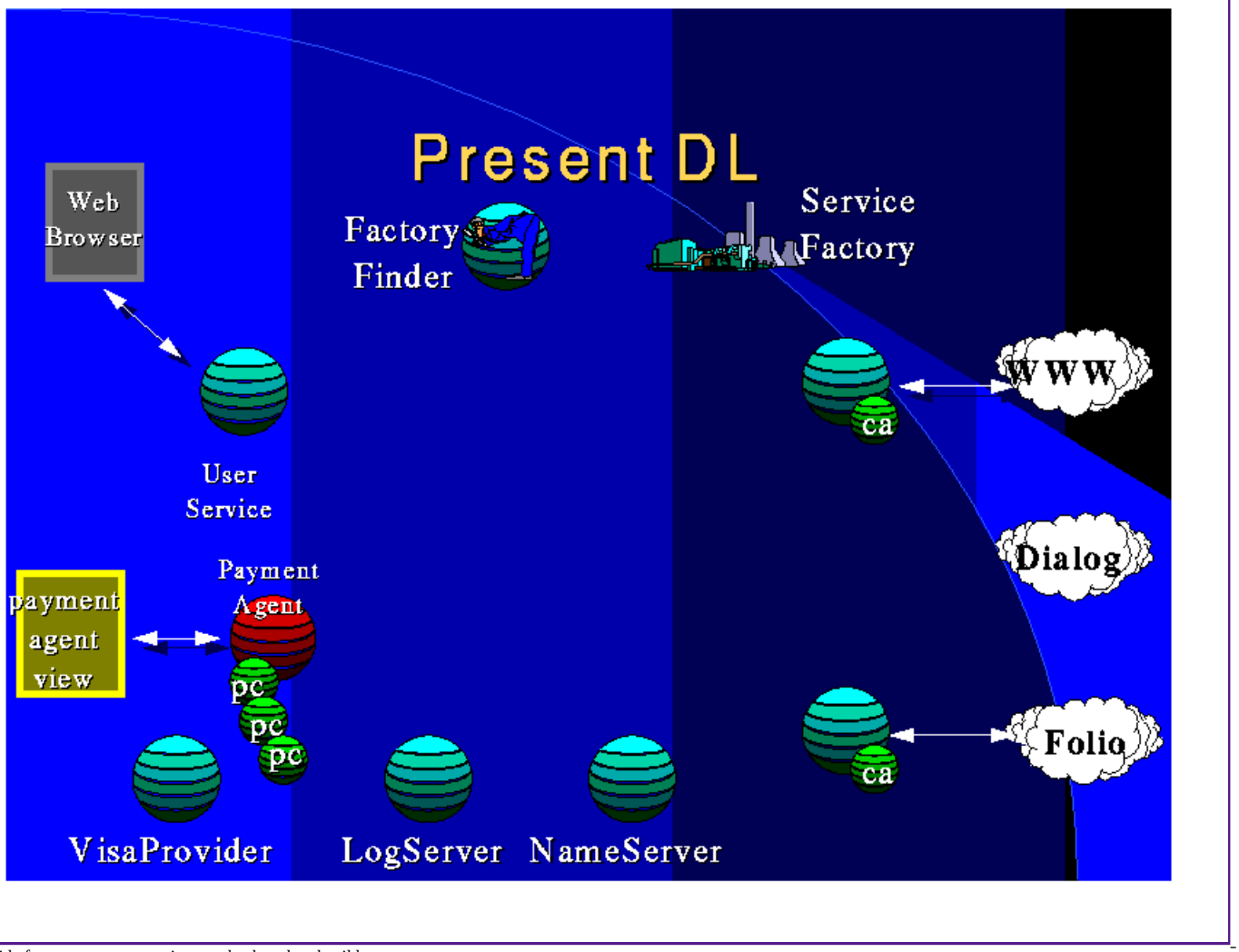


Click slide for next, or goto previous, or back to thumbnail layout.

Digital Libraries Webmaster
Webmaster@diglib.stanford.edu

COS - Common Object Services: Slide 11 of 49.

Click slide for next, or goto previous, first, last slides or back to thumbnail layout.

