ACM SIGCSE '97 Workshop
Digital Libraries for CS Education

INSTRUCTORS: Edward A. Fox, John Urquhart
SCHEDULE: Saturday, Mar. 1, 1997, 1:30 p.m. - 4:30 p.m.
LOCATION: Fairmont Hotel, San Jose, CA, USA
PARTICIPANTS: Participants are encouraged from those working in fields like computing, library science, information science, multimedia, and publishing. Those with little background on digital libraries are invited.

PART I.
URL: http://ei.cs.vt.edu/~fox/FoxMM96tutA.pdf is identical except for 1st page
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. They should become familiar with the underlying technologies for digital libraries, such as information retrieval, hypertext and electronic publishing.
Planned Activities: There will be coverage of the following PROJECTS: NETLIB (numerical analysis); CORE (chemistry); TULIP (material science & engineering, with Elsevier, OCLC); IBM digital libraries products and projects; Hyper-G/HyperWave (clients and servers); BEV HistoryBase; CS technical reports (CS-TR, WATERS, NCSTRL) and related efforts; CS education (ACM literature, courseware on IR, multimedia, hypertext, history); Digital Library Initiative (CMU, Michigan, Stanford, UC Santa Barbara, Illinois); ETD (electronic theses and dissertations).

PART II.
URL: http://ei.cs.vt.edu/~fox/FoxMM96tutB.pdf is identical except for 1st page
Objectives: Attendees should be able to help in design, development, evaluation, and standardization efforts related to digital libraries. They should understand research and development issues, including principles and guidelines for design of scalable, sustainable DLs.
Planned Activities: There will be coverage of the following SOURCES: Digital Library conferences; IITA meetings (e.g., May 1995 workshop); Allerton Institutes (from U. Illinois, NSF); D-Lib (research, magazine, working groups); D-Lib research articles (architecture, metadata, URNs, use); Virginia Tech information (DL page, Sourcebook); Virginia Tech projects (Envision, ILDLCS, WWW traffic analysis/mod/sim); Z39.50 (overview, OCLC, CNIDR); Library of Congress; CNRI (architecture, handles); UMBC agents, preservation, TEI, ...

PART III. Discussion
A general discussion of the use of digital libraries in CS education will involve all participants.

Introduction

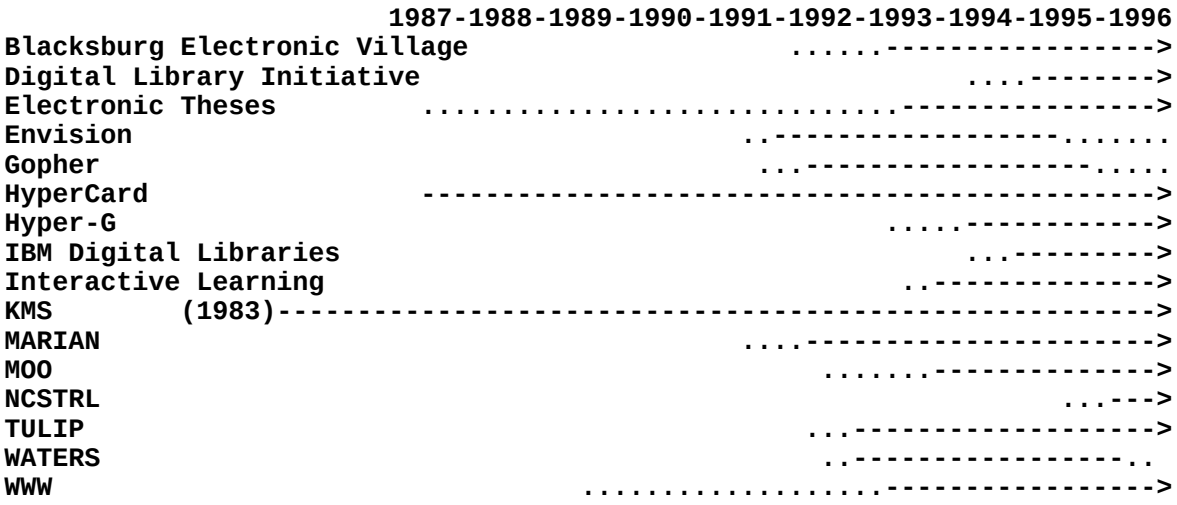


Figure 1: Timeline of Recent Information and Digital Library Systems

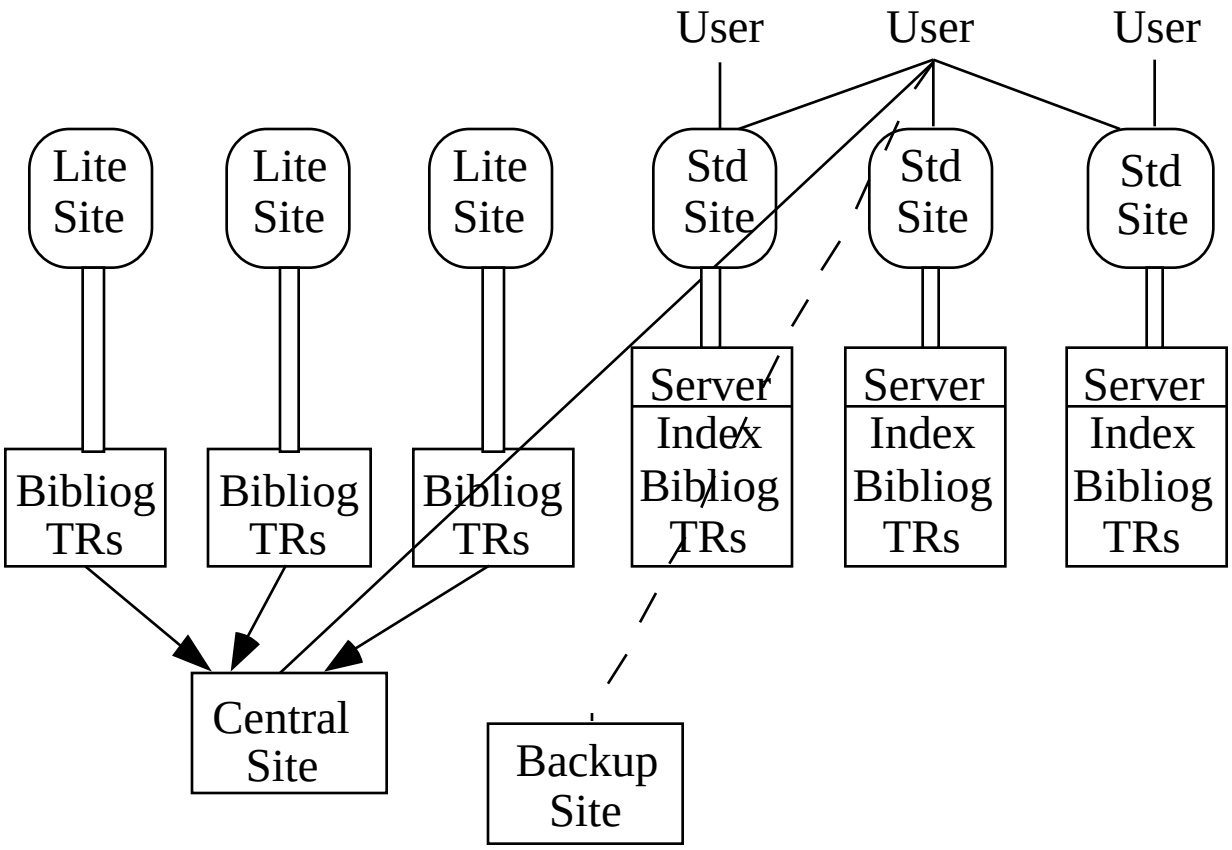


Figure 2: NCSTRL Architecture

TABLE OF CONTENTS

Fig. 1: Timeline of Recent Information & DL Systems [link]
Fig. 2: NCSTRL Architecture [link]
SELECTED PAGES FROM WWW ---
I. SOURCES:
* Conferences
- ACM SIGIR DL Page http://info.sigir.acm.org/sigir/digital_lib.html
- DL'97: Philadelphia, July 23-26 (before ACM SIGIR'97)
- DL'96: Bethesda, March (1st ACM ...)
- DL'95: Austin, June [link] http://csdl.tamu.edu/DL95/
- earlier in DL'9x and ADL'9x series from 1994-1996
- IITA Digital Libraries Workshop, 1995 [link]
  http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html
- Allerton, 1995 [link] http://edfu.lis.uiuc.edu/allerton/95/
* Journals
- J. of Visual Communication and Image Representation, 1996
- IEEE Computer, May 1996
- Communications of the ACM, Apr. 1995
- J. American Society for Information Science, Sept. 1993
- (new) British Computer Society J. of Digital Information
* WWW
- D-Lib http://www.dlib.org/
  + Research Projects (incl. DLI) ([link]) http://WWW.dlib.org/projects.html
- D-Lib Articles ([link]) http://www.dlib.org/title-index.html ([link])
- D-Lib Working Groups [link] http://www.dlib.org/groups.html
  + Metadata [link] http://www.dlib.org/metadata/overview.html
  + Naming [link] http://www.dlib.org/naming/overview.html
  + Repository Interfaces [link]
    http://www.dlib.org/repository/overview.html
  + Social Aspects [link] http://www.dlib.org/social/overview.html
- D-Lib Magazine
  + Agents [link] http://www.dlib.org/dlib/July95/07birmingham.html
  + Architecture (incl. handles) [link]
    http://www.cnri.reston.va.us/home/dlib/July95/07arms.html
  + Metadata [link] http://www.dlib.org/dlib/July95/07weibel.html
  + Uniform Resource Names (URNs) [link]
    http://www.dlib.org/dlib/february96/02arms.html
  + Use [link] http://www.dlib.org/dlib/october95/10bishop.html
- Virginia Tech ptrs [link] http://scholar.lib.vt.edu/digilib/
- Digital Library Source Book, 1993 [link] http://fox.cs.vt.edu/DLSB.html

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

Getting a physical copy of the DL 95 Proceedings

Sponsors and cooperating institutions

From the Conference Chair, David M. Levy

From the Program Chair, Richard Furuta

Conference Committee

Attendee List

Full Papers

Delivering Technology for Digital Libraries: Experiences as Vendors,
William T. Crocca and William L. Anderson

InterPay: Managing Multiple Payment Mechanisms in Digital Libraries,
Steve B. Cousins, Steven P. Ketchpel, Andreas Paepcke, Hector Garcia-Molina, Scott W. Hassan, and Martin Roescheisen

Providing Government Information on the Internet: Experiences with THOMAS,
W. Bruce Croft, Robert Cook, and Dean Wilder

A New Zealand Digital Library for Computer Science Research,
Ian H. Witten, Sally Jo Cunningham, Mahendra Vallabh, and Timothy C. Bell

Cataloging in the Digital Order,
David M. Levy

Collection Maintenance in the Digital Library,
Mark S. Ackerman and Roy T. Fielding

The Digital Research Library: Tasks and Commitments,
Peter S. Graham

Management of the Nationale HPCC Software Exchange--A Virtual Distributed Digital Library,
Shirley Browne, Jack Dongarra, Ken Kennedy, and Tom Rowan

Digital Library Research at Loughborough: The Last Fifteen Years,
Cliff McKnight

User Needs Assessment and Evaluation for the UC Berkeley Electronic Environmental Library Project,
Nancy A. Van House

A Hypertextual Interface for a Searcher's Thesaurus,
Eric H. Johnson and Pauline A. Cochrane

Automatic Extraction of Hypermedia Bundles from the Digital Library,
Hugh Davis and Jessie Hey

Automatic Creation and Maintenance of an Organizational Spatial Metadata and Document Digital Library,
Charles Kacmar, Dean Jue, David Stage, and Christie Koontz

Use of the ISite Z39.50 Software to Search and Retrieve Spatially-referenced Data,
Douglas D. Nebert and James Fullton

Enhancing Usability of Network-based Library Information System---Experimental Studies on User Interface for OPAC and of a Collaboration Tool for Library Services,
Shigeo Sugimoto, Seiki Gotou, Yanchun Zhao, Tetsuo Sakaguchi, and Koichi Tabata

Early Prototypes of the Repository for Patterned Injury Data,
Prasun Dewan, Kevin Jeffay, John Smith, David Stotts, and William Oliver

Digital Libraries and Sustainable Development?,
Amanda Spink

Using Online Information Resources: Reaching for the *.*'s,
Roberta Lamb

Digital Libraries: Issues and Architectures,
Peter J. Nuernberg, Richard Furuta, John J. Leggett, Catherine C. Marshall, and Frank M. Shipman III

SCAM: A Copy Detection Mechanism for Digital Documents,
Narayanan Shivakumar and Hector Garcia-Molina

Penstation: Easy Access to Relevant Facts without Retrieving,
Kazunori Muraki and Kenji Satoh

Short Papers

On-the-fly Hyperlink Creation for Page Images,
Eytan Adar and Jeremy Hylton

Four Lessons Learned from Managing World Wide Web Digital Libraries,
Robert Pettengill and Guillermo Arango

Public Access to EPA Superfund Records - A Digital Alternative,

Verne E. McFarland and Steven Wyman

Automating the Structural Markup Process in the Conversion of Print Documents to Electronic Texts,
Casey Palowitch and Darin Stewart

The Knowledge Manager as a Digital Librarian: An Overview of the Knowledge Management Pilot Program at the MITRE Corporation,
Kathleen M. Flynn

Querying, Navigating, and Visualizing a Digital Library Catalog,
Aravindan Veerasamy and Shamkant Navathe

Service Models, Operational Decisions and Architecture of Digital Libraries,
Yuehong Yuan, Stephen Roehrig, and Marvin Sirbu

Author Index

Keyword Index

Frank Shipman / shipman@cs.tamu.edu / 95-06-08

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
Definitions and Roles of Digital Libraries
Defining Interoperability in the Digital Library Environment
Infrastructure Requirements for Digital Library Research
Research Issues and Priorities
1. Interoperability
2. Description of Objects and Repositories
3. Collection Management and Organization
4. User Interfaces and Human-Computer Interaction
Conclusions

Executive Summary
Appendix 1 - List of Perticipants
Appendix 2 - Strawman Report
Appendix 3 - Report of the working groups
- 3-1 - The Publishing Perspective
- 3-2 - The Commercial Perspective
- 3-3 - The Library Perspective
- 3-4 - The Internet Perspective
- 3-5 - The Multimedia Perspective

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

activity. Informed by insights gained from current research, this workshop offered an opportunity to consider questions such as interoperability objectives that might be defined among projects now underway.

The workshop was organized by Hector Garcia-Molina of Stanford University and Clifford Lynch of the University of California Office of the President. The IITA working group, which sponsored the meeting, reports to the National Science and Technology Council (NSTC) through the High Performance Computing, Communications, and Information Technology subcommittee of the Committee on Information and Communication. The workshop was attended by some 60 leading digital library researchers and developers and by representatives from a wide range of federal government organizations concerned with research and development and policy formulation related to digital libraries (see Appendix 1 for a roster of attendees).

Workshop attendees were asked to consider the following questions as a point of departure in developing the research agenda:

1. What is a Digital Library? How does it differ from an information repository or from today's World Wide Web? How many Digital Libraries will there be, and how will they interlink? How might this look to users?

2. What Digital Library infrastructure is needed? What does "infrastructure" consist of in this context and how does it differ from the broader applications support infrastructure for the emerging NII? What is the relationship between infrastructure and standards? Who will use this infrastructure? When must it be defined, and what parts are most urgently needed? How does the infrastructure relate to intellectual property management and publisher concerns?

3. How can a Digital Library be evaluated? How will we know in three to four years if current research projects have been successful in developing effective digital library services for their user communities?

To further frame and stimulate discussion, Hector Garcia-Molina prepared a position paper discussing the issues and distributed it prior to the workshop (see Appendix 2).

Participants spent the majority of the workshop in one of five groups; unlike many workshops, in which each group is assigned a different set of issues, here each group approached the full spectrum of questions from a specific, unique viewpoint and generated a summary of their discussions that reflected that viewpoint. After a presentation from the five group leaders representing each group's approach to the issues, each participant selected his or her group. The five groups and their leaders were

Bill Arms,
Corporation for National Research Initiatives:
The Publishing Perspective

Michael Lesk,
Bellcore:
The Commercial Perspective

Bruce Schatz,
University of Illinois Urbana Champaign:
The Library Perspective

Mike Schwartz,
University of Colorado:
The Internet Perspective

Terry Smith,
University of California, Santa Barbara:
The Multimedia Perspective

The reports of these five groups appear in Appendix 3. This summary of the workshop extracts common themes and also key points of disagreement from the work of the five groups and places them in broader context. The report is not a consensus document; while it draws heavily on the five group reports and has also benefited greatly from comments from attendees, it does not attempt to reflect completely any of the five group reports.

This report addresses responses to the first two questions posed to the attendees (the definition of a digital library and infrastructure needs to support digital librariesand discusses the research agenda. The third question posed to the attendees -- how to evaluate Digital Library projects - -- did not receive much attention from most of the groups; it is to be the subject of a separate workshop on User Evaluation Methods to be held October 29-31 at the Allerton Center under the auspices of The University of Illinois Urbana-Champaign and NSF. Some groups did identify the need for consistent instrumentation and data gathering across projects to facilitate evaluation. In addition, several groups stressed the need to make the transition from a systems technology framework to one driven by user access and collection organization in developing future digital library technology and systems. This view is perhaps most eloquently stated in the reports of the Internet working group and the Library working group.

Definitions and Roles of Digital Libraries

Considerable work has already been done on operational definitions of Digital Libraries and their relationship to traditional library institutions, as well as to the broader systems of scholarly and commercial publishing (see, for example, Communications of the ACM, April 1995). Much of the discussion in this workshop was motivated by questions of scaling, interoperability and needed support infrastructure.

Digital libraries were viewed as systems providing a community of users with coherent access to a large, organized repository of information and knowledge. One group made the provocative proposal that this organization of information was characterized by the absence of prior detailed knowledge of the uses of the information. The ability of the user to access, reorganize, and utilize this repository is enriched by the capabilities of digital technology; the Multimedia group provided particularly vivid examples of these possibilities.

Several groups pointed out that, in fact, digital libraries would, for the foreseeable future need to span both print and digital materials and that the central issue was to provide a coherent view of a very large collection of information. In this sense, an emphasis on content solely in digital format is too limiting. Really, the objective is to develop information systems providing access to a coherent collection of material, more and more of which will be in digital format as time goes on, and to fully exploit the opportunities that are offered by the materials that are in digital formats. Additionally, the comprehensiveness and value of the collection accessible through a digital library system can be strengthened by the ability to integrate materials in digital formats that have not been well-represented, easy to access, or effectively usable in traditional library collections, such as multimedia, geospatial data, or numerical datasets. There is, in reality, a very strong continuity between traditional library roles and missions and the objectives of digital library systems.

Participants in the workshop repeatedly underscored this continuity, and emphasized that the traditional library institutional missions of collection development, collection organization, access, and preservation must extend to the digital library environment. Digital libraries will be a component in the broader range of future library services, and librarians will play a central role in developing and managing digital libraries.

While there would be many digital repositories, a given digital library system should provide a coherent, consistent view of as many of these repositories as possible. From the user's perspective, to the extent possible, there should appear to be a single digital library system. Users increasingly have access to various types of digital collections and information systems: personal information resources, workgroup and organizational information collections and collaboration environments, and more public digital libraries. Defining the boundaries and characteristics of these information spaces and exploring ways in which they can be fused into a coherent whole is a central problem that cuts across all aspects of the research agenda. From the

user's perspective, the digital library system needs to extend smoothly from personal information resources, workgroup and organizational systems, and out to personal views of the content of more public digital libraries.

Some groups raised, but did not resolve, the question of the extent to which the digital library system should incorporate support for publishing, annotation, and integration of new information, and the extent to which additions to repositories within the digital library system should be mediated by librarians. It is clear that the development of digital libraries is closely linked to the changes that are occurring in modes of scientific and scholarly communication; the extent to which the digital library should actively embrace -- and perhaps even drive -- these changes remains to be fully explored.

Libraries -- digital or traditional -- exist to serve diverse purposes and constituencies. To some extent, each discipline, constituency, and collection creates its own organization of information. In the digital library world this differentiation among library collections, organization, and services may become more visible. One of the key challenges is to retain this diversity, which is responsive to unique constituencies, and at the same time permit information to be effectively shared across disciplines and constituencies. This is an essential component of the interoperability questions that formed a major focus for the workshop. Workshop participants represented many of these diverse perspectives: university research libraries, archives, libraries supporting teaching, public libraries, and libraries of the performing arts.

Defining Interoperability in the Digital Library Environment

Defining interoperability proved difficult. It is clear that this is still a central research problem in its own right, and one that merits continued attention. Discussions of infrastructure focused on common tools, enabling technologies and standards that would provide a basis for further exploration of interoperability issues, particularly by encouraging and facilitating the growth of digital libraries on the Internet. Considerable effort was spent on identifying infrastructure that was either unique or particularly critical to progress in digital libraries, as opposed to more general-purpose infrastructure that a range of NII applications, including digital libraries, might share. One clear theme was that an understanding of interoperability issues required operational experience which could only be gained by large-scale deployment of digital library systems. Speculation about interoperability in the abstract is of very limited value.

Participants expressed a full spectrum of views on interoperability. At one end of the spectrum is the use of common tools and interfaces that provide a superficial uniformity for navigation and access but rely almost entirely on human intelligence to provide any coherence of content. At the opposite end of the spectrum is deep semantic interoperability. The precise definition of deep semantic interoperability was the subject of some debate, but deals with the ability of a user to access, consistently and coherently, similar (though autonomously defined and managed) classes of digital objects and services, distributed across heterogeneous repositories, with federating or mediating software compensating for site-by-site variations. It also extends beyond passive digital objects to actual services offered by specific digital library systems. Deep semantic interoperability is a "grand challenge" research problem; it is extraordinarily difficult, but of transcendent importance, if digital libraries are to live up to their long-term potential. An intermediate position between these two extremes advocates primarily syntactic interoperability (the interchange of metadata and the use of digital object transmission protocols and formats based on this metadata rather than simply common navigation, query, and viewing interfaces) as a means of providing limited coherence of content, supplemented by human interpretation.

Note that the term "digital object" here is intended only to describe, in the broadest sense, the type of information objects that may comprise a digital library -- textual, audio, video, numeric, computer programs, or multimedia composites of such components. It is not intended either to endorse or preclude an object-oriented architectural framework for digital library systems (in the sense of object-oriented programming or object-oriented databases, for example).

Infrastructure Requirements for Digital Library Research

The most urgent infrastructure need is to establish common schemes for the naming of digital objects, and the linking of these schemes to protocols for object transmission, metadata, and object type classifications. The consensus of the groups was that naming schemes for digital objects that allow global unique reference represented perhaps the most immediate infrastructure deployment priority in order to facilitate resource sharing, linkages, and interoperation among digital library systems and to facilitate scale-up of digital library prototypes. It was recognized that the design of large-scale naming systems and their integration into the larger digital library framework will continue to be an important research area, but that infrastructure support needs to be put in place quickly for at least an interim system, and that in fact experience with such an interim system would inform further research.

The deployment of a public key cryptosystem infrastructure -- including the development of a system of key servers and the definition of standards and protocols -- was also identified as essential to progress in digital libraries; this is necessary to support digital library needs in areas such as security and authentication, privacy, rights management, and payments for the use of intellectual property. While the need for public key cryptosystem infrastructure is hardly unique to digital libraries, the importance of the digital library services and components which depend on this infrastructure mean that its absence represents a significant barrier. In particular, until these problems are addressed, it seems unlikely that we will see commercial publishers and other information suppliers making large amounts of high-value copyrighted information broadly available to digital library users. This in turn will constrain the development of research prototypes and may be a distorting factor in studies of user behavior.

Research Issues and Priorities

The working groups outlined a wide range of important research issues; most groups were less successful at prioritizing them, beyond the immediate infrastructure needs already discussed. The five key research areas that emerged from the workshop are described below; arguably, the first three are of most central and immediate importance, specifically to the development of digital libraries, though the long-term importance of research in the fifth area (economic, social, and legal issues) cannot be overemphasized. The distinctions among the five areas are to some extent arbitrary; for example, progress on interoperability (the first area) depends critically on progress in our ability to describe successfully objects and repositories (the second area).

1. Interoperability

The difficulty in defining the objectives for interoperability have already been discussed; clarifying these objectives, mapping the spectrum of interoperability, and establishing the key challenges at points along this specturm are key research issues in their own right.

The more technical interoperability research involve protocol design that supports a broad range of interaction types, inter-repository protocols, distributed search protocols and technologies (including the ability to search across heterogeneous databases with some level of semantic consistency), and object interchange protocols. Interoperability is not simply a matter of providing coherence among passive object repositories. Digital library systems offer a range of services, and these services must be projected in an interoperable fashion as well. One particular issue that emerged was that existing Internet protocols (such as HTTP, the basis of the World Wide Web) are clearly inadequate. Research must move beyond the current base of deployed protocols and systems. This raises complex questions about how to deploy prototype systems and the tradeoffs between advanced capabilities and ubiquity of access.

The practical question of the nature of the installed technology base and the need to support this installed base will increasingly frame and influence interoperability research. Access to digital libraries is not an end in itself for most users, but rather a support service; many will be willing to sacrifice advanced functionality for consistency, stability, and ability to use familiar, common access tools. Just as the installed base has become the greatest barrier to meaningful large-scale trials of new approaches that improve existing services (as opposed to providing entirely new services which do not compete with an installed base) in the overall Internet environment, user expectations and the installed base will ultimately impede progress in fundamental technology research within the large-scale experiments necessary to gain insights into interoperability among

digital libraries. Managing this tension will be a critical element in the continued development of the community's research agenda.

It should be noted that, at this relatively early stage in the evolution of digital library technology, it is of vital importance that projects strive for approaches that incorporate high functionality and extensibility. A high level of functionality in the standards and protocols used, even if not fully exploited initially, will postpone the time when the inertia of the installed base begins to confine research opportunities. Careful design of extensibility in digital library systems will facilitate continued research progress and understanding of the impact of new approaches on the user community without the need to attempt to displace an installed base.

2. Description of Objects and Repositories

In order to provide a coherent view of collections of digital objects, they must be described in a consistent fashion which can facilitate the use of mechanisms such as protocols that support distributed search and retrieval from disparate sources. Research in description of objects and collections of objects provides the foundation for effective interoperability. Interoperability at the level of deep semantics will require breakthroughs in description as well as retrieval, object interchange, and object retrieval protocols.

Issues here include the definition and use of metadata and its capture or computation from objects, the use of computed descriptions of objects, federation and integration of heterogeneous repositories with disparate semantics, clustering and automatic hierarchical organization of information, and algorithms for automatic rating, ranking, and evaluation of information quality, genre, and other properties. Other key issues involved knowledge representation and interchange, and the definition and interchange of ontologies for information context. The idea of active "information matchmaking" emerged in several group reports.

Research is also needed to understand the strengths and limitations of purely computer-based technologies for describing objects and repositories, and the appropriate roles for the efforts of human librarians and subject experts in the digital library context as a complement to these technology-based approaches.

3. Collection Management and Organization

Collection management and organization research is the area where traditional library missions and practices are reinterpreted for the digital library environment. Progress in this area is essential if digital library collections are to meet successfully the needs of their user communities.

Policies and methods for incorporating information resources on the network into managed collections, rights management, payment, and control issues were all identified as central problems in the management of digital collections. Approaches to replication and caching of information and their relationship to collection management in a distributed environment need careful examination. The authority and quality of content in digital libraries is of central concern to the user community; ensuring and identifying these attributes of content calls for research that spans both technical and organizational issues. Research is also needed to clarify the roles of librarians and institutions in defining and managing collections in the networked environment.

With the enhanced potential to support nontextual content effectively in the digital library environment, issues in nontextual and multimedia information capture, organization, and storage, indexing and retrieval are clearly key research areas. However, textual digital documents remain a vitally important research area in their own right, and are far from fully understood. The role of knowledge bases in digital libraries remains a poorly explored but potentially important question.

The preservation of digital content for long periods of time, across multiple generations of hardware and software technologies and standards is essential in the creation of effective digital libraries. This is an extraordinarily difficult research problem which has not received sufficient attention.

4. User Interfaces and Human-Computer Interaction

While user interfaces and human-computer interaction issues are an extensive field of research in their own right, there are some specific problems that are central to progress in digital libraries.

Display of information, visualization and navigation of large information collections, and linkages to information manipulation/analysis tools were identified as key areas for research. The use of more sophisticated models of user behavior and needs in long-term interactions with digital library systems is a potentially fruitful area for research. The necessity for a more comprehensive understanding of user needs, objectives, and behavior in employing digital library systems was stressed repeatedly as a basis for designing effective systems. Finally, it was observed that digital library systems must become far more effective in adapting to variations in the capabilities of user workstations and network connections (bandwidth) in presenting appropriate user interfaces; new technologies such as personal digital assistants and nomadic computing models will emphasize this need.

5. Economic, Social, and Legal Issues

Digital libraries are not simply technological constructs; they exist within a rich legal, social, and economic context, and will succeed only to the extent that they meet these broader needs. Rights management, economic models for the use of electronic information, and billing systems to support these economic models will be needed. User privacy needs to be carefully considered. There are complex policy issues related to collection development and management, and preservation and archiving. Existing library practice may shed some light on these questions. The social context of digital documents, including authorship, ownership, the act of publication, versions, authenticity, and integrity require a better understanding. Research in all of these areas will also be needed if digital libraries are to be successful.

Conclusions

This workshop has made substantial progress in refining and focusing a research agenda for digital libraries, as well as in developing insights into questions about interoperability among digital libraries and the infrastructure necessary to support such interoperability. Interoperability is likely to continue to be a useful organizing theme in refining this agenda in the coming years. The outcomes of the workshop also suggest that a focus on broad architectural issues in digital libraries will be fruitful. Several working groups commented on the need to develop component software strategies that would facilitate the transfer of technology among the current digital library pilot projects and from these projects to other new digital library research efforts. The Internet working group went further in suggesting that the development of a broadly available software base for the digital library community would contribute to rapid progress, and we believe that this suggestion deserves careful consideration.

Scaling was identified as a major area of concern. The common vision is one of tens of thousands of repositories of digital information that are autonomously managed yet integrated into what users view as a coherent digital library system. Accommodating this very large number of repositories -- a very different environment than that in which today's handful of pilot projects operate -- will clearly have major implications for infrastructure definition and design. We must move rapidly towards an infrastructure that can support and facilitate research towards this common vision. The full range of issues here are unclear. Some immediate needs are evident; these are reflected in the emphasis on establishing naming systems for digital objects as a high priority, for example.

We don't know how to approach scaling as a research question other than to build upon experience with the Internet. However, attention to scaling as a research theme is essential and may help in further clarifying infrastructure needs and priorities, as well as informing work in all areas of the research agenda outlined above. For example, reliability questions are poorly understood; in a sufficiently large system, some components will inevitably be out of service during the processing of any given query. The need to support large-scale deployment projects (in terms of size of user community, number of objects, and number of repositories) and to study subsequently the effectiveness and use of such systems was emphasized repeatedly. It is clear that limited deployment of prototype systems will not suffice if we are to understand understand the

research questions involved in digital libraries.

Research in scale-up is very difficult to perform except by building and deploying a large-scale digital library system. Establishing infrastructure and tools to facilitate experimentation with large-scale systems is essential, as is funding to study use and behavior of large-scale systems once deployed through this infrastructure. The Internet as a context for deploying digital library systems offers an unprecedented opportunity -- not only technically by providing connectivity to an enormous potential user base but also culturally, given the Internet community's models and traditions of technology diffusion through the distribution of publicly available prototype software -- to move ahead large-scale experiments. Research efforts should exploit these opportunities.

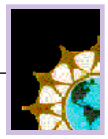
Finally, it seems clear that the inevitable presence of large amounts of commercially valuable, proprietary information in the future -- which can be viewed as another form of scale-up in digital libraries -- will also shape the research agenda in new ways. The near-term focus is on overcoming the infrastructural barriers to supporting proprietary information (such as authentication, billing, and rights management). There are research issues in the design of such an infrastructure, but also operational and policy problems impeding deployment. While some of the research issues are complex and will require ongoing exploration, putting at least the first steps towards the necessary infrastructure in place to accommodate such commercially valuable information is a high priority in advancing the research agenda and addressing scale-up issues. It will also stimulate commercial developments that will complement existing research initiatives. The development of an increasingly rich marketplace of information resources under a wide range of economic and legal constraints will create new opportunities in all areas of the research agenda presented above, and will allow us to explore vital new research questions in the development of description, navigation, access, and resource discovery technologies and systems that can function in this broader environment.

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



Introduction, Ann P. Bishop
Session 1 - Migrating Foundational Study Approaches to the Virtual Environment
Special Presentation: Findings from Digital Library Studies
Annelise Mark Petjersen, *Designing for Retrieval in Library Collections: Lessons from Book House*
Michael Twidale, *How to Study and Design for Collaborative Browsing in the Digital Library*
Session 2 - Co-Design in Digital Libraries
Session 3 - Work Practice and Institutional Change
Session 4 - Electronic Information Seeking and Use
Special Presenation: Social and Organization Issues in Classification (notes only) - S. Leigh Star and Geof Bowker
Session 5 - Users, Diversity, and Change
Session 6 - Wrap-up

List of Participants

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

Digital Library Research

Here are pointers to some of the major cooperative projects and associated activities in digital library research.

- Federally-funded cooperative projects
- Coordinating and funding bodies

Federally funded cooperative projects

The NSF/DARPA/NASA Digital Library Initiative (DLI). Six federal funded projects in digital library research, with partnerships led by universities. The individual projects are listed below.

University of California, Berkeley: An Electronic Environmental Library Project. (A DLI project.)

University of California, Santa Barbara: The Alexandria Project: Towards a Distributed Digital Library with Comprehensive Services for Images and Spatially Referenced Information. (A DLI project.)

Carnegie Mellon University: Informedia: Integrated Speech, Image and Language Understanding for Creation and Exploration of Digital Video Libraries. (A DLI project.)

University of Illinois at Urbana- Champaign: Building the Interspace: Digital Library Infrastructure for a University Engineering Community. (A DLI project.)

University of Michigan: The University of Michigan Digital Library Project. (A DLI project.)

Stanford University: Stanford University Digital Libraries Project. (A DLI project.)

The Computer Science Technical Reports Project (CSTR). A collaboration involving CNRI, five universities, and the Library of Congress.

D-Lib. A forum for researchers and developers of advanced digital libraries.

Coordinating and funding bodies

NASA's Digital Library Technology Project. A project that supports the development of new technologies to facilitate public access to NASA data via computer networks.

The Coalition for Networked Information. A joint project of the Association of Research Libraries, CAUSE, and EDUCOM to promote information resources in networked environments.

The Internet Engineering Task Force. The protocol engineering and development arm of the Internet.

The World Wide Web Consortium. The W3 Consortium exists to develop common standards for the evolution of the World Wide Web.

D-lib magazine
The Magazine of Digital Library Research

ISSN 1082-9873

Access to Stories, Briefings, and Editorials by Title

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

You may also browse by author.

1990 Census LOOKUP: Mining a Mountain of Data, Deane W. Merrill, Nathan G. Parker, Harvard H. Holmes, Chris Stuber, Valerie J. Gregg

A

Access and Discovery: Issues and Choices in Designing DIFWICS, Jeremy Hylton
Access to Digital Objects: A Communications Law Strategy, Patrice A. Lyons
Accessing the Visible Human Project, Michael J. Ackerman
Advanced Web Presentation through Data Modeling: An Open Architecture for the Personalized Webs of the Future, Leon Shklar
Agent-based Architecture for Digital Libraries, William P. Birmingham

B

Berkeley Digital Library SunSITE, Roy Tennant

Briefings:

- ARTFL, Andrea Doane
- Brief Update on the Alexandria Digital Library Project: Constructing a Digital Library for Geographically-Referenced Materials, Terence R. Smith
- CyberStacks(sm): A 'Library-Organized' Virtual Science and Technology Reference Collection, Gerry McKiernan
- Digital Libraries in the Classroom, Elliott Soloway
- EduPort, Miriam Masullo
- Highlights related to the Government Information Locator Service (GILS): Toward a Global Information Locator, Eliot Christian
- Integrated Document Access (IDA) Project, Margaret Colmer
- Making a Digital Library: The Chemistry Online Retrieval Experiment -- A Summary of the CORE Project (1991-1995), Richard Entlich, Lorrin Garson, Michael Lesk, Lorraine Normore, Jan Olsen, Stuart Weibel
- OCLC Internet Cataloging Project, Erik Jul
- Oregon State University's Government Information Sharing Project, Jacquelyn Miller
- Project Muse: 43 Humanities and Social Sciences Journals to Come on the Network
- SuperJournal, David Pullinger, Christine Baldwin
- TURNIP: The URN Interoperability Project, Renato Iannella

- UK Electronic Libraries Programme, Chris Rusbridge
- Z39.50 and the World Wide Web, Sebastian Hammer, John Favaro

C

Caching for Large Scale Systems, Lessons from the WWW, Robert E. McGrath
Content Ratings and Other Third-Party Value- Added Information: Defining an Enabling Platform, Martin Röscheisen, Terry Winograd, Andreas Paepcke
Creating a Networked Computer Science Technical Report Library, James R. Davis

D

Digital Libraries and Corporate Technology Reuse, Jonathan T. Hujsak
Digital Libraries: Searching Is Not Enough; What We Learned On-Site, Andreas Paepcke

E

Economic Framework for Pricing and Charging in Digital Libraries, J. Sairamesh, C. Nikolaou, D. Ferguson, Y. Yemini

Editorials:

- Double-Edged Sword of Access
- Future is a Complex Place
- Getting Used to Technology and Revolutions in the Making
- Levels of Abstraction
- Taking the Measure of the Net
- Telling Time
- Text Is More Than Just Words on a Page, Susan Hockey
- Tragedy of the Commons, Revisited (Again)
- When is Honesty the Best Policy?
- Where Are We in Space?
- Word (or two) of welcome

F

French Minitel: Is There Digital Life Outside of the "US ASCII" Internet? A Challenge or Convergence? Jack Kessler

G

Global Change Data and Information System-Assisted Search for Knowledge (GC-ASK) Project, Roberta Y. Rand

H

Historical Collections for the National Digital Library: Lessons and Challenges at the Library of Congress, *(Part I)*, Caroline R. Arms
Historical Collections for the National Digital Library: Lessons and Challenges at the Library of Congress *(Part II)*, Caroline R. Arms

I

Image Browsing in the Alexandria Digital Library (ADL) Project, B.S. Manjunath

K

Key concepts in the architecture of the digital library , William Y. Arms

M

Metadata: the foundations of resource description, Stuart Weibel
Model Editions Partnership: Historical Editions in the Digital Age, David Chesnutt

N

Need for a Common Infrastructure: Digital Libraries and Electronic Commerce, Daniel Schutzer
Netlib Mathematical Software Repository, Shirley Browne, Eric Grosse, Tom Rowan
New Center at Columbia University for Digital Library Research: Fostering Interdisciplinary Research and Bridging Cultural Clashes, Judith Klavans
News-on-Demand: An Application of Informedia® Technology, Alexander G. Hauptmann, Michael J. Witbrock and Michael G. Christel

O

Options for the Future, Joshua Lederberg

R

Recent Developments in GALEN II: Evolution of a Digital Library for the Health Sciences, John A. Kunze, Brian N. Warling
The Red Sage Project: An Experimental Digital Journal Library for the Health Sciences, A Descriptive Overview, Richard E. Lucier and Peter Brantley
Research in Support of Digital Libraries at Xerox PARC; Part I: The Changing Social Roles of Documents, Marti A. Hearst

S

SCAM Approach to Copy Detection in Digital Libraries, Narayanan Shivakumar, Hector Garcia-Molina
Secure Repository Design for Digital Libraries, Carl Lagoze
SunSITE: Serving Your Internet Needs Since 1992, Judson Knott, Paul Jones

T

Task Force on Archiving of Digital Information, John R. Garrett
Text Is More Than Just Words on a Page, Susan Hockey *(guest editorial)*

U

Uniform Resource Names: A Progress Report, The URN Implementors
User-Centered Iterative Design for Digital Libraries: The Cypress Experience, Nancy A. Van House, Mark H. Butler, Virginia Ogle, Lisa Schiff

W

What Do People Want from Information Retrieval? (The Top 10 Research Issues for Companies that Use and Sell IR Systems), W. Bruce Croft
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

Y

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

address both, hence the distinction made in the second definition stated in the beginning of this report.