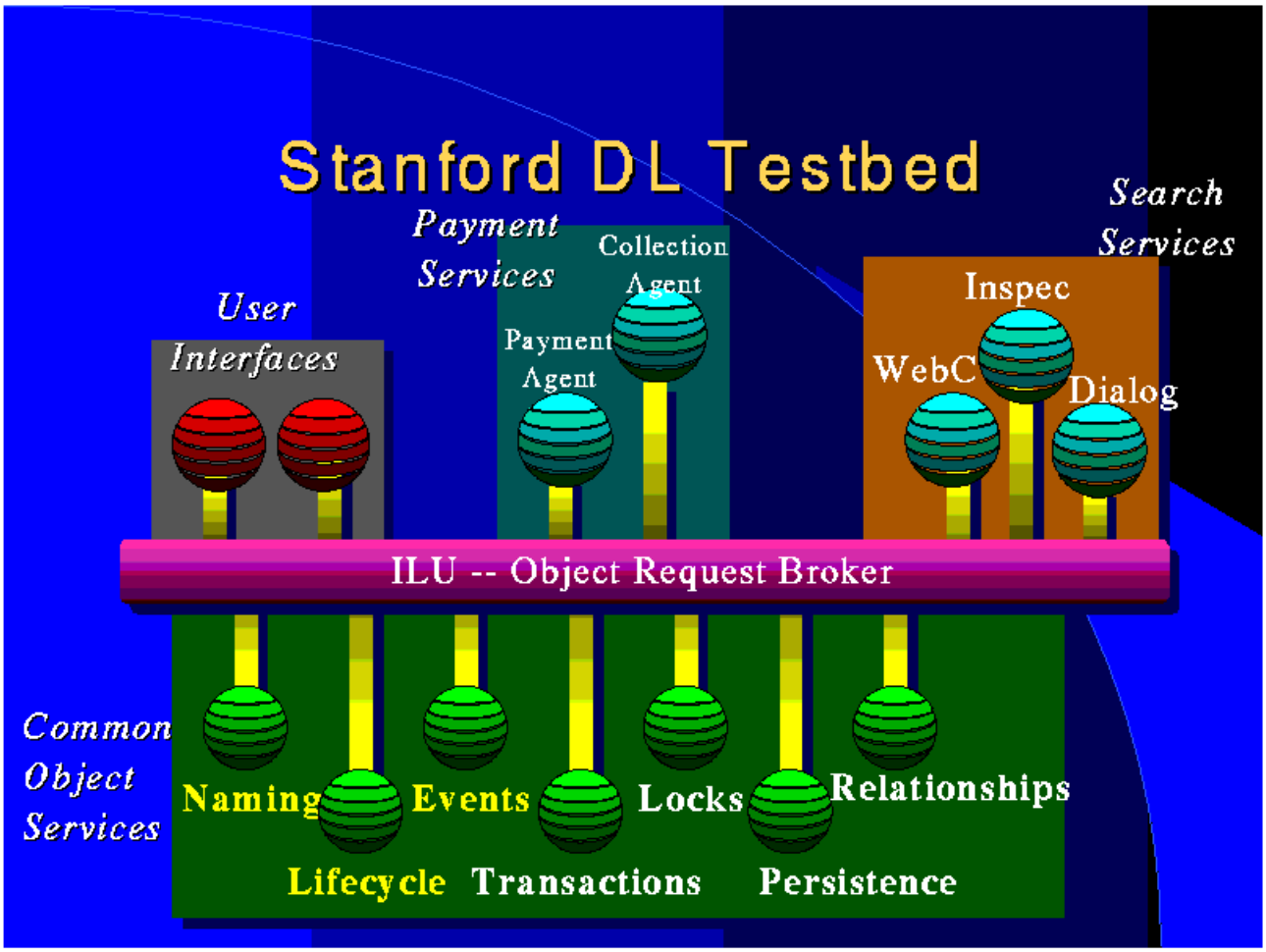


Click slide for next, or goto previous, or back to thumbnail layout.

Digital Libraries Webmaster
Webmaster@diglib.stanford.edu

COS - Common Object Services: Slide 2 of 49.

Click slide for next, or goto previous, first, last slides or back to thumbnail layout.



Click slide for next, or goto previous, or back to thumbnail layout.

Digital Libraries Webmaster
Webmaster@diglib.stanford.edu

COS - Common Object Services: Slide 15 of 49.

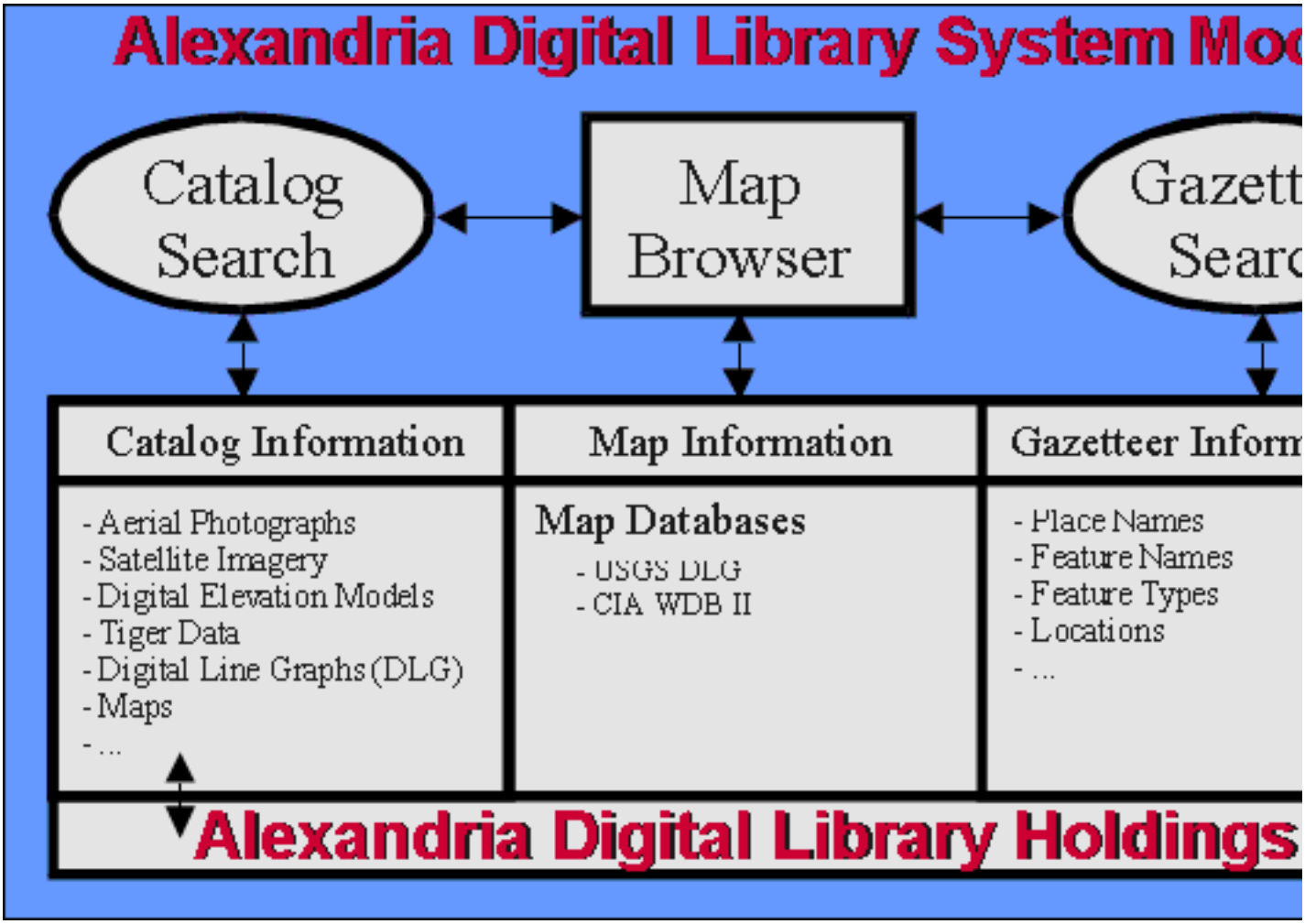
Click slide for next, or goto previous, first, last slides or back to thumbnail layout.

Click slide for next, or goto previous, or back to thumbnail layout.

Digital Libraries Webmaster
Webmaster@diglib.stanford.edu

Tutorial Table of Contents

Prev Next



Conceptual model of the Alexandria Web interface

Building the Interspace:Digital Library Infrastructure for a University Engineering Community

Bruce Schatz, Principal Investigator
University of Illinois DLI project
dli@uiuc.edu

DLI Project-Wide Workshop
November 9, 195 Santa Barbara, CA

Research on the Net

- The Past: Access

The Net fetches documents

- The Present: Organization

The Net searches repositories

- The Future: Analysis

The Net correlates information

- From the Internet (data transmission)
- to the Interspace (information manipulation)

Project Goals

- Semantic Federation (research)
- Distributed Repostories (infrastructure)
- Scientific Literature (testbed)

- evaluate large testbed
- perform technology research

Organizations

- Testbed

Grainger Engineering Library Information Center

(part of University Library, *UL*)

- Infrastructure

NCSA Software Development Group

Semantic Federation from Distributed Repositories of Scientific Literature

Bruce Schatz, Principal Investigator
University of Illinois DLI project
dli@uiuc.edu