III. Research agenda for Social Aspects Of Digital Libraries

We will present the research agenda with respect to the two definitions of digital libraries outlined above. These two definitions converge in a model of the life cycle of information and information processes.

The model covers the sequence from the creation of information (author, artist, memo-writer, data-generation scientist, publisher, etc.), through the searching for information, and the utilization of it, often for very different purposes than it was originally created. An exit from the loop is given to indicate that we do not need to save everything created in digital form -- indeed, we need criteria and mechanisms to decide what to keep and what to destroy. The model addresses the social context for all aspects of the cycle -- people create information for one purpose, search for it for another, and utilize for another. We need to organize, describe, and represent for multiple uses but we must design based on an understanding of what those uses might be. Similarly, we need searching and utilization interfaces that support many perspectives and purposes, with a variety of functional capabilities -- but all must be based on some understanding of the underlying tasks/roles that the information will play in a social context.

Home W Groups

clb/wya
Last revised: March 18, 1996

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

- Introduction
- Agents
- What the architecture provides
- The Conspectus and the conspectus language
- Status and summary
- Acknowlegements
- References

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

span multiple databases, each residing in a different place in cyberspace.

Furthermore, the creativity of users of digital libraries will spawn thousands of different, specialized services (e.g., notification and translation, even special collections of information). Perhaps most importantly, methods of organizing information will transcend a single "digital library," in that it is unlikely that a single indexing or naming scheme (e.g., the Dewey Decimal System) will be used across the multiplicity of digital libraries that are sure to emerge. Thus, we must create flexible software architectures that can federate as many content suppliers, information-organizational schemes, and service providers as possible, and yet scale to the extremely large size needed to support the digital libraries of the future.

Considering this view of digital libraries, we have developed some guidelines and objectives for our system. First, the guidelines:

- Given that many digital libraries will emerge, we want to make ours as attractive to users and content and service providers as possible. Thus, we intend to make the fewest and least-restrictive standards.
- The only way to ensure intellectual property in the future is to provide incentives for its creation. Thus, we intend to make support for economic incentives an integral part of the UMDL architecture. This covers a wide range of issues, from definition and protection of intellectual property rights through payment for the use of intellectual property. Please note: *we are not establishing policies related to rates or payment, rights of users to access the contents of the library, or other related issues such as "fair use"*. We simply plan to have the machinery in place to support whatever policies may arise in the future.
- The elements of the library (services and collections of information) are autonomous in that these elements will make decisions based on their own perspective. In other words, there is no central authority that can press an element into service. All elements are considered peers, and thus interaction is achieved entirely through negotiation processes.

Broadly speaking, the objectives of the architecture are to provide services that fall under the following categories:

- Registration: maintaining a comprehensive list of all the agents (collections, user interface, and others) in the UMDL.
- Brokering and teams of agents formation: finding potential information sources and support services (e.g., translation of query languages) to fulfill a user's information needs.
- Commerce support: providing mechanisms to support commerce for information goods, and protecting intellectual property and privacy.

Furthermore, we require that the architecture have the following properties:

- Modular, in that new elements can be added or removed without effecting the operation of other elements;
- Scaleable, to allow for the potential of millions of constituent elements;
- Extensible, to allow new elements (collections, data types, services, etc.) to be easily added to the digital library.
- In the remainder of this paper, an overview of the UMDL architecture is given. We describe the notion of software agents and types of agents in UMDL, and then describe how agents interact to provide service.

Agents

The architecture is based on the notion of a software agent. An agent represents an element of the digital library (collection or service), and is a highly encapsulated piece of software that has the following special properties:

- Autonomy: the agent represents both the capabilities (ability to compute something) and the preferences over how that capability is used. Thus, agents have the ability to reason about how they

use their resources. In other words, an agent not have to fulfill every request for service, only those consistent with its preferences. A traditional computer program does not have this reasoning ability.
- Negotiation: since the agents are autonomous, they must negotiate with other agents to gain access to other resources or capabilities. The process of negotiation can be, but is not required to be, stateful and will often consist of a "conversation sequence", where multiple messages are exchanged according to some prescribed protocol, which itself can be negotiated.

Autonomy is critical to scaling UMDL to a large size because autonomy implies local or decentralized control. As a result, we do not have to update some "master" program everytime a new agent is added to UMDL. The effects of adding or removing an agent are propagated locally using a set of protocols. Thus, there is no need for global coordination among all agents [4]. The notion of decentralized control of autonomous agents is similar to the way our economy works. Each of us is similar to an agent in that the decision about how money is spent is done individually. These spending decisions do not require communication across the entire economy (e.g., when ones buy a car, she do not need to tell the whole country or even the car manufacturer, just the car dealer), nor does one need to get permission from a central authority. Similarly, UMDL agents can make decisions and form teams at a local level, without requiring interaction with all agents in the system or with a central authority.

Negotiation is complementary to autonomy, in that autonomous agents must be capable of making binding commitments for the system to work. Thus, when agents negotiate and strike a deal (i.e., something of value is exchanged for something else of value), the agents are bound to fulfill that deal. It is possible, and even likely, that some deals will allow agents to back out. This "feature", however, must be explicitly negotiated in our system.

The UMDL is populated by three classes of agents:

- UIAs (User Interface Agents) provide a communication wrapper around a user interface. This wrapper performs two functions. First, it encapsulates user queries in the proper form for the UMDL protocols. Second, it *publishes* a profile of the user to appropriate agents, which is used by mediator agents to guide the search process.
- Mediator agents [8], of which there are many types, perform a variety of functions: essentially, all tasks that are required to refer a query from a UIA to a collection, monitor the progress of the query, transmit the results of a query, and perform all manner of translation and bookkeeping. Presently, two types of mediators populate the UMDL. Registry agents capture the address and contents of each collection. Query-planning agents [5] receive queries and route them to collections, possibly consulting other sources of information to establish the route. Another special class of mediators currently being developed, called facilitators [7], mediate negotiation among agents [3].
- CIAs (Collection Interface Agents) provide a communication wrapper for a collection of information. While performing translation tasks similar to those performed by the UIA for a user interface, the CIA also publishes the contents and capabilities of a collection in the *conspectus* language (described in the next section, "What the architecture provides").

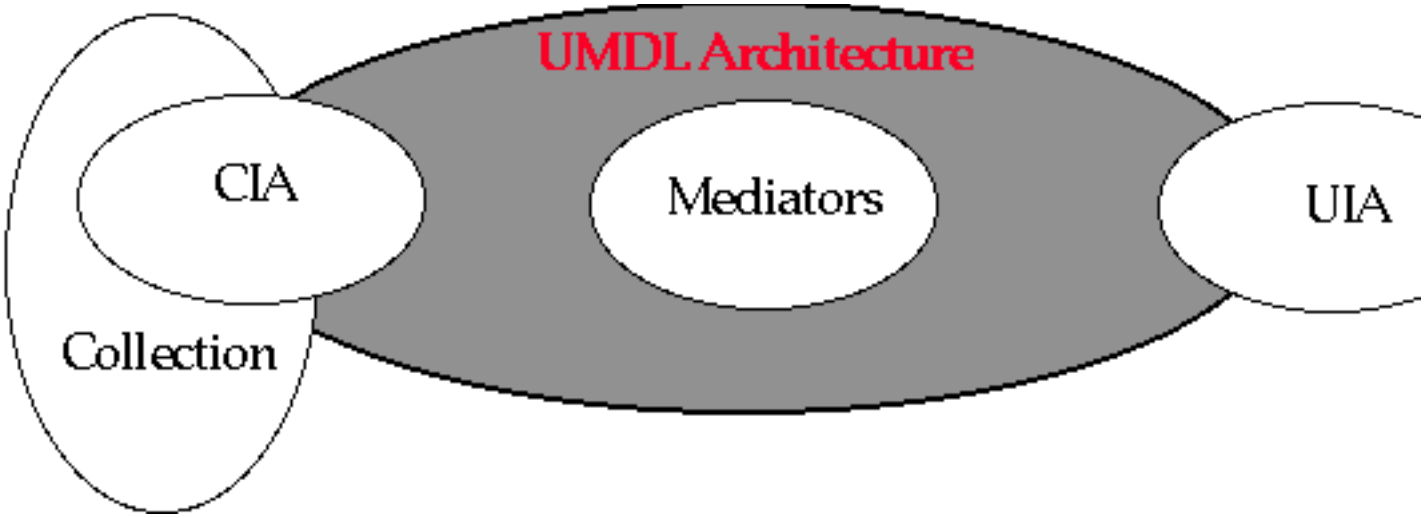


Figure 1: UMDL agent types

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.

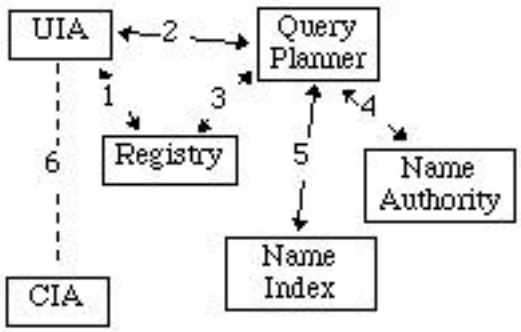


Figure 2: Example search by author

We separate the activities of agents UMDL into two types: that used to organize agents to perform the team building (called *architectural*), and that used to perform the actual task (called *task*), such as actually querying a database. Strictly separating these activities allows us to reduce the commitments that an agent must make to operate in our system (i.e., a CIA is not required to support all query languages used in UMDL, only those it chooses to support.). Thus, we require only that agents use a language, called the *conspectus language*, designed to support architectural activities (see the next section, "The Conspectus and the conspectus language"); the decision to support any particular task language is left up to individual agents.

The distinction between architecture and task has advantages and disadvantages. The advantages include minimal standards, and therefore increased flexibility in creating agents. Furthermore, the agents themselves are smaller, and therefore easier to build and maintain. A disadvantage is that not all agents will have access to all other agents. For example, if a CIA supports only Z39.50, but a UIA uses some other language X (and no mediator exists that can translate X to Z39.50), then that UIA cannot access the CIA. We see, however, no practical solution to this problem at this time.

Since it is impossible to create an architecture that has everything, we prefer flexibility over guaranteed interoperability among all agents. Task languages that will undoubtedly evolve over time, as we learn more about digital libraries. By being neutral on which languages are supported, we avoid having to rewrite significant portions of our software as the languages change.

The Conspectus and the conspectus language

The space of information in UMDL is potentially enormous, as is the possibility of bringing the system to its knees with rogue query processes. To limit queries to potentially applicable CIAs, we reason about the contents of each collection to derive an estimate of their likely usefulness. This leads us to a two-level partition of the information space:

- Conspectus: includes, among other things, the content of the collection, the search capabilities of the search engine(s) associated with the collection, and the structure of the material (documents) in the collection.
- Collection: the set of actual documents in a collection. These documents are in native formats, and the search engines are engaged through native query languages.

The conspectus is an abstracted description of the aggregate of collections populating the UMDL. Additionally, the conspectus is a *normalized* description of content. This is important, as various collections will have different methods for describing the same thing (e.g., title as TI or TL). To help normalize terms, we are using a variety of thesauri developed by various researchers around the world.

The conspectus is written in a language that we have defined (the UMDL conspectus language, UCL). Although we retain complete control over the UCL, the actual conspectus expressed in UCL will be specified

by the separate collections. Our aim is that UCL (and its associated resources, such as various thesauri and cataloging systems) provide sufficient structure for developing compatible representations of collections. Thus, the conspectus provides interoperability for various search and retrieval methods through a common representation over collections.

Since the conspectus will be large both in scope and in size, it will be distributed and hierarchically organized. We expect to create special mediator agents whose sole responsibility is to maintain the integrity of the conspectus.

Agents communicate using patterns of messages, where the content of the message is specified by UCL and sets of *performatives* describing the purpose of the communication (e.g., to ASK or TELL something) [6]. The messages transmitted between the agents describe capabilities, services, and other primitives. For example, all agents use the ASK performative to make requests to the registry for notification about classes of agents with certain capabilities. The registry agent continues sending information about these agents, as they come on-line, until the UNASK performative is received.

Another example performative set is TELL, which is typically used in response to an ASK. The registry agent uses TELL to send the names of agents that correspond to some capability specification. The registry agent uses the UNTELL performative to express that an agent is no longer available, or that its capabilities have changed.

Protocols specify communication patterns among agents. In order to participate in UMDL, an agent must use our protocols. Since these protocols are minimally restrictive in how a task is accomplished, we believe they are not a significant impediment to the development of agents by third parties. Standardizing the protocols, but not the task languages, strikes a balance between flexibility and ease of integration into the UMDL environment.

The agent-identification protocol (used by both the CIA and query planner in the example depicted in Figure 2) provides a way for agents to locate other agents with specific capabilities (Figure 3). The requesting agent (*R*) uses the ASK performative to describe the specific capabilities to the registry agent. The registry agent executes a lookup operation to match the specifications to the agents it knows about. Any matches are sent to the requesting agent via the TELL performative. The ASK performative implies a standing request for information about agents, so that the registry agent continues to send *R* information about other agents as they advertise their capabilities. When *R* receives information about an agent (*A*) from the registry agent, it has the option of storing that information in its local knowledge base for future use.

If *R* no longer wants to receive information about an agent, then it uses the UNASK performative to communicate this desire to the registry agent. Upon receipt of the UNASK performative, the registry agent stops sending information to *R*. If *A* is no longer available, or has a change in capabilities, then the registry agent sends the UNTELL performative to all agents who received a TELL performative about *A*. Thus, the registry agent must keep track of the agents to which it sent the TELL performative.

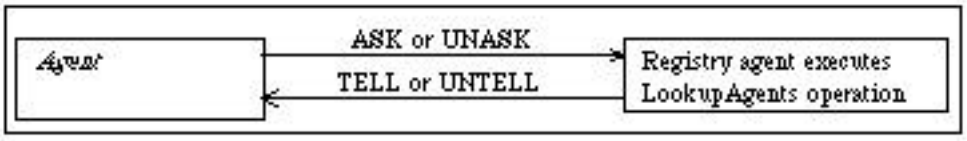


Figure 3: Agent-identification protocol.

The performative and protocol features of the UMDL architecture are general enough to accommodate a variety of actions within the library. As illustrated here, the same protocol can be used by several different agents to achieve their objectives. We expect that once we have established a basic set of protocols, including those for negotiations about intellectual property, they will become relative stable even though the variety of information

and services in the library will grow enormously. In fact, the stability of these protocols is the foundation for growth of the system.

Status and summary

The UMDL is operational, and can be accessed through http://www.sils.umich.edu/Catalog/UMDL.html . The current system has about 50 CIAs and basic search support. We expect to have subscription, notification, and known-item search running by the end of the calendar year. Two task languages are supported: Z39.50 and FTL (a locally created query language).

The current system demonstrates that the agent architecture approach outlined in this article is viable, and paves the way for more interesting experiments with scaling both the total number of agents as well as the types of services and collections available. It is interesting to note that the architecture was able to handle the addition of new services (new collections and a notification service) without modifications to existing agents and protocols, thus demonstrating properties of scaleability, extensibility, and modularity.

Acknowledgments

The members of the UMDL architecture group contributed many of ideas presented here. In particular, Fritz Freiheit provided helpful suggestions to drafts of this paper.

The UMDL project is funded under a joint initiative of NSF/ARPA/NASA; we are grateful for their financial support, and their enthusiasm for the initiative. The views expressed in this paper are those of the author only, and do not necessarily represent the views of the funding agencies (nor of the UMDL project).

References

1. Birmingham, W. P., K. M. Drabenstott, C. O. Frost, et al. (1994). The University of Michigan Digital Library: This is not your father's library. *Digital Libraries '94*, College Station, TX.
2. Birmingham, W. P., E. H. Durfee, T. Mullen, et al. (1995). The distributed agent architecture of the University of Michigan Digital Library. *AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, Stanford, CA, AAAI Press.
3. D'Ambrosio, J. and W. P. Birmingham (1995). Preference-directed design. *AI in Engineering, Design, Analysis, and Manufacture* . To appear.
4. Darr, T. P., and W. P. Birmingham (1994). Automated design for concurrent engineering. *IEEE Expert* **9**(5): 35-42.
5. Durfee, E. H. and T. A. Montgomery (1991). Coordination as distributed search in a hierarchical behavior space. *IEEE Transactions on Systems, Man, and Cybernetics*, Special Issue on Distributed Artificial Intelligence, **21**(6):1363-1378.
6. Finin, T., R. Fritzon, D. McKay, et al. (1994). KQML as an agent communication language. *Third International Conference on Information and Knowledge Management*, ACM Press.
7. Mullen, T., and M. P. Wellman (1995). A simple computational market for network information services. *First International Conference on Multi-agent Systems*, San Francisco, CA.
8. Wiederhold, G. (1992). Mediators in the architecture of future information systems. *Computer* **26**(3): 38-49.

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

- 1. The technical framework exists within a legal and social framework
- 2. Understanding of digital library concepts is hampered by terminology
- 3. The underlying architecture should be separate from the content stored in the library
- 4. Names and identifiers are the basic building block for the digital library
- 5. Digital library objects are more than collections of bits
- 6. The digital library object that is used is different from the stored object
- 7. Repositories must look after the information they hold
- 8. Users want intellectual works, not digital objects
- Reference

d-Lib forum d-Lib magazine

General Principles

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

Early networked information systems were developed by technical and professional communities, concentrating on their own needs. The emphasis was on making information available to colleagues and the public, without charge. The digital library of the future will exist within a much larger economic, social and legal framework.

For example, musical works and their performance represent the livelihood of composers and musicians. Their artistic reputations often depend on their work not being changed in storage or transmission. They require payment, as do recording studios and concert halls. Such work will only be part of the digital library, if the library supports their interests.

The legal system's task is to codify this rapidly changing economic and social framework. The relevant areas of law include copyright, performance, and other intellectual property, libel and obscenity, communications law, privacy, and international law.

The Kahn/Wilensky architecture can not write the law, but it provides a technical design that matches the legal structure that is expected to emerge. The architecture respects the creators and owners of intellectual property. It allows the preservation of rights that can last for more than one hundred years, and recognizes that digital works may include material from many sources, with separate property rights.

Society expects the creators of works to be responsible for their content, and for those who make decisions about content to behave responsibility. However, the digital library will not thrive if legal liability for content is placed upon parties whose only function is storage and transmission. Therefore, the architecture establishes clear boundaries between the areas of responsibility of the various parties.

2. Understanding of digital library concepts is hampered by terminology

Terminology proves to be a barrier in describing a digital library. Some words have such strong social, professional, legal, or technical connotations that they obstruct discussion between people of varying backgrounds. Simple words mean different things to different people. For example, the words "copy" and "publish" have different meanings to computing professionals, publishers, and lawyers. Common English usage is not the same as professional usage, and the versions of English around the world have subtle variations of meaning.

Certain words cause such misunderstandings that they are best expunged from any precise discussion of the on-line digital library. The list includes "copy", "publish", and "document". Other words have to be used very carefully and their exact meaning made clear whenever they are used. An example is "content".

In the Kahn/Wilensky architecture, items in the digital library are called "digital objects". They are stored in "repositories" and identified by "handles". Information about the digital object is known as "properties" or "metadata".

3. The underlying architecture should be separate from the content stored in the library.

A conventional research library stores more than books, and the digital library stores more than digitized text. Almost every type of information can be represented in digital form, including text, pictures, musical works, computer programs, databases, models and designs, video programs, and compound works combining many types of information.

The underlying architecture of the digital library, as described by Kahn and Wilensky, specifies those characteristics that apply to all types of material. For example, every object needs to have a name or identifier; the actions of adding objects to the library or deleting them apply to all material; general purpose methods of security can be provided.

This underlying architecture is a base for extensions that can be tailored for various types of information. The

extensions typically include specific formats, protocols, and rights management that are appropriate for the type of material. For example, the extensions for digitized movies will be very different from those for video games; texts are usually described by bibliographic terms, such as author and title, which are of little relevance to a computer program; a protocol designed for interaction with a database is unlikely to be useful in manipulating graphic designs.

Separating general functions from those specific to the type of content has other benefits. It encourages different markets to emerge, and allows a legal framework in which storage, transmission and delivery of digital objects is separate from activities to create and manage the intellectual content.

4. Names and identifiers are the basic building block for the digital library

Names are a vital building block for the digital library. Names are needed to identify digital objects, to register intellectual property in digital objects, and to record changes of ownership. They are required for citations, for information retrieval, and are used for links between objects.

These names must be unique. This requires a administrative system to decide who can assign them and change the objects that they identify. They must last for very long time periods, which excludes the use of an identifier tied to a specific location, such as the name of a computer. Names must persist even if the organization that named an object no longer exists when the object is used. There need to be computer systems to resolve the name rapidly, by providing the location where an object with a given name is stored.

The Corporation for National Research Initiatives has implemented a handle system which satisfies these requirements. A "handle" is a unique string used to identify digital objects. The handle is independent of the location where the digital object is stored and can remain valid over very long periods of time. A global handle server provides a definitive resource for legal and archival purposes, with a caching server for fast resolution. The computer system checks that new names are indeed unique, and supports standard user interfaces, such as Mosaic. A local handle server is being added for increased local control.

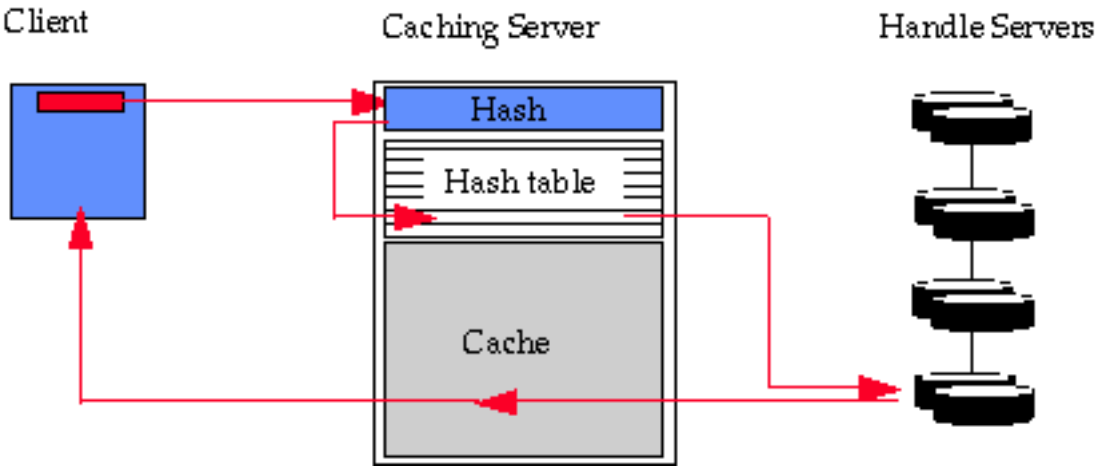


Figure 1. The CNRI handle system

5. Digital library objects are more than collections of bits

In the digital library, information is stored as "digital objects". A primitive idea of a digital object is that it is just a set of bits, but this idea is too simple. The content of even the most basic digital object has some structure, and information, such as intellectual property rights, must be associated with the digital object. Figure 2 shows that a digital object in a repository has two parts, content and associated data, sometimes called "metadata".

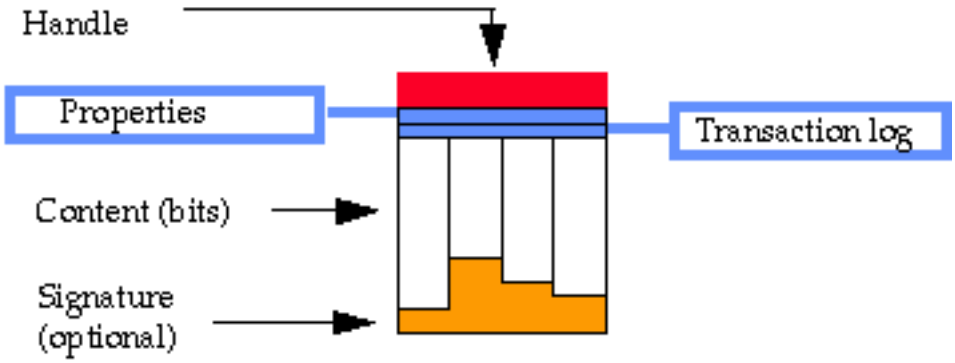


Figure 2. Parts of a digital object

To enable the content to represent useful information, its type must be known. Thus part of the content may be of type text (perhaps encoded in a mark-up language), while another part may be of type audio. A single digital object may contain many types of content. It turns out that arbitrarily complex data types can be constructed from a few basic types, notably bit-sequences, handles and other digital objects. By combining these in various combinations, any digital content can be represented.

To manage valuable intellectual property, certain metadata is required. This is shown in the figure. It always includes a unique identifier (the handle). It may also include properties such as rights and access methods. For example, one property states whether a digital object is mutable, in that it may be altered after being placed in a repository. Another is a digital signature or other method of validating that an object has not been changed. Frequently, it is useful to keep a log of all transactions associated with each digital object.

6. The digital library object that is used is different from the stored object

In the digital library, what you store is not what you get. The architecture must distinguish carefully between digital objects as they are created by an originator, digital objects stored in a repository, and digital objects as disseminated to a user.

The user receives the result of executing a program on the stored object. This may be a simple program, such as a file transfer program, or something very complex. For example, an image is stored in a library as a set of wavelets. To use it, the stored wavelets are used to generate an image with the characteristics requested. This is transmitted over the network to a user's computer, where it can be further processed or displayed.

Some classes of digital objects can be provided it to a user in more than one way. For example, the score of a musical work is held in the library. One form of use is to transmit a representation of the score to the user's computer. Alternatively, the user could request the repository to execute a synthesizer program, which would perform the score, and transmit the digitally encoded audio over the network. For some types of object, such as a data base or a video game, the use consists of an interaction between the user and the execution of the program.

Legal scholars see an interesting parallel between the computer viewpoint of executing a program to supply a digital object to a user and the legal concept of performance. This may prove to be the correct framework for managing rights in a digital library.

7. Repositories must look after the information they hold

A repository stores digital objects, both the content and the metadata.

A digital object as stored in a repository may be very different from the digital object that is made available to

users' computers. Different repositories will have very different internal organizations, but for each digital object every repository will have a properties record, which holds attributes of the object, and a transaction log.

Since digital objects may contain valuable intellectual property, the stored form of a digital object within the repository includes information that allows for it to be managed within economic and social frameworks. The repository maintains this information, provides basic reference information, and provides security to ensure that only valid operations are carried out on the digital objects.

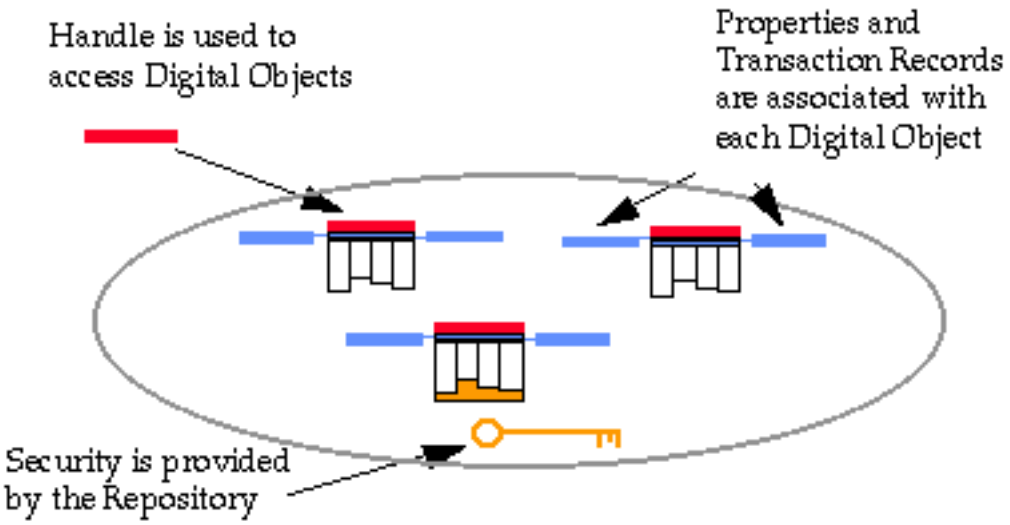


Figure 3. A repository

The internal organization of a repository and the way that digital objects are stored are hidden from the user. A simple protocol is provided for interactions with the repository. This protocol is called the "repository access protocol." The basic commands in this protocol are those to access a digital object and its metadata, and the service request to disseminate a digital object. In addition there are commands to add and delete digital objects.

8. Users want intellectual works, not digital objects

Digital objects are the basic building blocks of the digital library, but users of the library usually want to refer to items at a higher level of abstraction. Common English terms, such as "technical report", "computer program", or "musical work", often refer to many digital objects that can be grouped together. The individual objects may have different formats, minor differences of content, different usage restrictions, and so on, but certain users are willing to consider them as equivalent.

Which digital objects should be grouped together can not be specified in a few dogmatic rules. The decision depends upon the context, the specific objects, their type of content and sometimes the actual content. The underlying architecture has to support two main needs. It must provide methods for grouping digital library objects and must provide means for retrieval.

The Kahn/Wilensky architecture supports these higher level ideas in several ways. One is to have a digital object containing several digital objects. Thus several formats of a text might be assemble into a single digital object. Another approach is to have these variants stored as separate digital objects, each with its own handle. These handles are contained in a digital object, known as a "meta-object", which acts like a catalog record. It contains a list of the variants with their handles and information about the differences amongst them.

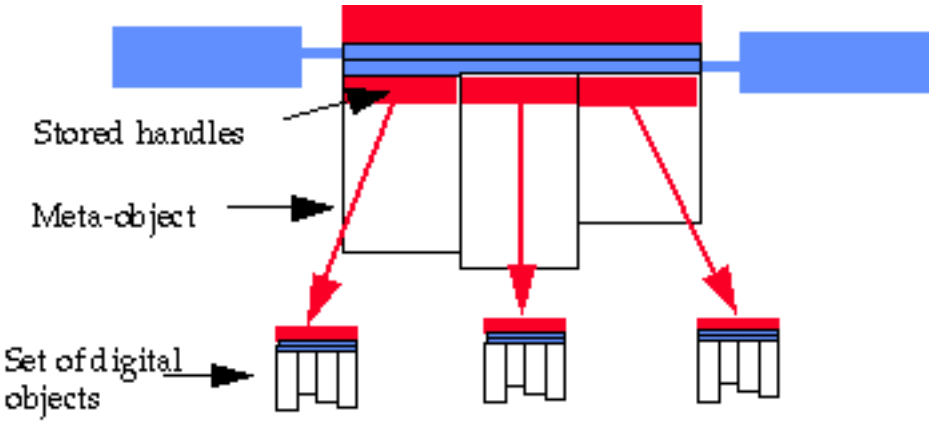


Figure 4. A digital object used as a catalog record

Reference

hdl:cnri.dlib/tn95-01 Kahn, Robert and Wilensky, Robert. "A framework for distributed digital object services". May, 1995. (http://WWW.CNRI.Reston.VA.US/home/cstr/arch/k-w.html)

---

d-Lib forum d-Lib magazine

---

hdl:cnri.dlib/july95-arms.html

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
- Underlying Assumptions
- Implementations
- Next Steps
- References

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.

An alternative solution that promises to mediate these extremes involves the creation of a record that is more informative than an index entry but is less complete than a formal cataloging record. If only a small amount of human effort were required to create such records, more objects could be described, especially if the author of the resource could be encouraged to create the description. And if the description followed an established standard, only the creation of the record would require human intervention; automated tools could discover these descriptions and collect them.

Can a simple metadata record be defined that sufficiently describes a wide range of electronic objects? The Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA) convened the invitational Metadata Workshop on March 1-3, 1995, in Dublin, Ohio to address this issue. Fifty-two librarians, archivists, humanities scholars and geographers, as well as standards makers in the Internet, Z39.50 and Standard Generalized Markup Language (SGML) communities, met to identify the scope of the problem, to achieve consensus on a list of metadata elements that would yield simple descriptions of data in a wide range of subject areas, and to lay the groundwork for achieving further progress in the definition of metadata elements that describe electronic information.

Goals

Goals of the workshop included fostering a common understanding of the problems and potential solutions among the stakeholders and promoting a consensus on a core set of metadata elements to describe networked resources.

Scope

Since the Internet contains more information than professional abstractors, indexers and catalogers can manage using existing methods and systems, it was agreed that a reasonable alternative way to obtain usable metadata for electronic resources is to give authors and information providers a means to describe the resources themselves. The major task of the Metadata Workshop was to identify and define a simple set of elements for describing networked electronic resources. To make this task manageable, it was limited in two ways. First, only those elements necessary for the discovery of the resource were considered. It was believed that resource discovery is the most pressing need that metadata can satisfy, and one that would have to be satisfied regardless of the subject matter or internal complexity of the object.

Secondly, the discussion was further restricted to the metadata elements required for the discovery of what were called **document-like objects**, or **DLOs** by the workshop participants. It was believed that DLOs are still the most common type of resource sought in the Internet and that whatever solution could be proposed for DLOs could be extended to other kinds of resources. More importantly, the likelihood of making progress on this challenging problem would be increased if attention could initially be restricted to something familiar.

DLOs were not rigorously defined, but were understood by example. For example, an electronic version of a newspaper article or a dictionary is a DLO, while an unannotated collection of slides is not. Of course, the crux of the problem is that in a networked environment, DLOs can be arbitrarily complex because they can consist of text with callouts to images, audio or video clips, or to other hypertext documents. The Metadata Workshop participants made no attempt to limit the complexity of DLOs, except to say that the intellectual content of a DLO is primarily text, and that the metadata required for describing DLOs will bear a strong resemblance to the metadata that describes traditional printed texts.

As a result of the restricted focus of the workshop, certain issues required for a complete description of DLOs, such as cost, archival status and copyright information, were eliminated from the scope of the discussion. Elements required for the description of objects other than DLOs, such as the elements required for the description of complex geological strata in a geospatial resource, were also beyond the scope of the discussion. The goal was to define a core set of metadata elements that would allow authors and information providers to describe their work and to facilitate interoperability among resource discovery tools. But because the core elements do not yield a complete description of objects in a networked environment, careful consideration was also given to mechanisms for extending the element set.

The primary deliverable from the workshop was a set of thirteen metadata elements, named the **Dublin Core Metadata Element Set** (or Dublin Core, for short). The Dublin Core was proposed as the minimum number of metadata elements required to facilitate the discovery of document-like objects in a networked environment such as the Internet. The syntax was deliberately left unspecified as an implementation detail. The semantics of these elements was intended to be clear enough to be understood by a wide range of users.

Below is a brief description of the elements in the Dublin Core **Dublin Core Element Description**

- **Subject:** The topic addressed by the work
- **Title:** The name of the object
- **Author:** The person(s) primarily responsible for the intellectual content of the object
- **Publisher:** The agent or agency responsible for making the object available
- **OtherAgent:** The person(s), such as editors and transcribers, who have made other significant intellectual contributions to the work
- **Date:** The date of publication
- **ObjectType:** The genre of the object, such as novel, poem, or dictionary
- **Form:** The physical manifestation of the object, such as Postscript file or Windows executable file
- **Identifier:** String or number used to uniquely identify the object
- **Relation:** Relationship to other objects
- **Source:** Objects, either print or electronic, from which this object is derived, if applicable
- **Language:** Language of the intellectual content
- **Coverage:** The spatial locations and temporal durations characteristic of the object

To make this discussion concrete, consider an electronic a record created with the relevant portions of the Dublin Core, and a sample syntax, that describes an electronic version of Maya Angelou's poem "On the Pulse of Morning". This description is based on a record created by the University of Virginia Library's Electronic Text Center. (For a description of that project, see Gaynor [Gaynor] .)

- **Subject:** Poetry
- **Title:** On the Pulse of Morning
- **Author:** Maya Angelou
- **Publisher:** University of Virgina Library Electronic Text Center
- **OtherAgent:** Transcribed by the University of Virginia Electronic Text Center
- **Date:** 1993
- **Object:** Poem
- **Form:** 1 ASCII file
- **Identifier:** AngPuls1
- **Source:** Newspaper stories and oral performance of text at the presidential inauguration of Bill Clinton
- **Language:** English

Underlying Assumptions

The discussions at the Metadata Workshop revealed several principles that should guide the further development of the element set. Adherence to these principles increases the likelihood that the core element set will be kept as small as possible, that the meanings of the elements will be understood by most users, and that the element set will be flexible enough for the description of resources in a wide range of subject areas. These principles are intrinsicality, extensibility, syntax independence, optionality, repeatability, and modifiability.

Intrinsicality

The Dublin Core concentrates on describing intrinsic properties of the object. Intrinsic data refer to the properties of the work that could be discovered by having the work in hand, such as its intellectual content and physical form. This is distinguished from extrinsic data, which describe the context in which the work is used.

For example, the "Subject" element is intrinsic data, while transaction information such as cost and access considerations are extrinsic data. The focus on intrinsic data in no way demeans the importance of other varieties of data, but simply reflects the need to keep the scope of deliberations narrowly focussed.

Extensibility

In addition to its use in dealing with extrinsic data, extension mechanisms will allow the inclusion of intrinsic data for objects that cannot be adequately described by a small set of elements.

Extensibility is important because users may wish to add extra descriptive material for site-specific purposes or specialized fields. In addition, the specification of the Dublin Core itself will change over time, and the extension mechanism will allow revisions while maintaining some backward compatibility with the originally defined element set.

Syntax Independence

Syntactic bindings are avoided because it is too early to propose formal definitions and because the Dublin Core is intended to be eventually used in a range of disciplines and application programs.

Optionality

All the elements are optional. The Dublin Core may eventually be applied to objects for which some elements have no meaning (who is the author of a satellite image?). It also seems counterproductive to mandate complex descriptions if the creators of the content are expected to provide the descriptive material. A simple description is better than no description at all.

Repeatability

All elements in the Dublin Core are repeatable. For example, multiple author elements would be used when a resource has multiple authors.

Modifiability

Each element in the Dublin Core has a definition that is intended to be self-explanatory. However, it is also necessary that the definitions of the elements satisfy the needs of different communities. This goal is accomplished by allowing each element to be modified by an optional qualifier. If no qualifier is present, the element has its common-sense meaning; otherwise, the definition of the element is modified by the value of the qualifier.

Qualifiers will be typically derived from well-known conventions in the library community or from the field of knowledge appropriate to the resource. Qualifiers are important because they give the Dublin Core a mechanism for bridging the gap between casual and sophisticated users. For example, the data in the **Subject** element consists of any word or phrase that describes the object's content. However, a professional cataloger may wish to supply the name of the authoritative source from which the subject terms are taken. In such a case, the element may be written as **Subject (scheme=LCSH)** , indicating that the subject terms are taken from the Library of Congress Subject Headings.

Implementations

One of the goals of the OCLC/NCSA Metadata Workshop was to promote prototype resource description projects based on a common model of resource description. A number of Metadata Workshop conferees represent organizations that have ongoing activities or are starting activities that will be influenced by the results of the workshop. These include:

- The OCLC Spectrum Project
  Contact:Diane Vizine-Goetz, *vizine@oclc.org*
- The OCLC Internet Resources Cataloging Project
  Contact:Erik Jul, *jul@oclc.org*
- Library of Congress
  Contact:Rebecca Guenther, *rgue@loc.gov*
- O'Reilly Associates
  Contact:Terry Allen, *terry@ora.com*
- Los Alamos National Laboratory and Indiana University
  Contact:Ron Daniel Jr.,*rdaniel@acl.lanl.gov*
  Contact:Pete Percival,*percival@bronze.ucs.indiana.edu*
- Bunyip Systems
  Contact:Chris Weider,*clw@bunyip.com*
- Georgia Institute of Technology
  Contact:Michael Mealling, *michael.mealling@oit.gatech.edu* , http://www.gatech.edu/iiir
- SoftQuad
  Contact: Yuri Rubinsky,*yuri@sq.com*
- Concordia University
  Contact:Bipin Desai, *bcdesai@cs.concordia.ca*,
  http://www.cs.concordia.ca/~faculty/bcdesai/cindi-system-1.0.html

Next Steps

Refinement and standardization of the metadata element set defined in this document will be an ongoing, dynamic process involving many stakeholder communities. No single forum will suffice to air all concerns and no single standard can be expected to accommodate the needs of all communities. The problem must be divided into manageable chunks and the process must engage the relevant stakeholder communities. Implicit in the present activity is the proposition that there are core elements common to many object types, and that a simple, extensible framework of such elements can be defined to support more complete resource descriptions.