DLI Project-Wide Workshop
November 10, 1995 Santa Barbara, CA

Levels of Federation

- Syntactic
  - connection protocols (translation gateways)
- Structural
  - field names (query normalization)
  - field values (tag normalization)
- Semantic
  - context (term co-occurrence)
  - meaning (content parsing)

Testbed Federation

- Index with Document Structure

Tag normalization for field values

- Deposit with common tags after transform

problems with sections and with authors

- Search across multiple repositories

Query normalization for field names

- Gateway maps multiple protocols

problems with distribution and definition

- Display integrates multiple views

multiple sources at multiple levels

Semantic Retrieval

- automatic indexing of concepts
  - find context of phrases within documents
  - generates a concept space based on term frequency

[text-only]

The LIBRARY of CONGRESS

AMERICAN MEMORY

Historical Collections for the National Digital Library

SEARCH
American Memory Collections

BROWSE
List of all American Memory Collections

LEARN
Organized help for using the collections

American Memory consists of primary source and archival materials relating to American culture and history. These *historical collections* are the key contribution of the Library of Congress to the National Digital Library. Most of these offerings are from the Library's unparalleled special collections.

Access Collections by Type


Prints & Photos


Documents


Motion Pictures


Sound Recordings

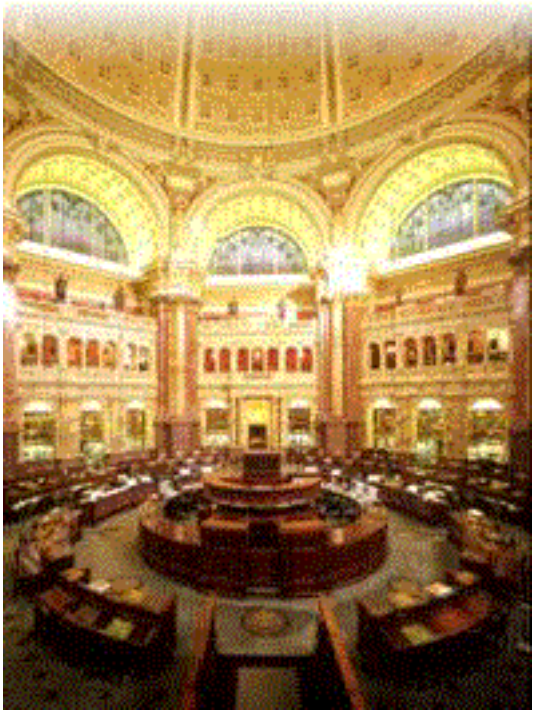
Showcase

Three new collections:

Evolution of the Conservation Movement, 1850-1920 (manuscripts, legal documents, photographs)
Gottscho-Schleisner (photographs)
Horydczak (photographs)

LIBRARY OF CONGRESS ❖ AMERITECH

National Digital Library Competition



***The*Library *of* Congress Home Page**

Library of Congress
Comments:lcweb@loc.gov (07/03/96)

Electronic Thesis and Dissertation Project

"For students, the electronic dissertation can be easier to prepare, more error free, less expensive, and more flexible in format. It can also allow more creativity on the part of the author by permitting inclusion of hypertext links and, soon, digital audio and video recordings." -- Dr. John Eaton

---

- **Sponsor:** Southeastern Universities Research Association (SURA)
- Funding Source: Department Number 041
- FRS Account Number: 4-35070

Students

 **Library Collection of ETDs**

 **Recent Library ETD Additions (*NEW*)**

 **Submission Guidelines**

 **LaTeX Templates and Instructions**

 **Frequently Asked Questions (FAQ)**

 **Currently Scheduled Workshops**

<ETD> Research

Virginia Tech <ETD> Submission Form

This form will not work properly unless
you are using Netscape Navigator 2.0 or better.

Instructions: Please fill out the form completely. Cut and paste, from your document and into the form, as necessary. Read the help file for help on cutting and pasting your abstract and for selecting keywords. Once you are done filling out the form read the copyright statement at the bottom of the page and if you agree to it click "Preview".

Document Type:
Select the type of document you are submitting.

- (•) Thesis
- ( ) Dissertation
- ( ) Special Report
- ( ) Major Paper
- ( ) Technical Report

Name:
Enter your name just as it appears on the title page.

Email address:
Enter your email address.

Adjust the width of your browser so that both ends of the "Title" field are visible.

Title:
Enter the title just as it appears on the title page.

Degree:
Enter degree.

Department:
Enter department.

Approved:
Enter the names and positions (ex: Chair, Co-Chair) of the people on your review committee.
Also enter the email address of the Chair, or the most senior Co-Chair.

Chair **email:**

Electronic Thesis and Dissertation Project

Frequently Asked Questions

about File Formats

What are the accepted file formats for ETDs?

Here is the current list of file formats we accept thus far. Please keep in mind that this list is growing.

Text

- ASCII (.txt)
- SGML according to the document type: "etd.dtd" (.etd)
  -- Note: We recommend Unicode for non-Roman characters.

Images

- PDF (.pdf)
  -- use Type I PostScript fonts
- JPEG (.jpg)
- CompuServe GIF (.gif)
- TIFF following version 6.0 or later, including CCITT G4 (.tif)
- CGM Computer Graphics Metafile (.cgm)
- PhotoCD
  -- Note: We recommend a minimum of 600 dpi resolution for images of pages with text.

Video

- MPEG (i.e., MPEG-1, MPEG-2) (.mpg)
- QuickTime - Apple (.mov)
- Audio Video Interleaved - Microsoft (.avi)

Audio

- MPEG-2
- CD-DA
- CD-ROM/XA (A or B or C)
- AIF (.aif)
- SND (.snd)
- WAV (.wav)
- MIDI (with timing information) (.midi)

Authoring

- Authorware
- Director (MMM, PICS)

Special

- Spreadsheet - Excel (.xcl)
- AutoCAD (.dxf)

Electronic Thesis and Dissertation Project

Principles

Principle 1:

We should allow parts of ETDs to be encoded according to all widely used international standards (e.g., JPEG). Other representations should be selected based on agreement of an ETD Standards Committee.

Principle 2:

References to objects outside the ETD should be restricted to widely used reference methods that are descriptive in nature (e.g., give a page range in a proceedings of a conference whose city, dates, sponsor, name and editors are provided) or that follow some persistent naming scheme (e.g., ISBN, ISSN plus vol/no/page, URN).

Principle 3:

Encoding should be done in a way that allows recovery of all critical details. Thus, if a page of text is encoded as a bitmap, the smallest characters on the page must be clearly readable without any ambiguity. If an image is encoded, the smallest details relevant to the author's purpose in inclusion of the image must be accurately rendered. Thus, the emphasis should be on creating an archival representation, not on one that is easily or quickly rendered with current devices.

Principle 4:

Suitable metadata must be provided for all digital objects, as called for in ETD requirements or in the author's discipline, e.g.:

- **author/creator,**
- **permission details if not by the ETD author,**
- **date of origination if not that of the ETD,**
- **any details of origination and/or capture that would be needed by one wishing to correctly render the digital object, e.g.,**
  - **scanner make and model used,**
  - **settings/calibration at time of capture.**

Principle 5:

Quality is important. If analog devices are involved, they should be calibrated and tested in advance so an accurate recording is made.

Principle 6:

World Wide Web (WWW) Traffic Analysis Research

Computer Science Department
Virginia Polytechnic and State University
Blacksburg, VA 24061-0106

The mission of our research group is to

- collect and make available to other researchers a collection of Web traffic traces from a variety of networks,
- work to make the use of proxy caches more effective through performance evaluation of different proposed cache designs, and
- production of tools to assist in the collection and analysis of Web traffic and in the evaluation of cache designs.

Resources currently available:

Marc Abrams, Stephen Williams, *Complementing Surveying and Demographics with Automated Network Monitoring,* to appear in *World Wide Web Journal*, June 1996.

Ghaleb Abdulla, Marc Abrams, Edward A. Fox, *Scaling the WWW,* submitted for publication, March 1996.

Stephen Williams, Marc Abrams, Charles R. Standridge, Ghaleb Abdulla, Edward A. Fox, *Removal Policies in Network Caches for World-Wide Web Documents,* Proceedings, ACM Sigcomm Conference, August 1996.

Marc Abrams, Stephen Williams, Ghaleb Abdulla, Shashin Patel, Randy Ribler, Edward A. Fox, "Multimedia Traffic Analysis Using Chitra95," *Proceedings: ACM Multimedia '95,* San Francisco CA, November 1995. pp 267-276.

(Also available as uncompressed postscript [2.25Mb] or gzip'd postscript [46Kb].) Discusses the use of Chitra95 for analyzing WWW trace data, and illustrates with the analysis of three educational workloads.

Marc Abrams, Charles R. Standridge, Ghaleb Abdulla, Stephen Williams, Edward A. Fox, "Caching Proxies: Limitations and Potentials," *Proceedings: 4th Inter. World-Wide Web Conference,* Boston, MA, Dec. 1995. pp 119-133.

(Also available as uncompressed postscript [0.64M] or gzip'd postscript [46K].) Contains a performance study of the effectiveness of proxy servers that cache documents requested by WWW clients, using workload data collected with Chitra95.

 *Experience in Network Delivery of Computer Science Courseware*, transparencies from presentation at the 2nd Annual SUCCEED Conference, N.C. State, March 1995. Describes using Chitra94 to analyze World-Wide Web traffic from Computer Science courses at Virginia Tech.

 WWW cache simulation used in certain papers listed above

 WWW data collection tools used in certain papers listed above, and slides from a presentation on some of the tools [Adobe pdf or postscript versions].

 Chitra, a tool that can be used to analyze trace data, including traces of traffic from the WWW. The trace formats currently supported are common log format" and the TCPdump tool. (The next release of the tool is expected in Spring 1996.)

Research team members:

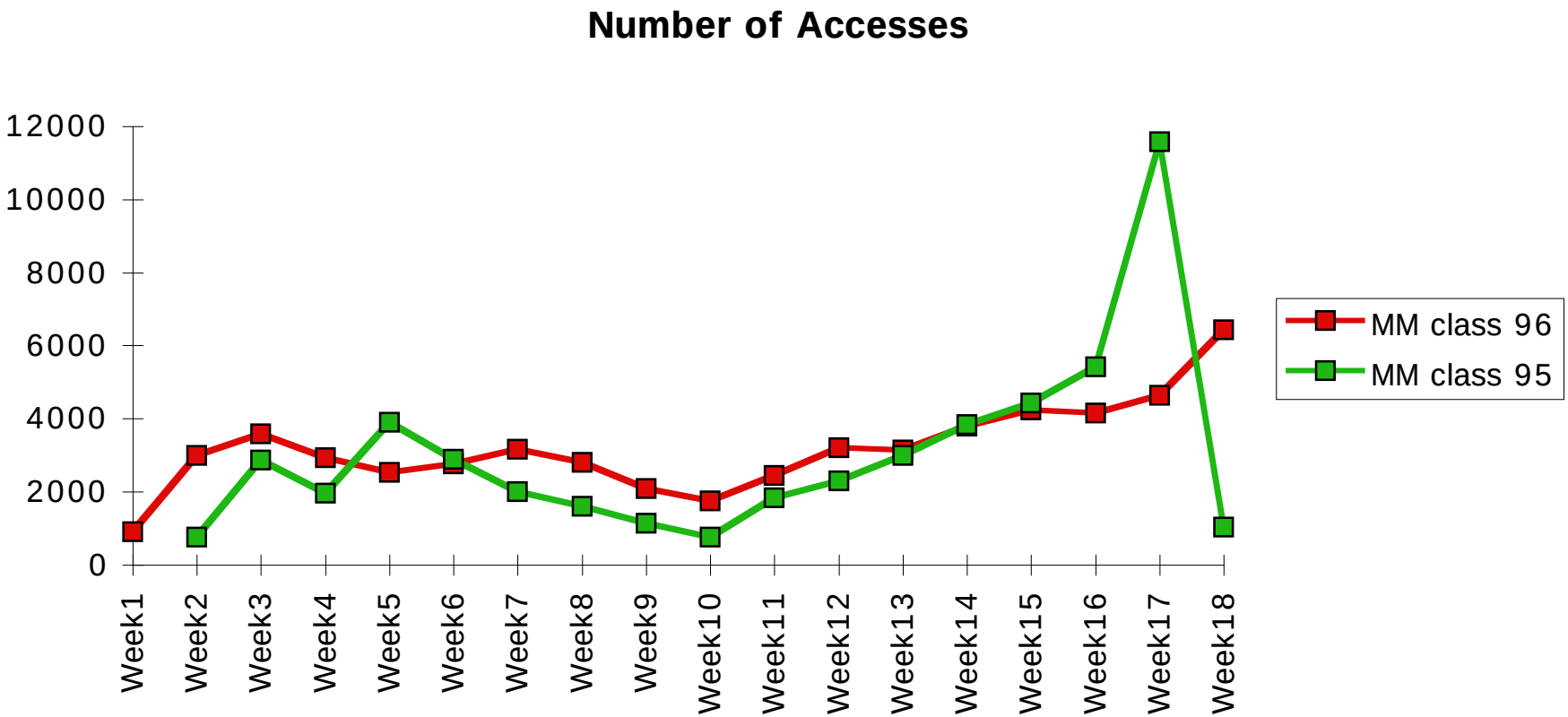
- Marc Abrams [URL: http://www.cs.vt.edu/vitae/Abrams.html]
  Computer Science Department
  Virginia Tech
  Blacksburg,VA 24061-0106 USA
  E-Mail:abrams@vt.edu