The initial objective--the specification of elements for the discovery of document-like objects--can be extended in a variety of directions:

- Expansion of the Dublin Core to include other object types, such as services or collections.
- Expansion of the Dublin Core to embrace functionality other than resource discovery, such as archival control and the authentication of users and charging mechanisms.
- Establishing standardized methods for extensibility.
- Refinement of existing work. The Dublin Core is an untested approach to the description of resources that will need to be modified with experience.

OCLC and NCSA will establish a workshop series to address aspects of this agenda. A Metadata Workshop Steering Committee will be established to define topics and assure appropriate representation of stakeholders. Design groups of perhaps a dozen or fewer individuals will be solicited to prepare discussion papers to focus workshop activities. Participants will be invited based on their publicly evident accomplishments in relevant areas or by reviewed application. Workshops will be limited to 50 or fewer participants and conducted in roughly the style of the March 1995 Workshop.

Other work will be done in coordination with IETF working group on Uniform Resource Identifiers (URIs) to assure that the results can be integrated into the emerging protocols for resource location and persistent naming.

Finally, active promotion of results will be carried out by establishing liaison with formal associations of stakeholders. In the library community, MARC standards evolve under the guidance of the Machine-Readable Bibliographic Information Committee (MARBI), composed of representatives of the Library of Congress and other stakeholders in the library community. A close relationship should be sustained between this committee

and the Metadata Work Group. Relationships should also be established with publishers, document vendors, SGML vendors and theoreticians working on the problem of text encoding. Other communities also have requirements that must be accommodated in any framework for resource description. These communities include the GIS community, government information providers and business communication groups.

References

[MARC]
Network Development and MARC Standards, Office, ed. 1994. USMARC Format for Bibliographic data. 1994. Washington, DC: Cataloging Distribution Service, Library of Congress.

[TEI]
Sperberg-McQueen, C. M., and Leu Burnard, ed. 1994. Guidelines for Electronic Text Encoding and Interchange. Chicago and Oxford: Text Encoding Initiative.

[Gaynor]
Gaynor, Edward. 1994. "Cataloging Electronic Texts: The University of Virginia Library Experience." Library Resources and Technical Services 38(4): 403-413 (October 1994).

---

d-Lib forum d-Lib magazine

---

hdl:cnri.dlib/july95-weibel

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.

Background

A good introduction to URNs is Internet RFC 1737, "Functional Requirements for Uniform Resource Names", by Karen Sollins and Larry Masinter, December 1994. The following is an extract from their introduction. It describes the function of URNs and, in particular, how they differ from the Uniform Resource Locators (URL) used by the World Wide Web.

> "A URN identifies a resource or unit of information. It may identify, for example, intellectual content, a particular presentation of intellectual content, or whatever a name assignment authority determines is a distinctly namable entity. A URL identifies the location or a container for an instance of a resource identified by a URN. The resource identified by a URN may reside in one or more locations at any given time, may move, or may not be available at all. Of course, not all resources will move during their lifetimes, and not all resources, although identifiable and identified by a URN will be instantiated at any given time. As such a URL is identifying a place where a resource may reside, or a container, as distinct from the resource itself identified by the URN."

The RFC concentrates on the relationship between a locator (URL) and a persistent name (URN), but naming questions arise in many other contexts. For example, the Resource Cataloging and Distribution System (RCDS), developed in the Computer Science department of the University of Tennessee, uses URNs to support cataloging, replication and caching (for high availability and fault-tolerance), and authenticity and integrity assurances using digital signatures. The paper "A Framework for Distributed Digital Object Services" by Robert Kahn and Robert Wilensky, May 1995, also identifies persistent names assigned to objects in repositories as a key component of a framework to manage intellectual property on networks.

A class of names with some characteristics similar to URNs are the domain names (such as "andrew.cmu.edu"), used to identify computer systems on the Internet. Domain names are supported by a well-tuned computer system, the Domain Name System (DNS). Several URN implementations build on domain names and DNS.

URN Requirements

RFC 1737 lays out functional requirements for URNs. It also makes recommendations about the form that such names might take. An updated version of RFC 1737 is under discussion, but, with some important clarifications, the following list of requirements has been widely accepted.

> "**Global scope**: A URN is a name with global scope which does not imply a location. It has the same meaning everywhere.
>
> "**Global uniqueness**: The same URN will never be assigned to two different resources.
>
> "**Persistence**: It is intended that the lifetime of a URN be permanent. That is, the URN will be globally unique forever, and may well be used as a reference to a resource well beyond the lifetime of the resource it identifies or of any naming authority involved in the assignment of its name.
>
> "**Scalability**: URNs can be assigned to any resource that might conceivably be available on the network, for hundreds of years.
>
> "**Legacy support**: The scheme must permit the support of existing legacy naming systems, insofar as they satisfy the other requirements described here. ...
>
> "**Extensibility**: Any scheme for URNs must permit future extensions to the scheme.
>
> "**Independence**: It is solely the responsibility of a name issuing authority to determine the

conditions under which it will issue a name."

Notice that these requirements focus on the URN, but make no assertions about the resource that it identifies. A URN may be globally unique and last for ever without any guarantee that the resource identified by the URN is unique or permanent.

Resolution

To use a URN, there must be a network-accessible service that can map the name onto the corresponding resource. This process is called **resolution**.

Frequently, the resolution system will return the current location of the resource or a list of locations. RFC 1737 concentrates on the case of a URN that resolves to a URL, but a URN can resolve to any network resource or service. For example, in RCDS, a URN may resolve to one or more location-independent file names (LIFNs), which can themselves be considered a specific type of URN. In the Kahn/Wilensky model a URN, known as a "handle", resolves to the name of the repository that holds the resource. In other contexts, a URN may resolve to a data structure containing meta-information about the resource.

The URN Framework

This section describes the URN framework that has emerged from the discussions of the past few months. Although many details remain, the level of agreement is promising.

General Principles

- **Naming schemes and resolution systems.** The framework distinguishes between naming schemes and resolution systems. A naming scheme is a procedure for creating and assigning unique URNs that conform to a specified syntax. A resolution system is a network-accessible service that stores URNs and resolves them.

- **Independence between naming schemes and resolution systems.** A naming scheme is not tied to a specific resolution system. Any resolution systems is potentially capable of resolving a URN from any given name scheme.

- **URN registries.** Since naming schemes and resolution systems are conceptually independent, mechanisms must be created so that the user of a URN can discover what resolution systems are able to resolve the URN. This is called a URN registry or simply a registry.

Multiple independent naming schemes and resolution systems are anticipated. Although the maintainer of a particular URN resolution system may also wish to maintain a registry, it is important to realize that registries and URN schemes are conceptually independent of one another. Any registry is capable of registering resolution services for any URN scheme, and a client may wish to consult multiple registries when attempting to resolve a name.

Syntax

The URN implementors have agreed on the following syntax, with one outstanding difference of opinion; opinions differ whether the leading characters "urn:" should be part of the name. This syntax is acceptable in all proposed naming schemes and resolution systems. There are many details that need to be discussed (for example the precise character sets allowed in URNs).

The following are examples of URNs:

urn:hdl:cnri.dlib/august95
urn:lifn:some.domain:anything-goes-here

urn:path:/A/B/C/doc.html
urn:inet:library.bigstate.edu:aj17-mcc *{Correction to this entry made with permission from the authors. Ed., 2/19/96.}*

Notice that the syntax of a URN explicitly indicates the naming scheme, by including a naming scheme identifier, "hdl", "lifn", "path", "inet", etc. This is followed by a colon and a string that has a syntax defined by the specific naming scheme.

As can be seen from the examples, the different naming schemes use different formats. Some naming schemes divide the name into two parts, a naming authority followed by a unique string, which is assigned by the naming authority. Thus the handle "cnri.dlib/august95" consists of a naming authority, "cnri.dlib" followed by a unique string, "august95". The path URN "/A/B/C/doc.html" consists of a naming authority (or path), "/A/B/C", and a unique string, "doc.html".

The Internet community is developing a general framework of uniform resource identification (URIs), of which URNs are a component. The URI framework was originally outlined in RFC 1630. Under the proposed framework, each participating naming schemes is a URI as defined in the RFC.

Management of Naming Schemes

The long term value of URNs requires the naming schemes to be well managed. Initially, a small number of schemes are under development. Hopefully, a small number of high quality naming schemes will be added in the future.

The criteria for an acceptable URN scheme will be outlined more formally as the URN framework is defined. They are likely to include a requirement that each naming scheme must have a verifiable management system to ensure the integrity of the naming scheme and of the URNs within it. This includes the process for assigning unique URNs within the naming scheme. It must also make sure that there is at least one resolution system able to resolve the names.

Those URN schemes that include naming authorities (e.g., handles, paths) will determine the names of the authority names themselves. Thus, it is possible that different organizations may get the same naming authority string under different naming schemes.

URN Registries

A URN registry is a network service that stores data about URN naming schemes, naming authorities, and resolution systems. A registry provides two types of service. It may provide rules for extracting the naming authority from URNs in a particular naming scheme. In this case, the first step of the URN resolution service may be to provide information on how to find the naming authority in the URN string. The second function is to know which resolution systems are capable of resolving a given URN, from the name scheme and, when appropriate, the naming authority.

The concepts of URN registries and resolution systems are not tied to any specific computing system or set of software. This is important since URNs are intended to be valid for long periods of time, much longer than any computer system can be expected to last. The format of data to be stored in a registry is currently under development. It has been given the working name NAPTR ("Naming Authority PoinTeR"). In practice, it is probable that several URN resolution systems will include URN registries, but every registry need not hold full information for all naming schemes. One proposed implementation is a modified version of DNS. Another uses the handle system.

Flexibility within the URN Naming Schemes and Resolution Systems

This report emphasizes the areas where the various URN developments are converging on a common framework. In a number of key areas, the URN implementors have carefully agreed to support flexibility

rather than to enforce unnecessary conformity.

The value of a naming scheme or a resolution system depends upon a number of assertions. Are the names unique? Can a resource have many names? Can it change? Is it guaranteed to exist? What is the retention scheme? Does a URN resolve to untyped data, typed data, entity-attribute pairs, a URL, the address of a repository, etc.? Within the general URN framework, such assertions about names, semantic decisions, and management issues may be enforced by the naming scheme or the resolution system, or they may be left to external systems. Variations in these important areas will give the schemes their distinctive features and will determine which are most suitable for specific application areas. The objective of the URN framework is to encourage wide flexibility within a stable system of naming and resolution.

URN Implementors

The following projects were represented at the University of Tennessee meeting in October 1995 and have continued to work together to reach agreement on the URN framework.

Resource Cataloging and Distribution Service (RCDS)
: This work is led by Keith Moore, Shirley Browne, Stan Green and Reed Wade of the University of Tennessee. Its aim is to provide transparent replication along with integrity/authenticity assurances, and alleviate the problem of huge demand for some random network resource.

The Handle System
: This work is led by David Ely and William Arms of the Corporation of National Research Initiatives. It is based on the ideas in the Kahn/Wilensky framework.

x-dns-2
: This is a scheme developed by Paul E. Hoffman of Proper Publishing and Ron Daniel, Jr. of Los Alamos National Laboratory. As the name implies it is based on the Internet domain name system (DNS).

URN Services
: This is a proposal by Keith E. Shafer, Eric J. Miller, Vincent M. Tkac, and Stuart L. Weibel of OCLC. It focuses on the syntax and functions of URNs.

Path URN
: This is another scheme that make use of DNS. It has been developed by Dan LaLiberte and Michael Shapiro at the National Center for Supercomputing Applications.

Whois++
: Several groups are working towards using Whois++ as an Internet Directory Service. Work done by Michael Mealling of Georgia Tech and Patrik Faltstrom and Leslie Daigle of Bunyip Information Systems, Inc., focuses on the distribution of URN resolution data and maintenance responsibility in a global publishing environment.

Contributors to this report

The following URN implementors contributed to this report: William Arms (CNRI), Leslie Daigle (Bunyip), Ron Daniel (Los Alamos National Laboratory), Dan LaLiberte (NCSA), Michael Mealling (Georgia Institute of Technology), Keith Moore (University of Tennessee), and Stuart Weibel (OCLC).

Home Magazine Comments
NEXT ▶

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,

however, are complicated by the newness of the systems, their ability to integrate a range of functions that were previously designed and evaluated separately, the heterogeneity of their user population, the physically distributed nature of usage, the ability to fragment and rearrange previously integrated documents and images, and the rapid versioning of digital objects. Appropriate user-centered research objectives, measures, and methods for the digital library are just beginning to emerge.

Results from user studies can help digital library designers and policymakers formulate appropriate goals, arrive at a more complete understanding of costs and benefits, design and allocate resources to both technologies and programs that offer the best means of achieving goals, and assess the degree to which network policies and programs have achieved their stated goals. Granted, determining (let alone predicting) impacts from information technology at the individual, organizational, and societal levels is notoriously difficult. But without such investigations, the views, needs, and experiences of individual information creators and consumers will be lost in the push and pull of constituencies with more powerful and direct voices in the process of building digital libraries, a process in which users themselves are all too apt to be considered mere passive consumers in the technology-implementation chain.

How can we learn more about the use and users of digital libraries? How can people involved in user-centered studies associated with the vastly different kinds of digital library initiatives described above share their ideas, concerns, methodological approaches, and findings? I would like to turn now to describing digital library user research, and mechanisms for sharing that research, that I have become involved with as a participant in the NSF/ARPA/NASA Digital Library Initiative.

Synchronizing Work Across the Six DLI Projects: The Role of the User Research Working Group

With encouragement from project sponsors, we have established an informal DLI-wide User Research Working Group. The motivation for the working group stems from our sense that the six DLI projects are similar enough that we can learn from each others' experiences. In addition, we have found that each user research group has different strengths. While our group at University of Illinois, for example, is especially strong in ethnographic approaches to studying system use, other groups have had more experience with conducting usability tests and designing system instrumentation. Common problems include the need to develop new methods for tracking distributed "virtual" users, difficulties in integrating and making sense of data vrom various quantitative and qualitative sources, and dealing with a large and heterogeneous user population.

For this first year, our interactions have been somewhat limited in scope and informal in nature. We get together twice a year at the DLI synchronization meetings and have set up a mailing list for working group members. At the Spring 1995 meeting, we discussed our basic approaches to user evaluation and the recognition that evaluation can proceed at different levels, to reach different goals. Summarizing our discussion, Karen Drabenstott of the University of Michigan suggested the following schema for evaluation levels:

- Adequacy of the collection, functionality, interface, usability.
- Search and retrieval performance and behavior.
- Effect on work of users, fundamental changes in processes
- Public policy implications

It was clear that the six projects are devoting varying amounts of attention to each of these evaluation levels.

We also realized that there were a number of unresolved issues confronting virtually all the user research groups. We discussed the way, for example, in which the new phenomena engendered by digital library technology lead to exploring unfamiliar methods and conceptual realms. Another major dilemma we all faced was figuring out how best to produce useful results for our system designers. Problems arose in juggling conflicting schedules, maintaining effective communication, and knowing how to make our findings operational. We concluded that there were new pulls on both system designers (new ways of thinking about

use and users) and social scientists (new approaches to studying systems), so that it was important to try to keep dialogue open among users, sponsors, social scientists, and computer scientists.

We agreed that members of our cross-project working group would present a status report on goals, methods, results, and problems for each synchronization meeting. We also agreed that we would try to facilitate cross-team sharing through posting our working papers, including instruments, on our project home pages, and that members from each project would complete brief user research "templates" (see Figure 1) to describe their work.

Figure 1: DLI User Research Template

- 1) Capsule summary (3-5 sentences) of the DLI project itself (i.e., what kind of DL your project team is building)
- 2) Age and experience levels of the users you're studying
- 3) Evaluation "levels" you're addressing in your user research:
    - a. Adequacy of the collection, functionality, interface, usability
    - b. Search and retrieval performance and behavior
    - c. Effect on iwork of users, fundamental changes in processes
    - d. Implications for public policy
- 4) Use settings you're studying (e.g., laboratories, public libraries, high schools)
- 5) Methods you're using
- 6) Theoretical considerations of particular import in your work

---

Based on information provided by members of each DLI project, I have prepared capsule descriptions of each project's user research efforts (http://anshar.grainger.uiuc.edu/dlisoc/home_page.html/user_research_wg).

Building Understanding Beyond the DLI Projects: The Allerton Institute

One recommendation that arose from the User Research Working Group at the Fall 1994 DLI synchronization meeting was to find a way that we could get together with other interested researchers to explore methodological approaches associated with the use of digital libraries. This recommendation has been realized in the 1995 Allerton Institute conducted by the Graduate School of Library and Information Science at the University of Illinois, titled "How We Do User-Centered Design and Evaluation of Digital Libraries: A Methodological Forum." The Institute, sponsored by the National Science Foundation (NSF), will be held on October 29-31.

As chairperson, my goal is to bring together an interdisciplinary group of researchers and practitioners involved in the design and study of information systems, in user-centered research in traditional libraries, and in a wide range of digital library projects. The purpose of the forum is to present both the range of user-centered methods available for studying digital libraries and rationales for choosing amongst them; we also want to look ahead to new methods and developments and map out the challenges that lie ahead. This methodological forum will give the 60 invited participants an opportunity to share their expertise, experiences, and ideas with their peers in a relaxed environment. Forum activities will be devoted to issues such as:

- What are appropriate measures for gauging digital library outcomes at the individual, group, institutional, and global levels?
- How can we best incorporate knowledge of user needs and behavior in designing digital library interactions and interfaces?
- What do we need to know about how people use electronic texts and how can we gain this knowledge and apply it to the development of digital libraries?
- What can we learn from studies of traditional library use?
- How can we develop an understanding of the computerization of library work that will help as digital

systems are incorporated into current institutional practices?
- How can we deal with the ethical, practical, and conceptual issues that arise in the remote observation of on-line (and off-line) behavior on a very large scale?
- How do we foster effective communication among digital library designers, users, and social science researchers?

Each participant submitted a brief discussion document outlining their work and the issues they were most eager to explore. These papers were used to develop the five major Institute sessions, which will focus on co-design approaches, work practice and institutional change, migrating foundational approaches to virtual library environment, electronic information seeking behavior, and understanding diversity and change. Participants include researchers from the fields of Computer Science, Sociology, Library and Information Science, Education, and Psychology who are involved in digital library projects in a wide range of settings. Presentations will be given by a number of participants, including Michael Twidale, Annelise Mark Pejtersen, William L. Anderson and Susan L. Anderson, F. W. Lancaster, Andrew Dillon, John M. Carroll, Brenda Dervin, Rob Kling, Chip Bruce, and Gary Marchionini. Discussion documents from participants, plus perhaps other related material from the Institute, will be made publicly available at some point after the Institute.

I hope that the user research efforts of the DLI projects, along with the ideas arising from the Allerton Institute, will help in building a framework for understanding the use and implications of digital information infrastructure, as our research methods, systems, and expectations of systems continually evolve. By situating the study of DLI testbed use within the broader context of professional work and social practices, I believe we will gain more robust insights into the functions and features that will make digital libraries more effective. In addition, we will get a sense of large-scale changes in work and cognition that occur as the nation's entire information infrastructure begins to change.