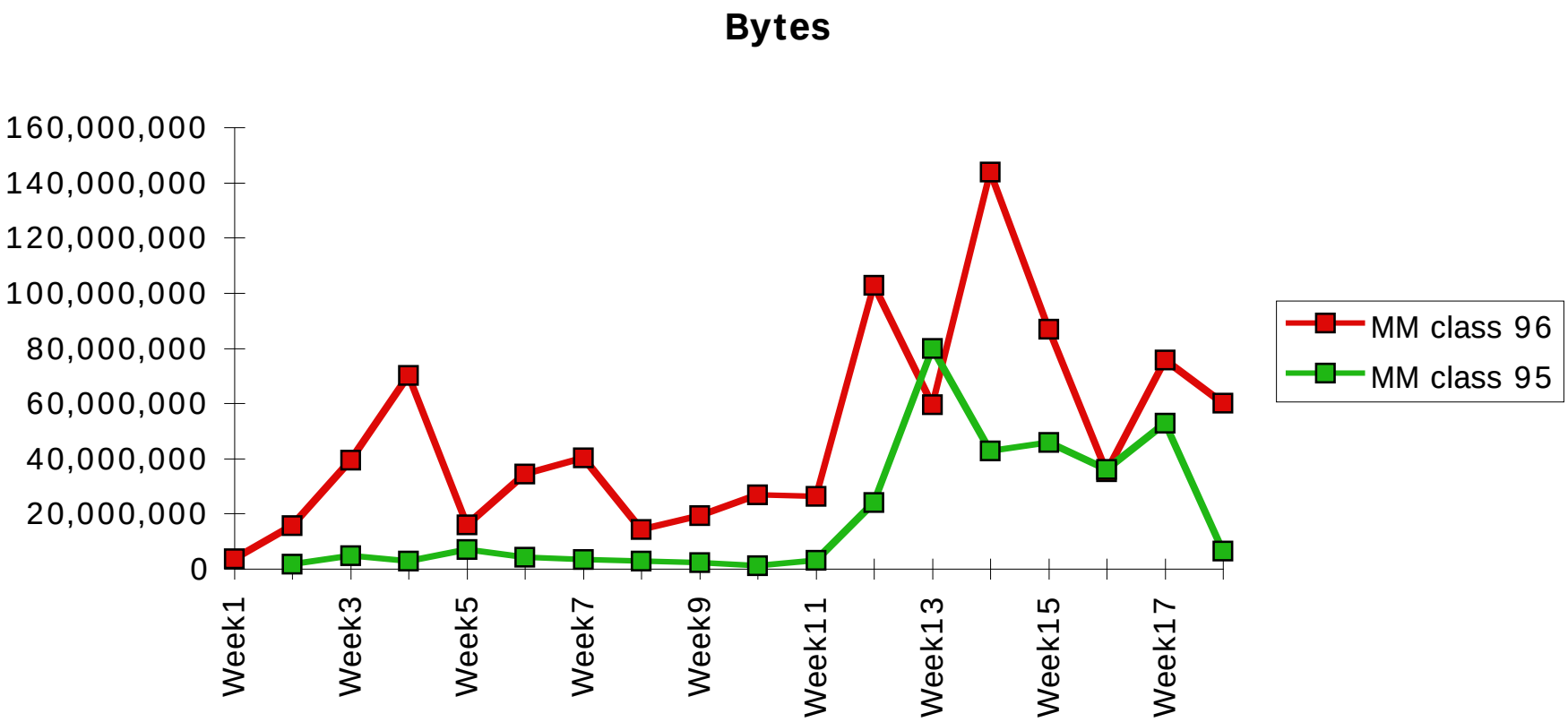
- Ghaleb Abdulla [URL: http://csgrad.cs.vt.edu/~abdulla/]
  Computer Science Department
  Virginia Tech
  Blacksburg,VA 24061-0106 USA
  E-Mail:abdulla@csgrad.cs.vt.edu

- Edward A. Fox [URL: http://fox.cs.vt.edu/]
  Computer Science Department
  Virginia Tech
  Blacksburg,VA 24061-0106 USA
  E-Mail:fox@vt.edu

- Charles R. Standridge
  FAMU-FSU College of Engineering
  Tallahassee, FL 32310, USA
  E-Mail:stand@evax11.eng.fsu.edu

- Stephen Williams [URL: http://csgrad.cs.vt.edu/~williams/]
  Computer Science Department
  Virginia Tech
  Blacksburg,VA 24061-0106 USA


Number of Accesses
12000
10000
8000
6000
4000
2000
0
Week1
Week2
Week3
Week4
Week5
Week6
Week7
Week8
Week9
Week10
Week11
Week12
Week13
Week14
Week15
Week16
Week17
Week18
MM class 96
MM class 95


Bytes
160,000,000
140,000,000
120,000,000
100,000,000
80,000,000
60,000,000
40,000,000
20,000,000
0
Week1
Week3
Week5
Week7
Week9
Week11
Week13
Week15
Week17
MM class 96
MM class 95

Z39.50 resources - a pointer page

****
The Library of Congress is the official maintanence agency for Z39.50. As such they are the place to go to get the most official current legal information related to Z39.50. This page you are reading may phase out (though not soon) as they develop their page (started July 1995).
****

This page is meant as a reference point for resources related to the Information Retrieval Service and Protocol standard, ANSI / NISO Z39.50. This standard was first successfully balloted in 1988; several companies implemented this standard or variants of this; but it did not develop large scale acceptance. A noteworthy implementation based on this standard is WAIS (Wide Area Information Services). Also see the **Profiles** section for more info on present development of WAIS within Z39.50.

The standard was significantly rewritten for its next version. This is ANSI/NISO Z39.50-1992 (Version 2). One important step in this version of the standard was alignment with ISO 10162/10163, the Search and Retrieval (SR) Service Definition and Protocol Definition. Also beginning with this version, the protocol data units are described in ASN.1 (A "Layman's Guide" to ASN.1 is available from RSA) -- The Version 3 ASN.1 is available as flat ascii as well as in a wonderfully useful HTML format. from Library of Congress's various servers.

The next version (Version 3) of the standard was balloted in December 1994, and officially accepted by ANSI in July 1995. The official version of the standard is available electronically, at the Library of Congress's ftp server (ftp.loc.gov). Note this is a copyrighted document - many thanks to whoever achieved this electronic availability. The official text is available in postScript and wordPerfect, in four parts:
postscript: Part1, Part2, Part3, and Part4.
WordPerfect: Part1, Part2, Part3, and Part4.

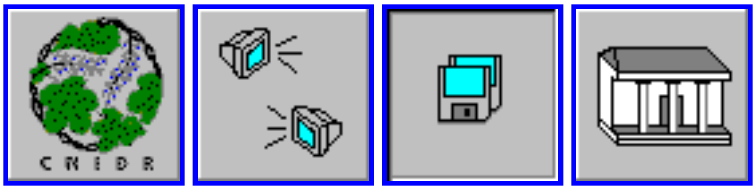
The Z39.50 ImplementorsGroup (ZIG) works closely with the standard's maintenance agency, the Library of Congress. This group meets 2 - 3 times a year and has discussions on its listserv Z3950IW@NERVM.NERDC.UFL.EDU. For meeting minutes, more about the LISTSERV, scheduled future meetings, and other related information check out the relevant sub-section at Library of Congress

Freely available implementations of Z39.50 and related code are starting to become available. Those I know of (let me know of others) are:

- CNIDR's Isite, Isearch, FreeWAIS, etc
- Index Data, a software development enterprise operating out of Copenhagen, Denmark has developed a Version 3 API toolkit to aid in the implementation of the ISO SR and Z39.50-1995 protocols. They say: "software is available free of charge, on a liberal license: Commercial re-use is explicitly permitted."
- National Library of Canada has made its client and server code available;
- NIST is making available a Z39.50 client/server package based on the PRISE search engines.
- OCLC has made its Z39.50 Client API available to the public
- University of California - Berkeley demonstration client/server protocol engine
- USGS is making available a freeware implementation of Z39.50 as an OLE add-on to WWW browsers. You can fetch the executable software, README.TXT, and source files by anonymous FTP to host www.usgs.gov, in the directory /gils/ciir/dtic_a02.
- Willow -- the Washington Information Looker-upper Layered Over Windows.

- John Lamp is doing a good job tracking sites with Z39.50 tools and resources.

Electronic documents of interest (let me know of more) are:



CNIDR Isite

CNIDR Isite is an integrated Internet publishing software package including a text indexer, a search engine and Z39.50 communication tools to access databases. Isite includes the CNIDR ZDist, Isearch and Search API distributions.

See what Isite can do for you

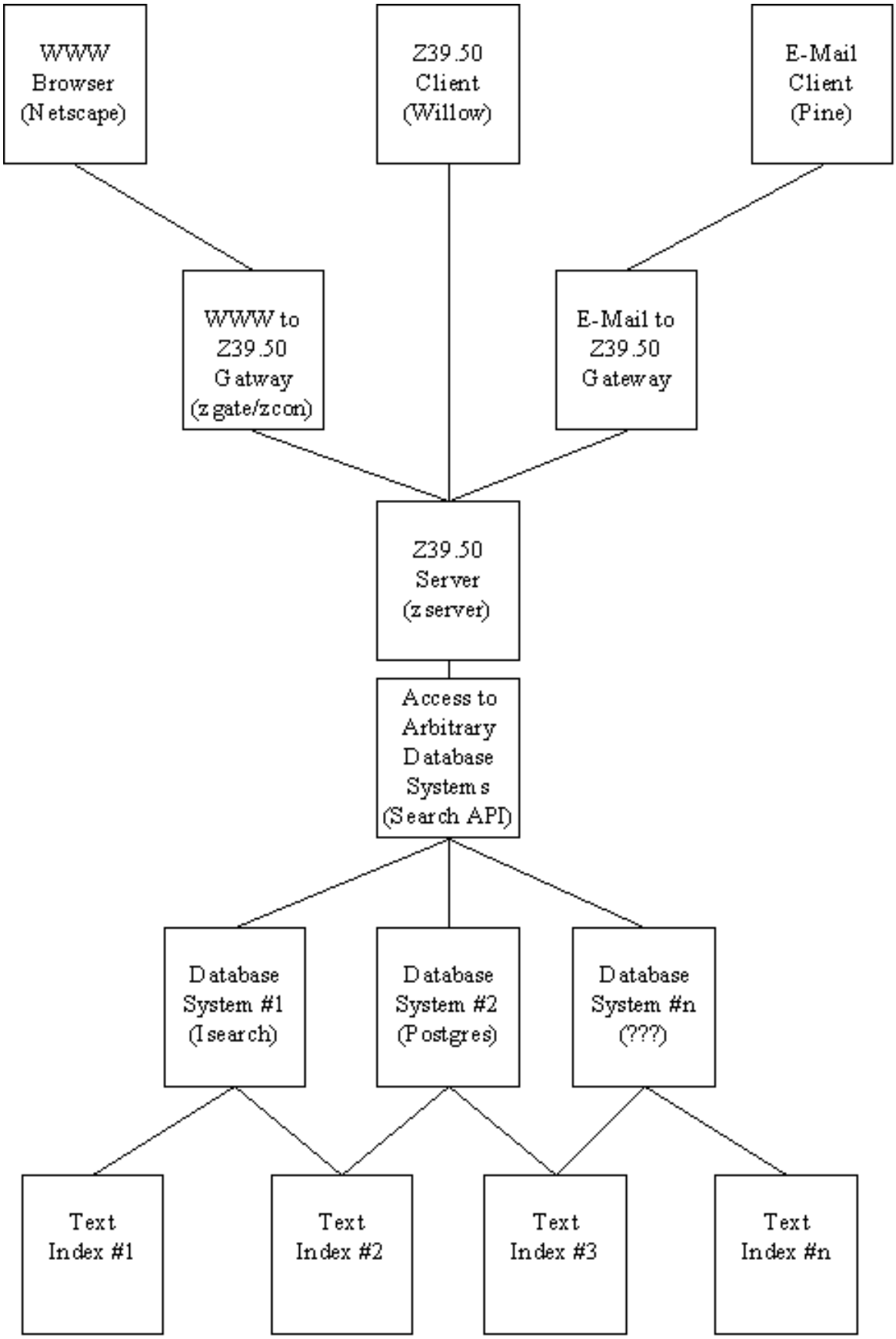
- **Help us to better serve you!**
- Diagram of Overall Architecture - Details available via Administrator's Guide below
- Demo of Stateful http to Z39.50 Gateway - Demonstrates access to various database systems
- Other systems using Isite
    - NASA Global Change Master Directory
        - Z39.50 Ranked Search
        - Z39.50 Boolean Search
        - Distributed Document Search
    - American Astronomical Society: Electronic Astrophysical Journal Letters
    - United Nations International Drug Control Programme
    - University of Tennessee Office of Research Services: Friends and Partners Cookbook
    - Microlytics, Inc.
    - Library of Congress Z39.50 Gateway
    - U.S. Department of Housing and Urban Development GILS Service
    - **YOUR LINK GOES HERE** - *Please* send me pointers to your Isite-based systems!!