References

- Bishop, A. P., & Bishop, C. M. (1995). The policy role of user studies. **Serials Review**, 21(1), 17-25.
- Buttenfield, B. P. (1995, draft). User evaluation for the Alexandria Digital Library Project. [Discussion document submitted for the 1995 Allerton Institute: "How We Do User- Based Design and Evaluation for Digital Libraries: A Methodological Forum"].
- Digital libraries [Special issue]. (1995). **Communications of the ACM**, 38(4).
- Finholt, Thomas A. (1995, draft). Understanding digital libraries: Possible lessons from the analysis of collaboratories. [Discussion document submitted for the 1995 Allerton Institute: "How We Do User-Based Design and Evaluation for Digital Libraries: A Methodological Forum"].
- Gaston, B. (1994, Sept. 27). NSF announces awards for digital libraries research. Washington, DC: National Science Foundation. (Available: http://www.grainger.uiuc.edu/dli/release.htm)
- Lasher, R. (1994, Oct. 11). Library issues for the Joint Initiative Digital Library Projects. Unpublished manuscript. (Available: http://www-diglib.stanford.edu/diglib/pub/dllibrary/library-issues.html)
- Reich, V. (1995, draft). Allerton discussion document. [Discussion document submitted for the 1995 Allerton Institute: "How We Do User-Based Design and Evaluation for Digital Libraries: A Methodological Forum"].
- Reich, V., & Weiser, M. (1994). Libraries are more than information: Situational aspects of electronic libraries. Serials Review, 20(3), 31-37. (Available: http://www.ubiq.com/hypertext/weiser/SituationalAspectsofElectronicLibraries.html)
- Van House, N. A. (1995). User needs assessment and evaluation for the UC Berkeley Electronic Environmental Library project. In F. M. Shipman, III, Richard Furuta, & David Levy (Eds.). **Proceedings of Digital Libraries '95: The second annual conference on the theory and practice of digital libraries** (pp. 71-76). College Station, TX: Texas A&M University. (Available: http://csdl.tamu.edu/DL95/papers/vanhouse/vanhouse.html)
- Van House, N. A. (1995, draft). Current project: UC Berkeley's NSF/ARPA/NASA Digital Libraries Project. [Discussion document submitted for the 1995 Allerton Institute: "How We Do User-Based Design and Evaluation for Digital Libraries: A Methodological Forum"].

- 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.

- Title Page
- Table of Contents
- Chapters 1-7 all together (1.26 M)
- Chapter 1: Future Directions in Text Analysis, Retrieval and Understanding (esp. white paper on A National Electronic Science, Engineering, and Technology Library)
- Chapter 2: July 1992 Workshop
- Chapter 3: December 1992 Workshop
- Chapter 4: Notable Events
- Chapter 5: Directory of Interested Parties
- Chapter 6: Summary and Recommendations
- Chapter 7: Glossary
- Index

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

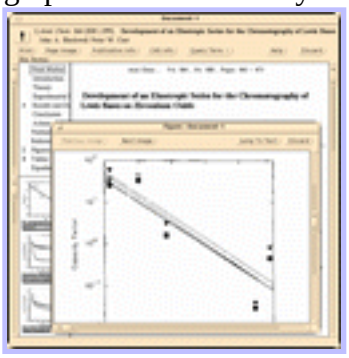
XSCEPTER provides "on-the-fly" formating of SGML as defined by configurable style guides and a DTD, to provide rapid display of scholarly data. The XSCEPTER interface is coupled with NEWTON, OCLC's proprietary search engine, to provide navigational capabilities of the CORE collection.


XSCEPTER main window


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


Result hit-list of database search


XSDOCUMENT view of article including dynamic table of content, table of graphics and "on-the-fly" rendering of SGML text.


Selection of image from table of graphics allowing hyperlink to referenced text or animation through image list.


Hyperlink of bibliographic reference to either location current document or retrieval of actual citation from database.

TULIP The University Licensing Program

When you scroll further down this page you'll find

- Introduction
- The TULIP Final report
- TULIP Newsletters
- The Journal Titles in TULIP
- The Universities involved in TULIP
- The Anonymous FTP facility for TULIP
- Contact information

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

The TULIP Final report

The final report for the TULIP project is currently available in draft version only.

A note on printing this report

For easy printing of the entire report by your Webbrowser, it is divided in three files

1. The top level document including the Table of Contents and the Executive Summary
2. The HTML document with all chapters I through VI This document is approximately 220 Kilobytes. The included artwork totals to approximately 380 Kilobytes.
3. The HTML document with appendices I through XI This document is approximately 35 Kilobytes. Note that this document is missing parts.

Click here to return to top of information

TULIP Newsletters

The following TULIP Newsletters are available for browsing

- TULIP Newsletter no. 6 - May 1995
- TULIP Newsletter no. 5 - September 1994
- TULIP Newsletter no. 4 - April 1994
- TULIP Newsletter no. 3 - January 1994
- TULIP Newsletter no. 2 - Augustus 1993
- TULIP Newsletter no. 1 - November 1992

Click here to return to top of information

Journals in TULIP in the field of Material Science

The participating universities have in common strength in the physical and engineering sciences. In looking within these disciplines for a target area, we wanted a field in which the researchers were comfortable with computer applications and had a higher than average installed base of workstations. An obvious choice might have been computer science itself, but we felt these users would be so atypical in their computer facility as to make it hard to generalize results to other disciplines. Materials science provided a field in which there was both a sufficiently large corpus of frequently-cited material within one publishing company and interested faculties. Therefore 83 journal titles were chosen from the collection of Elsevier Science journal titles.

Click here to return to top of information

Universities participating in TULIP

- **University of California** (all campuses)
  - Berkeley
  - Davis
  - Irvine
  - Los Angeles

    - Riverside
    - Santa Barbara
    - Santa Cruz
    - San Diego
    - San Francisco
- Carnegie Mellon University (Pittsburgh, PA)
- Cornell University (Ithaca, NY)
- Georgia Institute of Technology (Atlanta, GA)
- University of Michigan (Ann Arbor, MI)
    - A demo version of the prototype of Michigan's library system based on World Wide Web
- Massachusetts Institute of Technology (Cambridge, MA)
- University of Tennessee (Knoxville, TN)
- Virginia Polytechnic Institute and State University (Blacksburg, VA)
- University of Washington (Seattle, WA)

Click here to return to top of information

TULIP through Anonymous FTP

TULIP has a number of files available for download through Anonymous FTP. Connect to ftp.elsevier.nl, enter "anonymous" as username and enter your email address as password. Change into the /TULIP directory and read the Readme.txt file.

Click here to return to top of information

For general information on the TULIP Program

Project Manager
Jaco Zijlstra (Singapore), voice +65-434 3714, fax +65-337 2230, e-mail J.Zijlstra@Elsevier.nl

Project Leader
Karen Hunter (USA), voice +1 212-633 3787, fax +1 212-633 3764, e-mail K.Hunter@Elsevier.com

Technical Coordinator
Paul Mostert (The Netherlands), voice +31 20-485 3574, fax +31 20-485 3706, e-mail P.Mostert@Elsevier.nl

Mail Address
Elsevier Science
655 Avenue of the Americas
New York, NY 10010
United States of America

Click here to return to top of information

Last modified on March 21, 1996 by Paul Mostert

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

Introducing OCLC SiteSearch

To the Next Stage of the Electronic Library

Welcome to the **OCLC SiteSearch®** family of software products. For librarians, SiteSearch brings you a significant step closer to the dream of a virtual library--seamless integration of local and remote information resources. You enhance access to your local collection, and your library becomes the doorway to the global information environment.

For end-users, SiteSearch means one interface, one access point, and one search process--all from one desktop. Users have the most comprehensive reference system available whether they are in the library, dormitory, home or office. Read on to see how SiteSearch enables your college or university and other groups to build databases and link your local area networks, OCLC reference services, and the World Wide Web into **a customized virtual library that accommodates both print and electronic information.**

What is SiteSearch?

SiteSearch is a complete set of software tools that lets you create an integrated information home for your users with your library. It is based on client server technology, and it works for individual libraries or library groups, such as regional, state, or local consortia.

SiteSearch allows you to:

- load commercial databases locally
- bring remote databases--commercial or public--to your library
- build your own unique, local databases
- index those databases
- provide one user interface to search your databases and any remote Z39.50 compatible databases

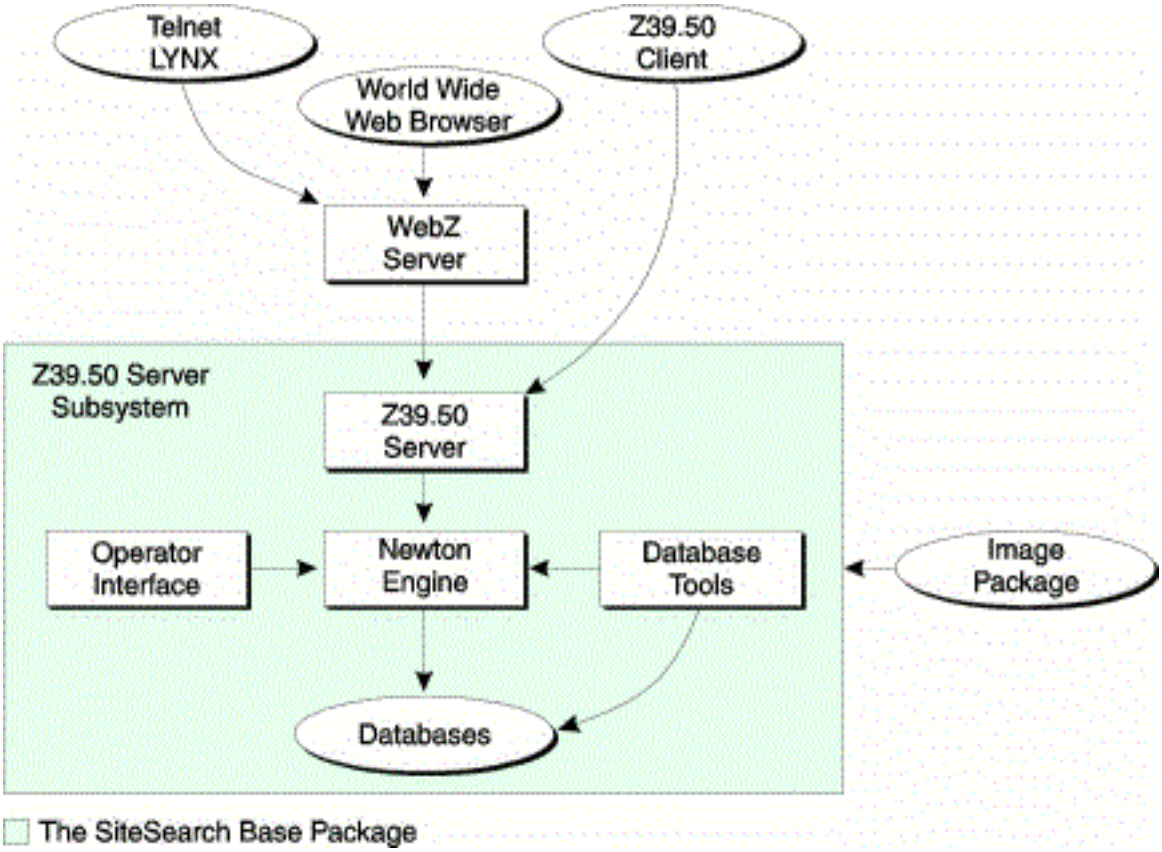
The SiteSearch software tools are:

- **Z39.50 Server System**
  - **Database loading, creation and maintenance software:** Helps you mount commercial databases locally and build and maintain unique, local databases. Lets you define indexing rules and display formats.
  - **Search engine:** Supports searching very large or very small databases of citations, full text, images, and sound with speed and efficiency. The search engine used is Newton, which is used for all of OCLC's online reference systems.
  - **Z39.50 server:** Manages connections to database servers and supports the Z39.50-1992 protocol for communication between the user interface and the search engine.
  - **Image support:** Provides tools for creating electronic image collections for applications

such as photo collections, archives and reserves.

- **Z39.50 client:** Provides a gateway to Z39.50 servers and access to OCLC SiteSearch databases for World Wide Web browsers, such as NetScape, Mosaic and Lynx. The client is called WebZ.

These components can be purchased individually or as an entire package with site licenses based on user population or simultaneous log ons.



Satisfy Your Users Information Demands

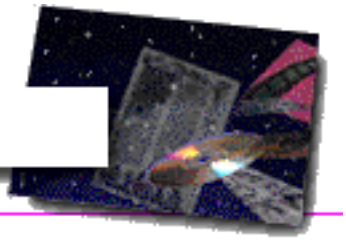
By linking resources from your library, remote sites and the World Wide Web, users will have a comprehensive and thorough information search. Their journey could take them to an electronic journal on the Internet, a database halfway around the world, or deep within your library's archives--all transparent to the user thanks to SiteSearch software.

Increase their awareness of your library's value. SiteSearch draws people to your library, both electronically and physically. Your library collection becomes more accessible, and your library becomes the starting point to the global information networks.

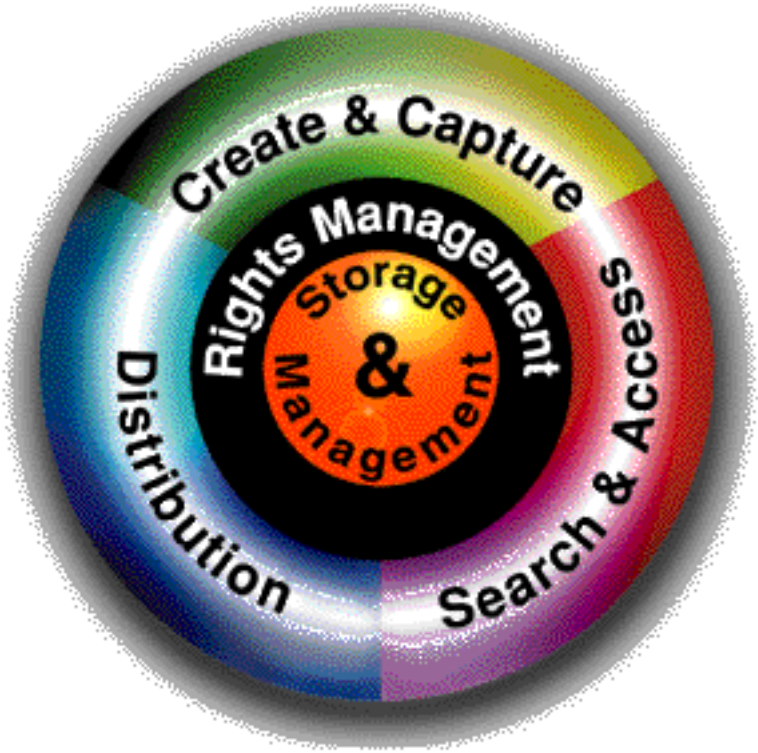
Enhance their library experience. Users will see your library as bringing some order to today's chaotic global information environment, where a myriad of systems and a mountain of information are available to them. Their confidence in their research will increase because they are searching and gathering information globally. And they can search at their convenience, wherever they are.

IBM

Software



IBM Digital Library



Case Histories

What's New

The Digital Future Is Now In Your Hands

From the boardroom to the classroom to your living room, multimedia is now a part of everyday life. The synthesis of video, audio and text on personal computers in homes and on networks is profoundly changing how people share, exchange and value information. And organizations that already own storehouses of information -- movies, music, manuscripts and more -- are tapping into vast new markets where digitized information can be available for an unlimited number of applications.

So What's Holding You Back?
Several obstacles seem to prevent content owners from becoming content providers. Digitizing existing materials like printed texts, audio recordings and reels of video or film is just the starting point. Once in digital form these materials need to be managed and maintained. Then deciding who gets access -- a few users over a local network or everybody in the world via the internet -- is a major issue. What about protecting your copyrights and intellectual property rights? And how do you provide fair compensation to intellectual property owners and providers for materials that are copied or downloaded?

Your Information. Your Assets. IBM Digital Library In Your Hands.
Pushing those obstacles out of the way just got a whole lot easier with IBM Digital Library. It's not a single product per se; you don't buy IBM Digital Library in a shrink-wrapped box. IBM Digital Library is a union of software and hardware backed by IBM intelligence and know-how. By harnessing diverse technologies and making them work as one, IBM Digital Library gives you the capability to capture assorted media in digital form, distribute multimedia content across public and private networks, search and access reams of information, manage multimedia databases of any scale, and protect your intellectual property rights. IBM

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:

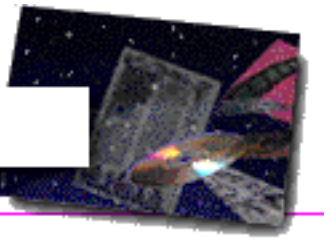
Your e-mail address:

Other links related to QBIC:

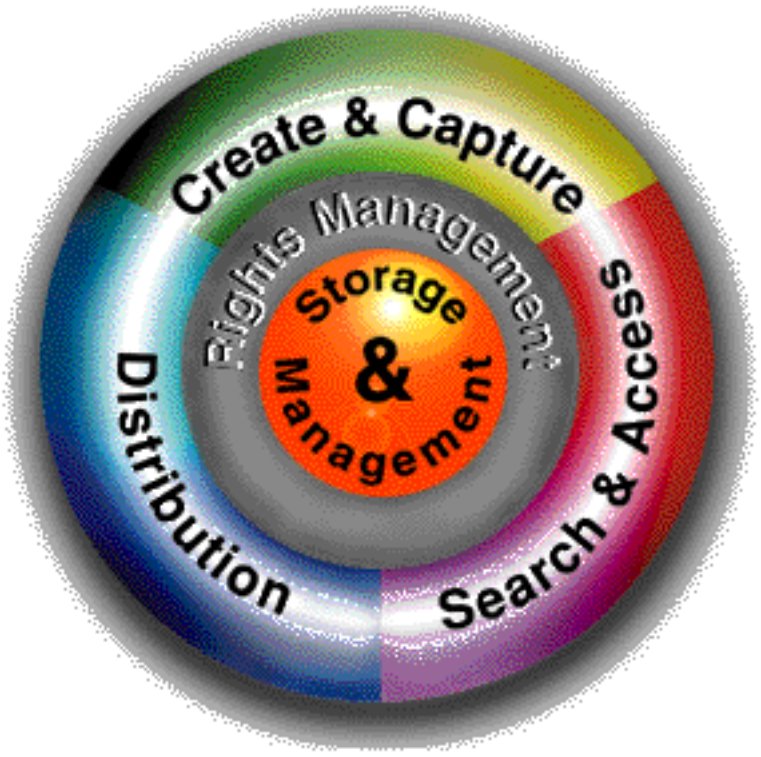
- A show and tell of the QBIC technology.
- Technical paper requests on QBIC.
- Ultimedia Manager 1.1 - A product that incorporates some of the QBIC technology.
- DB2 Extenders, which will soon include QBIC technology.