Download a copy

- Stable Version - includes precompiled binaries
- Untested Versions - require a C++ compiler

Read the documentation

- Isite Administrator's Guide - Refers to stable versions
- Untested Isite Administrator's Guide - Refers to untested versions
- Isearch Tutorial - Step-by-step guide on building databases with Isearch
- Z39.50 Maintenance Agency - Everything you always wanted to know about Z39.50 and more!
    - Includes electronic copies of the ANSI/NISO Z39.50 standard
    - Includes implementor agreements
    - Includes various papers written by experts in the field
    - Includes lots of other stuff you will need to get the most out of Isite
- BSn Doctypes - Many of the input files supported by the Isearch indexer are documented here


WWW Browser (Netscape)
Z39.50 Client (Willow)
E-Mail Client (Pine)
WWW to Z39.50 Gatway (zgate/zcon)
E-Mail to Z39.50 Gateway
Z39.50 Server (zserver)
Access to Arbitrary Database Systems (Search API)
Database System #1 (Isearch)
Database System #2 (Postgres)
Database System #n (???)
Text Index #1
Text Index #2
Text Index #3
Text Index #n

Corporation for National Research Initiatives

Key Architectural Issues in The Digital Library

William Y. Arms

Acknowledgments

- This is work in progress.
- This is a personal interpretation of ideas developed by the CSTR Project.
- CSTR is a joint project of CNRI with Carnegie Mellon, Cornell, MIT, Stanford and UC Berkeley, funded by ARPA.
- For background information, see the CSTR home page.
- The architecture is more fully described in a paper by Robert Kahn and Robert Wilensky.

Key Issues and CSTR Terminology

This set of WWW pages looks at the following six key issues in the architecture of the digital library.

- Items in the library - digital object.
- Identifiers - handle.
- Storage - repository.
- Sets of objects - composite and meta-object.
- Information about objects - properties.
- Semantic layering (schema) - data model.

The architecture under development is an open architecture. In general, it allows these topics to be considered separately.

The CSTR Architecture and the World Wide Web

Many of the concepts in the CSTR architecture can be partially implemented within the framework of the World Wide Web and fit with recent IETF discussions.

Corporation for National Research Initiatives

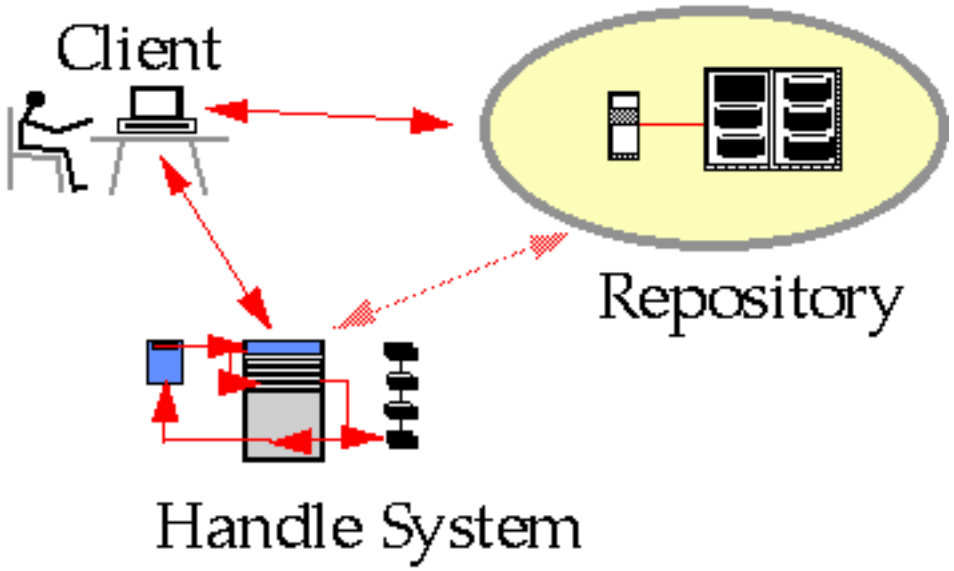
Digital Object Architecture Project

Principal investigators

Robert E. Kahn
William Y. Arms

Summary of the project

This project continues the architectural work of the DARPA-funded Computer Science Technical Reports (CS-TR) project. That project developed a Framework for Distributed Digital Object Services and implemented some key components. This project continues research and development of this framework and two extensive testbeds at the Library of Congress.



The basic entity in the system architecture is the "digital object", which contains copyright material or other material in which other rights and interests are manifest. There may also be rights and interests associated with digital objects themselves. The major components of the system are: (a) repositories of digital objects that allow network based deposit and access, (b) handle servers that record the location of digital objects over long periods of time, (c) registration and recordation mechanisms to keep track of rights and interests associated with digital objects, and (d) client software to enable use of these components over the network. Digital object

fingerprints are used in the registration system to permit validation of the objects at a later time.

The first testbed is with the Copyright Office at the Library of Congress. This is a system to register electronic materials for copyright and recordation of changes in copyright ownership. The second testbed is with the National Digital Library Program at the Library of Congress. This is a very large scale project to convert historic materials from the library's collections to digital form and make them available to the world.

Background papers

- A Framework for Distributed Digital Object Services by Robert Kahn and Robert Wilensky, May 1995
- Key Concepts in the Architecture of the Digital Library by William Y. Arms, D-Lib Magazine, July 1995
- "Implementation Issues in an Open Architecture Framework for Digital Object Services" by Carl Lagoze and David Ely. Cornell Computer Science Technical Report TR95-1540
- "A Design for Inter-Operable Secure Object Stores (ISOS)" by Carl Lagoze, Robert McGrath, Ed Overly, Nancy Yeager. Cornell Computer Science Technical Report TR95-1558
- Uniform Resource Names: A Progress Report The URN Implementors, D-Lib Magazine, February 1996
- Historical Collections for the National Digital Library: Lessons and Challenges at the Library of Congress Caroline R. Arms, D-Lib Magazine, April 1996. Part 2

Funding

Funding for this work is provided by the Defense Advanced Research Projects Agency (DARPA) and the Library of Congress.

Home

---

wya
6/30/96

UMBC **AgentWeb**

UMBC

An Honors University in Maryland

***L*aboratory for *A*dvanced *I*nformation *T*echnology**

UMBC AgentWeb

Intelligent Software Agents

Information and resources about intelligent information agents, intentional agents, software agents, softbots, knowbots, infobots, etc. Send comments and suggestions to Tim Finin (finin@umbc.edu).

- **About the AgentWeb...**
  - What's new... NEW
  - Current AgentNews webletter NEW
  - About the AgentNews webletter and mailing lists
  - AgentWeb help...
  - AgentWeb salon ... NEW
  - About the UMBC Laboratory for Advanced Information Technology
- **Agent basics ...**
  - Introductory material
  - Agent FAQ
  - Agent theory - philosophy, formalisms, ...
  - Agent technology - systems, tools, languages, standards, ...
  - Mobile agents ... NEW
- **Agent resources ...**
  - Agent papers
  - Agent events, conferences, workshops, ...
  - Agent mailing lists and newsgroups
  - Agent courses and seminars
  - Other agent related web resources
- **Who is doing what ...**
  - Agent-related R&D groups and companies
  - Agent-related projects
  - Example Agents

A PURL is a **P**ersistent **U**niform **R**esource **L**ocator. Functionally, a PURL is a URL. However, instead of pointing directly to the location of an Internet resource, a PURL points to an intermediate resolution service. The PURL resolution service associates the PURL with the actual URL and returns that URL to the client. The client can then complete the URL transaction in the normal fashion. In Web parlance, this is a standard HTTP *redirect.*

The OCLC PURL Service has been strongly influenced by the active participation of OCLC's Office of Research in the IETF Uniform Resource Identifier working groups. There is nothing incompatible between PURLs and the ongoing URN work. PURLs satisfy many of the requirements of URNs using currently deployed technologies and can be transitioned smoothly into a URN architecture once it is deployed.

Further Information and Resources

- A brief introduction to PURLs
- A longer introduction to PURLs
- Frequently Asked Questions
- Download the PURL software NEW
- PURL-L mailing list
- More info

Interacting with This Resolver

- Create your first PURL
- Register as a user
- Create PURLs, domains, groups
- Modify PURLs, domains, groups, users
- Search this resolver
- Power user's page (all features)

As of *Sat Jul 13 13:26:28 PDT 1996* : PURLs Created = **6768** , PURLs Resolved = **473905** and Unique Client Systems = **13121** (See the complete Database Stats for more details.)

The PURL Team
purl@oclc.org

Corporation for National Research Initiatives

Handles and the Handle System

Forms for handle administration

Forms to add and edit handles, to create naming authorities, and to set up groups of administrators are available through the Handle Administration Page.

Information about the Handle System

Technical information

- An overview of the system.
- Implementation of the Handle Management System.
- FTP server to download documentation and code.
- Browsers that support handles.

Architectural considerations

- The use of handles within a a framework for distributed digital object services.
- Handles as a key concept in the digital library.
- The IETF's work on Universal Reference Names.

Presentations and demonstrations

- An architectural overview.
- The handle system.
- D-Lib Magazine with handles.

A brief introduction to Handles

A **handle** is a unique identifier for a digital object. This object can be stored in a digital library repository, in an ftp archive, in a World Wide Web server, or any other digital store. Handles can also be used for other forms of identification, such as electronic mail addresses. A high performance Handle Management System is publicly available on the Internet. The useful properties of handles include the following.

- Handles are guaranteed to be unique.
- Handles are permanent. Therefore, they can be used to identify objects for purposes of copyright or archiving.
- Handles are location independent. The object may be moved to a different location without changing its handle. This enables handles to be used to refer to an object, for example, in a bibliographic citation.

A handle has the syntax:

naming authority / string
or: hdl://*naming authority / string*

The **naming authority** is a globally unique name. The **string** is unique for that naming authority.

Return to CNRI home page

---

hdl://cnri/handle-intro

wya
Last revised: November 11, 95

Digital Libraries - Implementation Principles

As we build digital libraries, it is important to consider key principles so that these libraries will be easily usable, and have long-term archival value.

1. Declarative representations of documents should be used.
2. Document components should be represented using natural forms, namely objects that can be manipulated by users familiar with those objects.
3. Links should be recorded, preserved, organized and generalized.
4. There should be a separation between the digital library and user interfaces to it.
5. Searching should make use of advanced retrieval methods.
6. Open systems that include the user, and where (some of) the functions of librarians are carried out by the computer, must be developed.
7. Task-oriented access to electronic archives must be supported.
8. A user-centered development approach should be adopted.
9. Users should work with objects at the right level of generality.

Preserving Digital Information: Final Report and Recommendations

May 20, 1996

At the end of 1994 the Commission on Preservation and Access (CPA) and RLG created a Task Force on Archiving of Digital Information charged with investigating and recommending means to ensure "continued access indefinitely into the future of records stored in digital electronic form." The 21-member task force, co-chaired with distinction by Donald Waters, Associate University Librarian, Yale University, and John Garrett, Chief Executive Officer of CyberVillages Corporation, recently completed their final report. RLG and CPA are making this widely available online and in print.

Electronic versions are available from RLG's FTP server (ftp.rlg.org) and this Web site:

HTML version
Adobe Acrobat version: /pub/archtf/final-report.pdf
Microsoft Word for Windows 6.0 version: /pub/archtf/final-report.doc
ASCII Rich Text Format version: /pub/archtf/final-report.rtf

Notes:

To download an Adobe® Acrobat® viewer to use as a helper application with your web browser, connect to the Adobe web site.

Copies of the printed, bound report are available for $15.00 (prepayment required) from the Commission on Preservation and Access, 1400 16th Street, N.W., Suite 740, Washington, DC 20036-2217.

RLG will be mailing the printed report to the member representative at each of our member institutions in North America and Europe as well as to each member liaison in our collaborative SHARES (Shared Resources) and PRESERV (Preservation) programs.

The task force's final report benefits from their action last September to make a draft version available online and to open a listserv for comments by the community. Many thanks to all of you who responded. That draft report can still be found on RLG's server and Web site:

Adobe Acrobat version: /pub/ArchTF/Draft-Report.pdf
Microsoft Word for Windows 6.0 version: /pub/ArchTF/Draft-Report.doc
ASCII version: /pub/ArchTF/Draft-Report.txt

RLG has already built into its agenda work on several of the task force's nine recommendations. (Our archival server and digital collections projects are directly related.) We will be following up on other recommendations with other stakeholders.

Please share your comments and advice with us regarding this report and the specific recommendations; you can send them by e-mail to Nancy Elkington, RLG member services officer and member of the task force.

Sincerely,

James Michalko
President

TEI Guidelines for Electronic Text Encoding and Interchange (P3)

Made available from the Electronic Text Center at the University of Virginia.

Search the *TEI Guidelines*.

Word or phrase (omit all quotes): Search

1-50

Other types of searches:
You may also combine words or phrases within a specified proximity, or locate segments such as sections where two words or phrases both occur.

Browse the *TEI Guidelines*.

Resources of Related Interest

- The Text Encoding Initiative Home Page
- Other Electronic Versions of the TEI Guidelines
- TEI P3 now available on CD-ROM
- The Electronic Text Center Introduction to TEI and Guide to Document Preparation.
- TEI DTD Browser, courtesy of CETH