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)

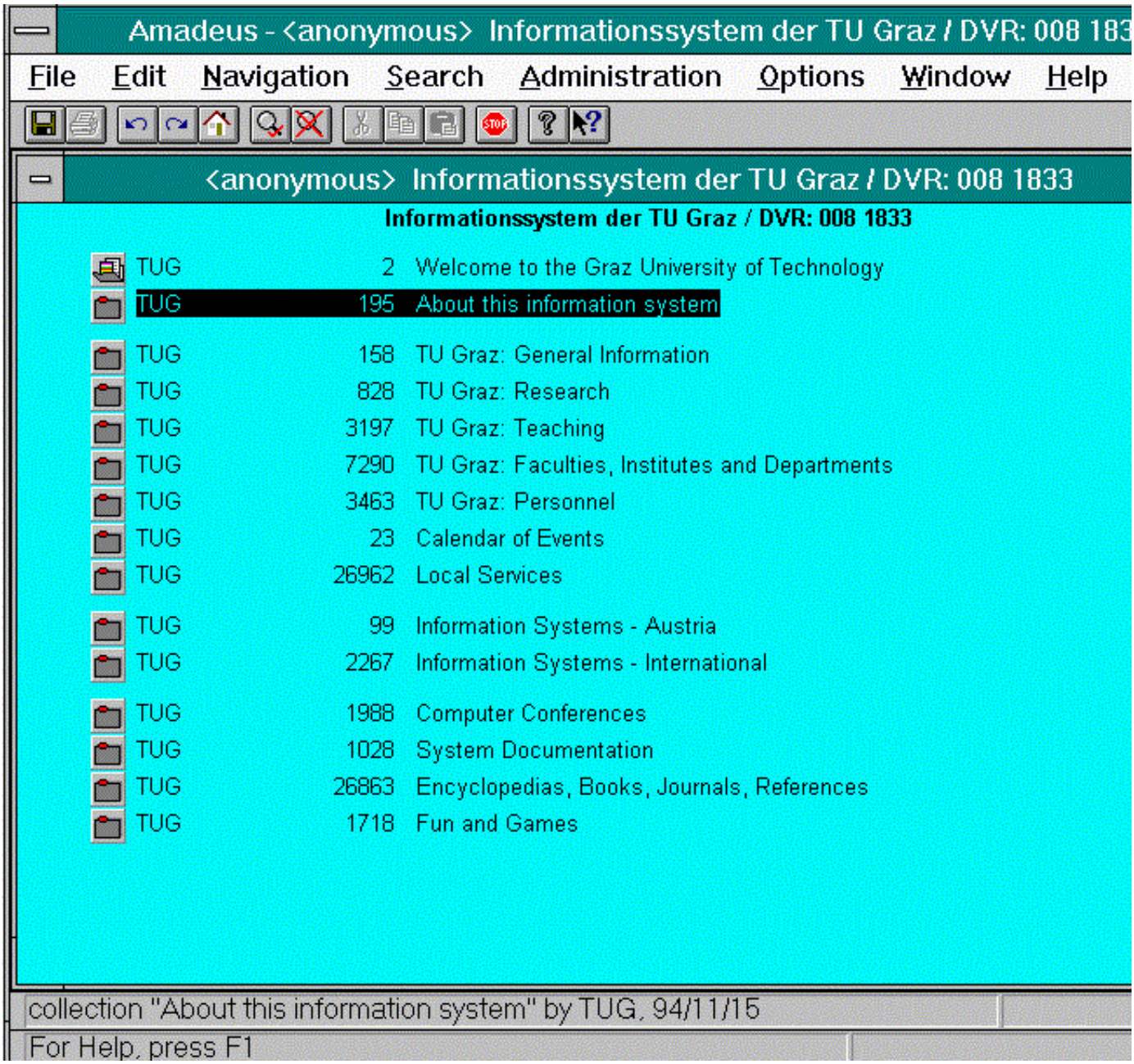
- Compatible with industry standards
  - HTML edit API tool
  - Views standard image formats (GIF, JPEG, TIFF)
  - Integrated movie (MPEG) and audio (AVI, AU) players
  - Comes with VRML (VRweb) and PostScript viewers
- Advanced navigation tools (local map, 3D landscape)
- Interactive client-server protocol 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

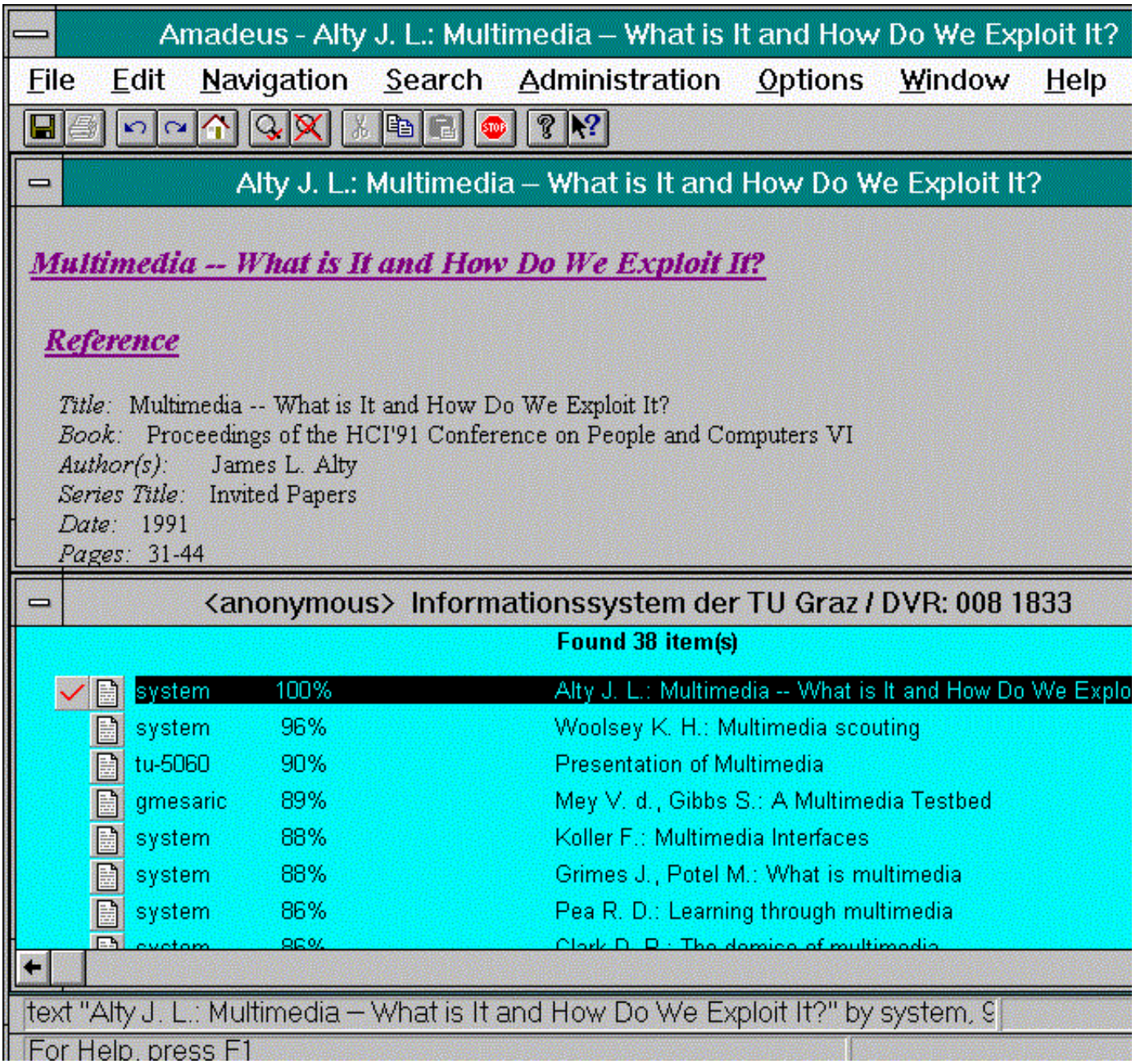
HyperWave Author for Windows (Amadeus)

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

HyperWave Author for UNIX (Harmony)


Amadeus - <anonymous> Informationssystem der TU Graz / DVR: 008 183
File Edit Navigation Search Administration Options Window Help
<anonymous> Informationssystem der TU Graz / DVR: 008 1833
Informationssystem der TU Graz / DVR: 008 1833
TUG 2 Welcome to the Graz University of Technology
TUG 195 About this information system
TUG 158 TU Graz: General Information
TUG 828 TU Graz: Research
TUG 3197 TU Graz: Teaching
TUG 7290 TU Graz: Faculties, Institutes and Departments
TUG 3463 TU Graz: Personnel
TUG 23 Calendar of Events
TUG 26962 Local Services
TUG 99 Information Systems - Austria
TUG 2267 Information Systems - International
TUG 1988 Computer Conferences
TUG 1028 System Documentation
TUG 26863 Encyclopedias, Books, Journals, References
TUG 1718 Fun and Games
collection "About this information system" by TUG, 94/11/15
For Help, press F1

- help screen


Amadeus - Alty J. L.: Multimedia – What is It and How Do We Exploit It?
File Edit Navigation Search Administration Options Window Help
Alty J. L.: Multimedia – What is It and How Do We Exploit It?
Multimedia -- What is It and How Do We Exploit It?
Reference
Title: Multimedia -- What is It and How Do We Exploit It?
Book: Proceedings of the HCI'91 Conference on People and Computers VI
Author(s): James L. Alty
Series Title: Invited Papers
Date: 1991
Pages: 31-44
<anonymous> Informationssystem der TU Graz / DVR: 008 1833
Found 38 item(s)
system 100% Alty J. L.: Multimedia -- What is It and How Do We Explo
system 96% Woolsey K. H.: Multimedia scouting
tu-5060 90% Presentation of Multimedia
gmesaric 89% Mey V. d., Gibbs S.: A Multimedia Testbed
system 88% Koller F.: Multimedia Interfaces
system 88% Grimes J., Potel M.: What is multimedia
system 86% Pea R. D.: Learning through multimedia
text "Alty J. L.: Multimedia – What is It and How Do We Exploit It?" by system, 9
For Help, press F1

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

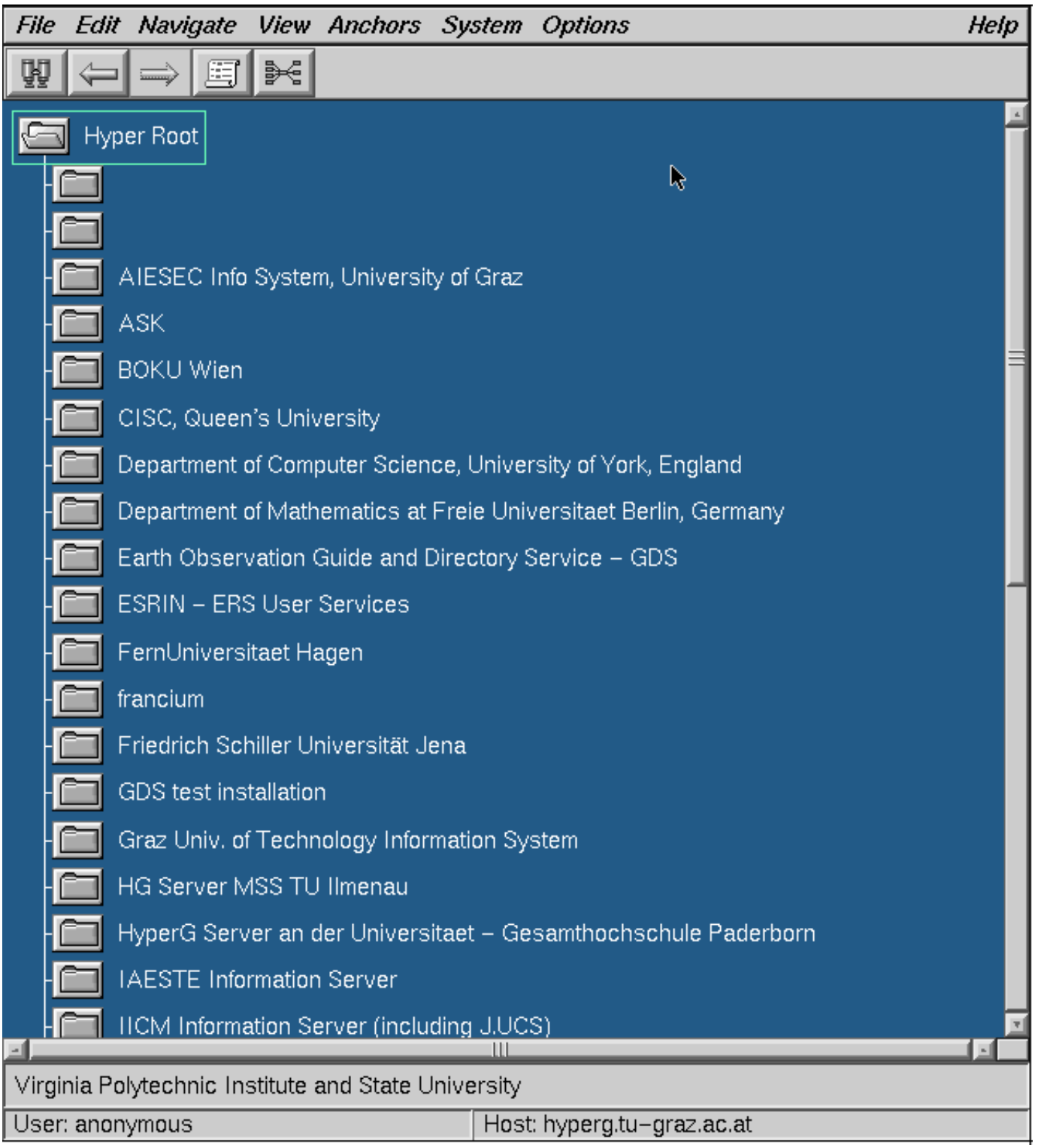
Parent(s):

Harmony's orientational aids

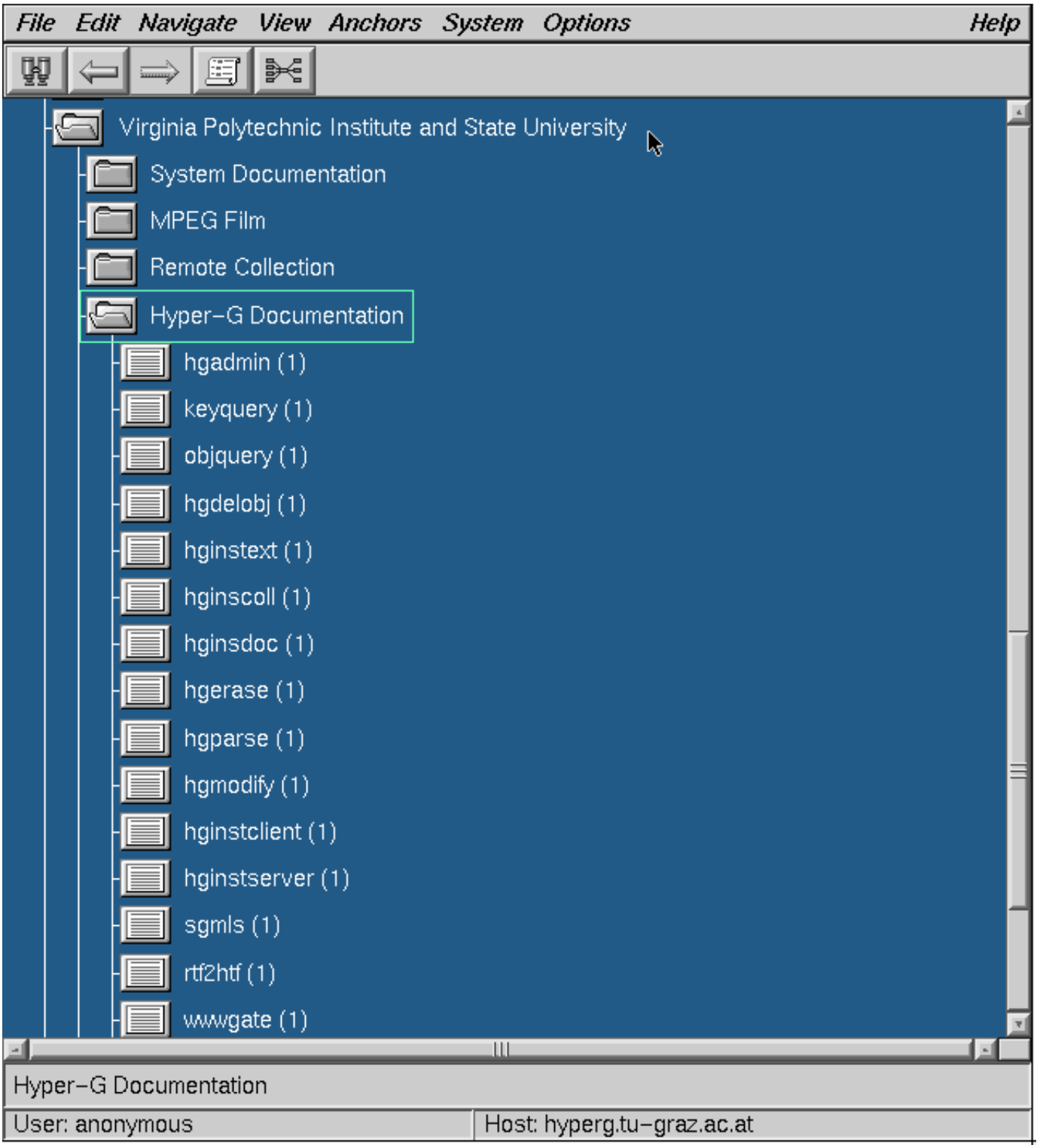
Hyper-G --- Harmony Illustrations

Illustrations of the use of Hyper-G with the Harmony (UNIX) client include:

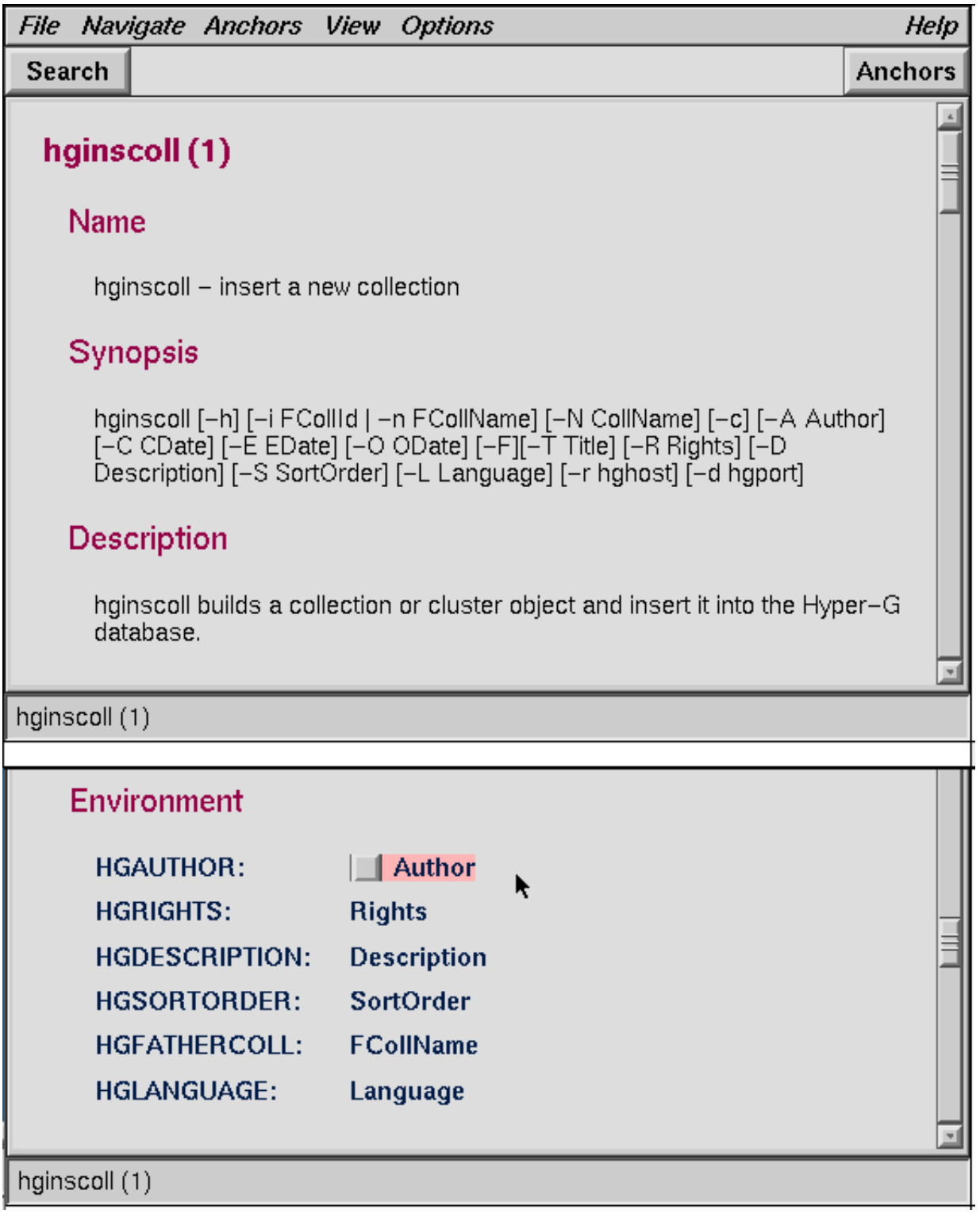
- connection to the global root



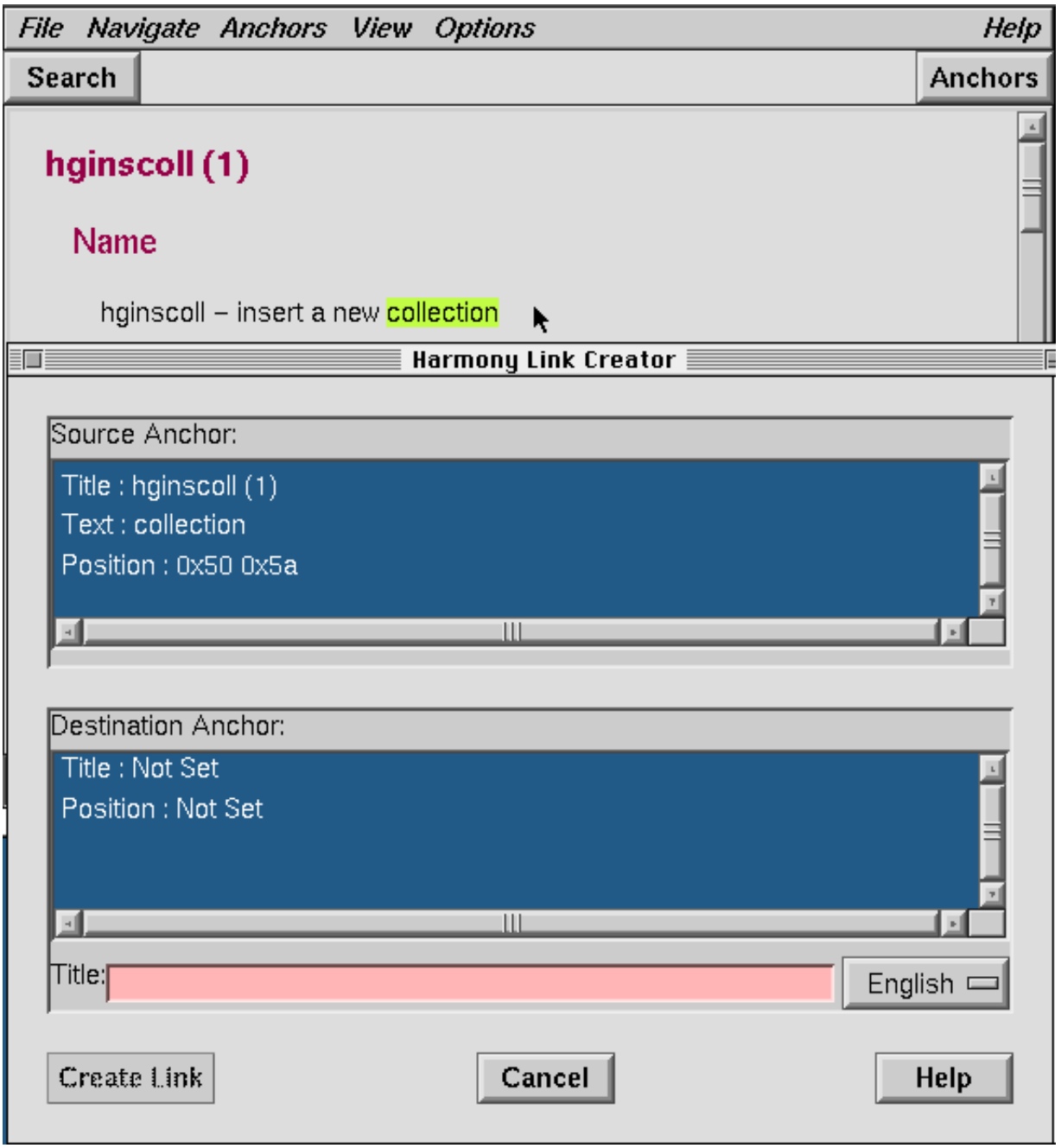
- expansion of the collection of nodes accessible from the root to those at the Virginia Tech server



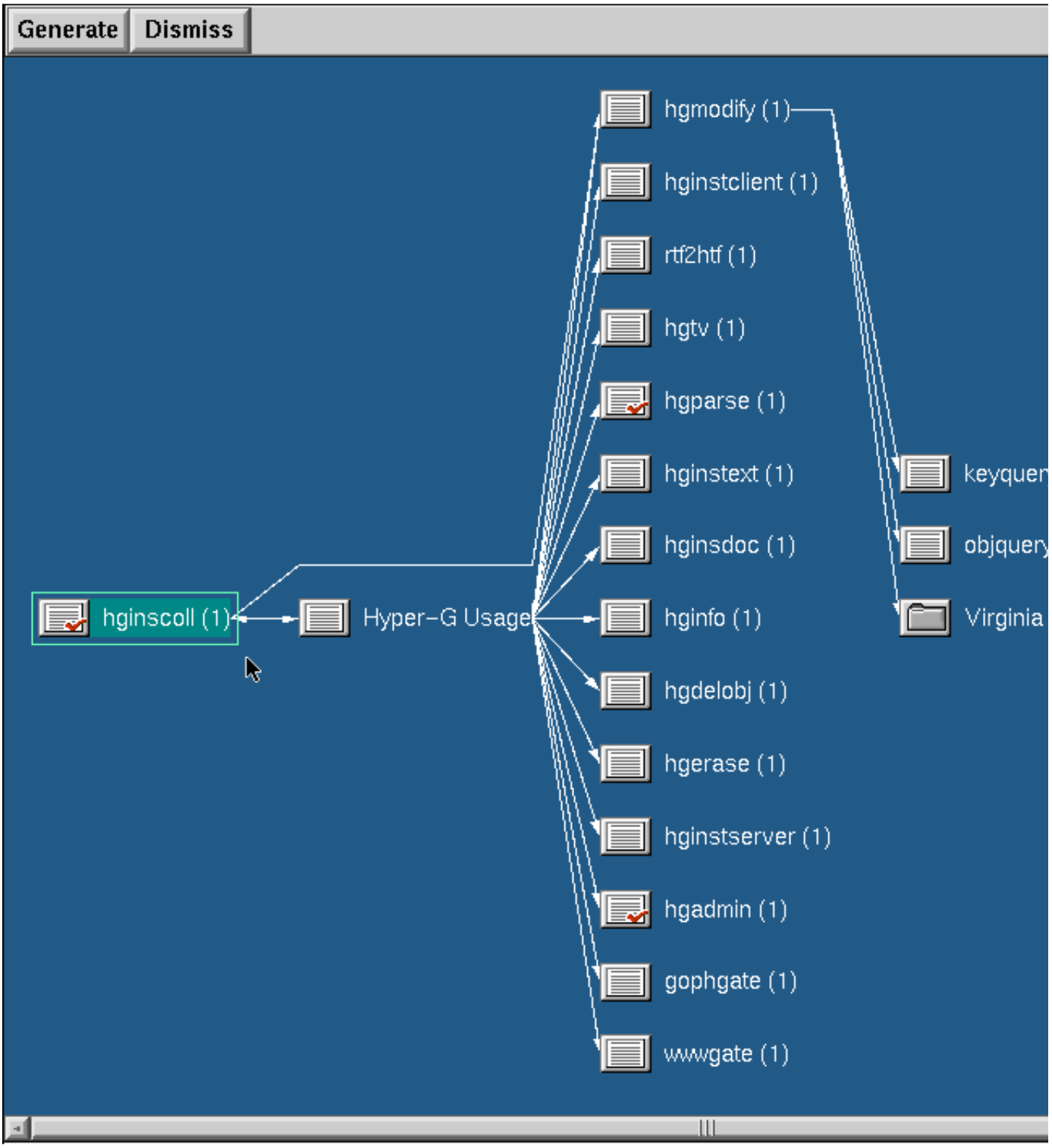
- viewing two text nodes and marking an anchor

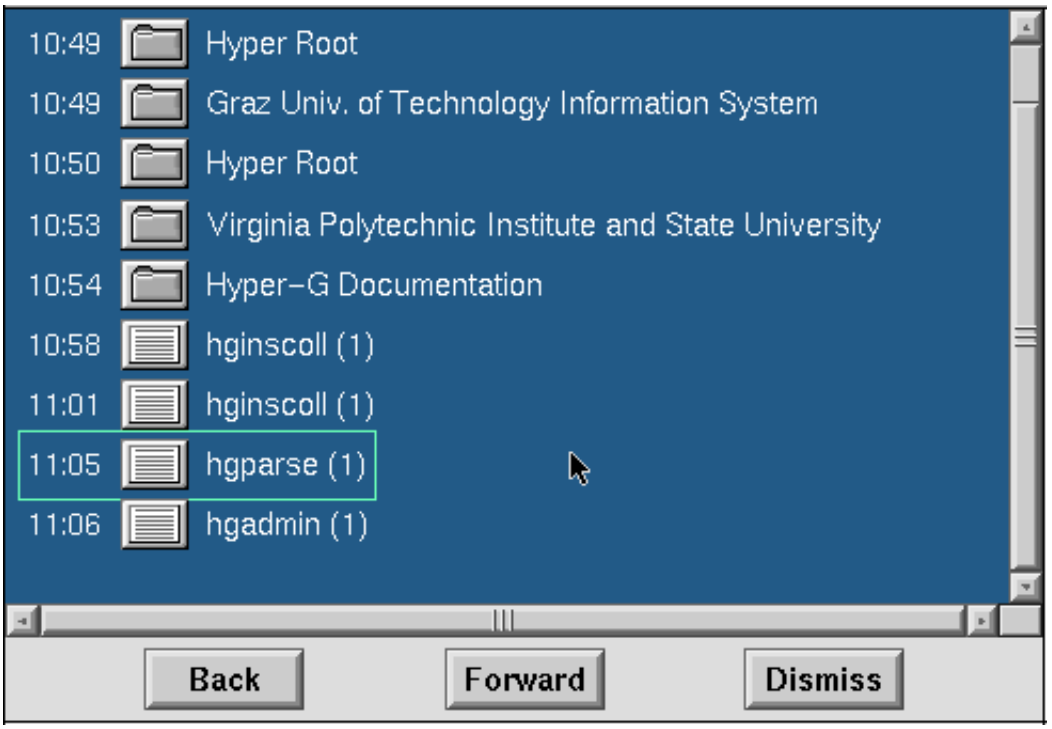


- making a link



- 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


10:49 Hyper Root
10:49 Graz Univ. of Technology Information System
10:50 Hyper Root
10:53 Virginia Polytechnic Institute and State University
10:54 Hyper-G Documentation
10:58 hginscoll (1)
11:01 hginscoll (1)
11:05 hgparse (1)
11:06 hgadmin (1)
Back
Forward
Dismiss

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

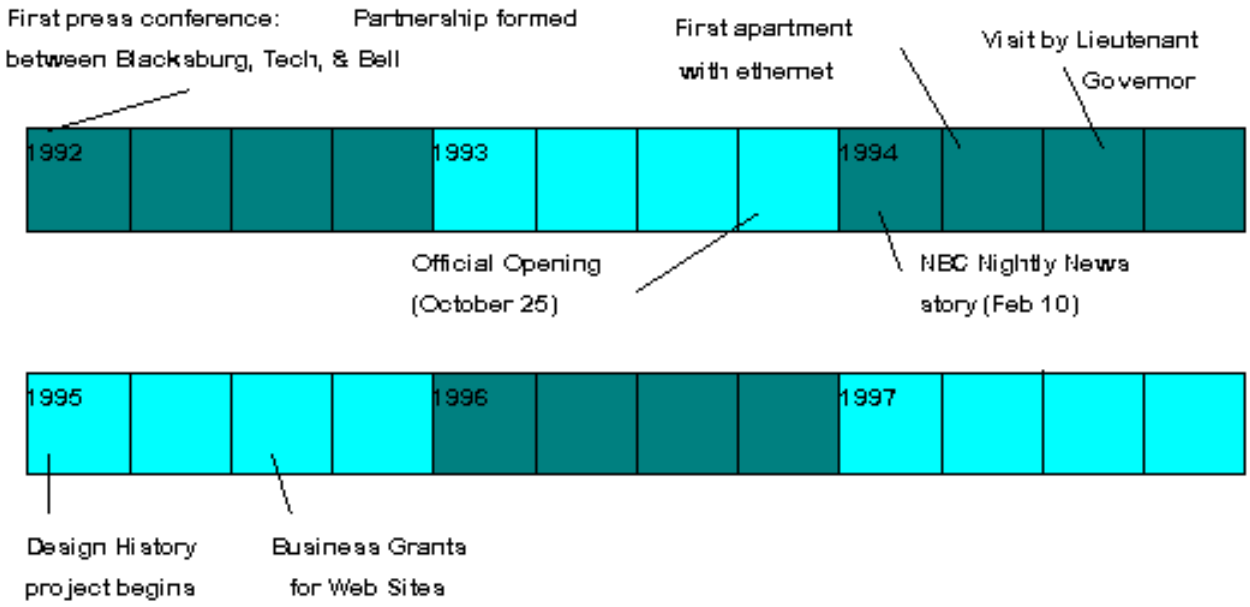
This project is supported by NSF Grant CDA-9424506. A copy of the grant proposal is online.

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

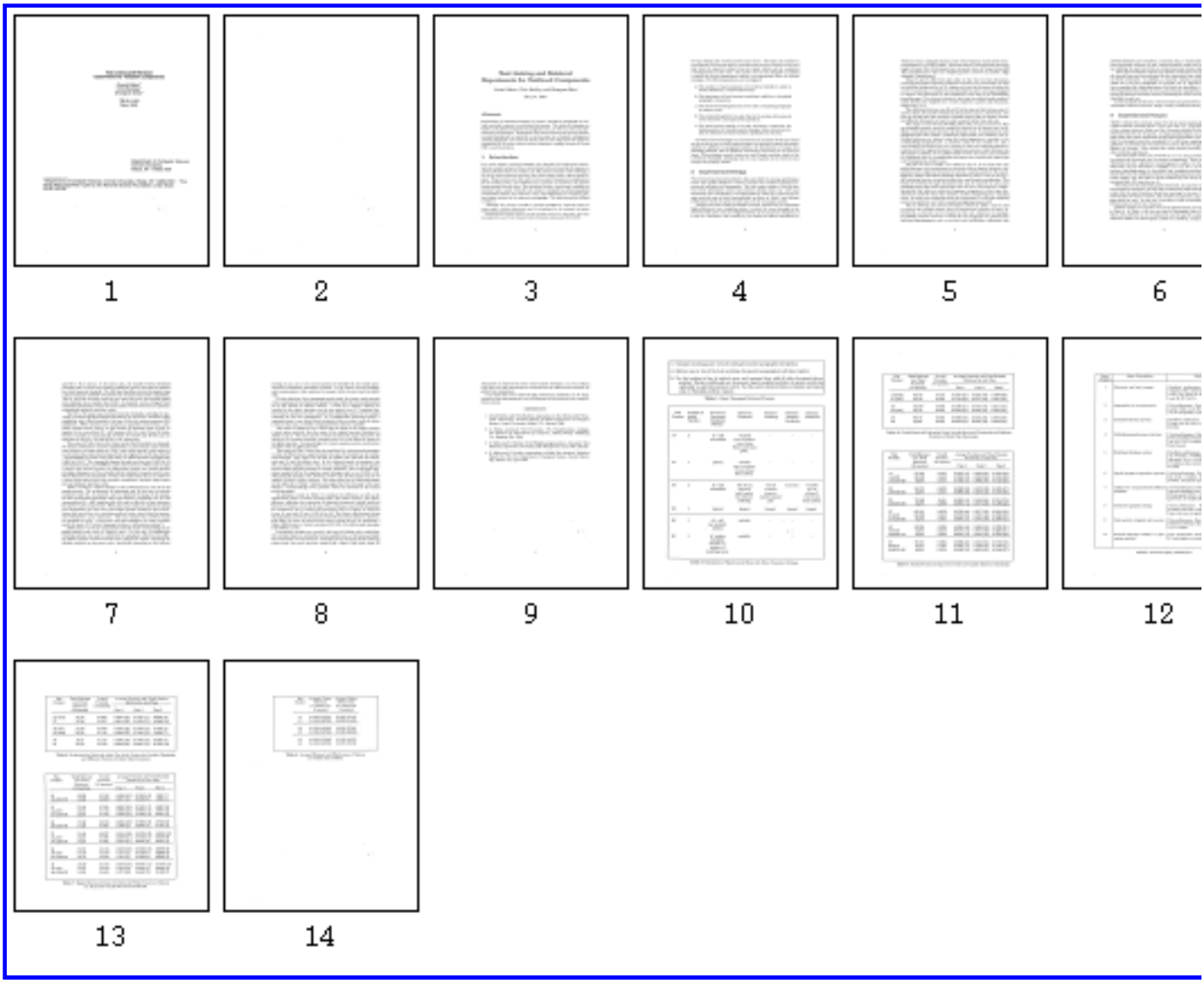
1 Display page

- 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.

[Prev section | Next section | Structure | Summary | Print/Download | Home page]

Text Linking and Retrieval Experiments for Textbook Components

Gerard Salton, Chris Buckley and Zhongnan Zhao
May 1990
Section 1 of 1. Select a page.



[Prev section | Next section | Structure | Summary | Print/Download | Home page]

Click with the mouse to zoom in on a section of the page.

Text Linking and Retrieval Experiments for Textbook Components

Gerard Salton*
Chris Buckley*
Zhongnan Zhao*

TR 90-1125
May 1990

Department of Computer Sci
Cornell University
Ithaca, NY 14853-7501

*Department of Computer Science, Cornell University, Ithaca, NY 14853-7501. T study was supported in part by the National Science Foundation under Grant IRI-87-02735.

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

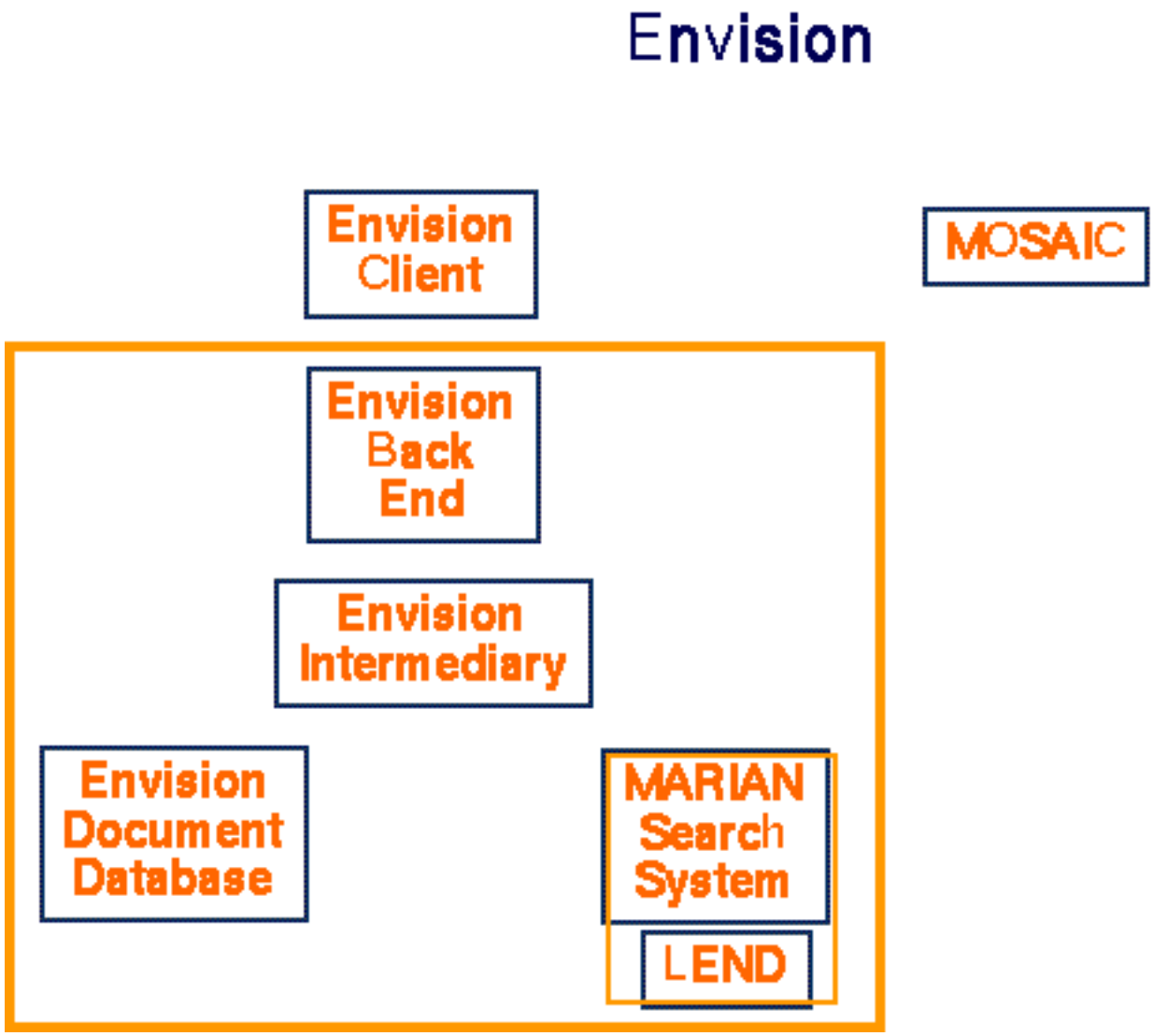
The Envision Project was funded as **A User Centered Database from the Computer Science Literature** by NSF for 1991-95. ACM has provided free access to their publications.

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

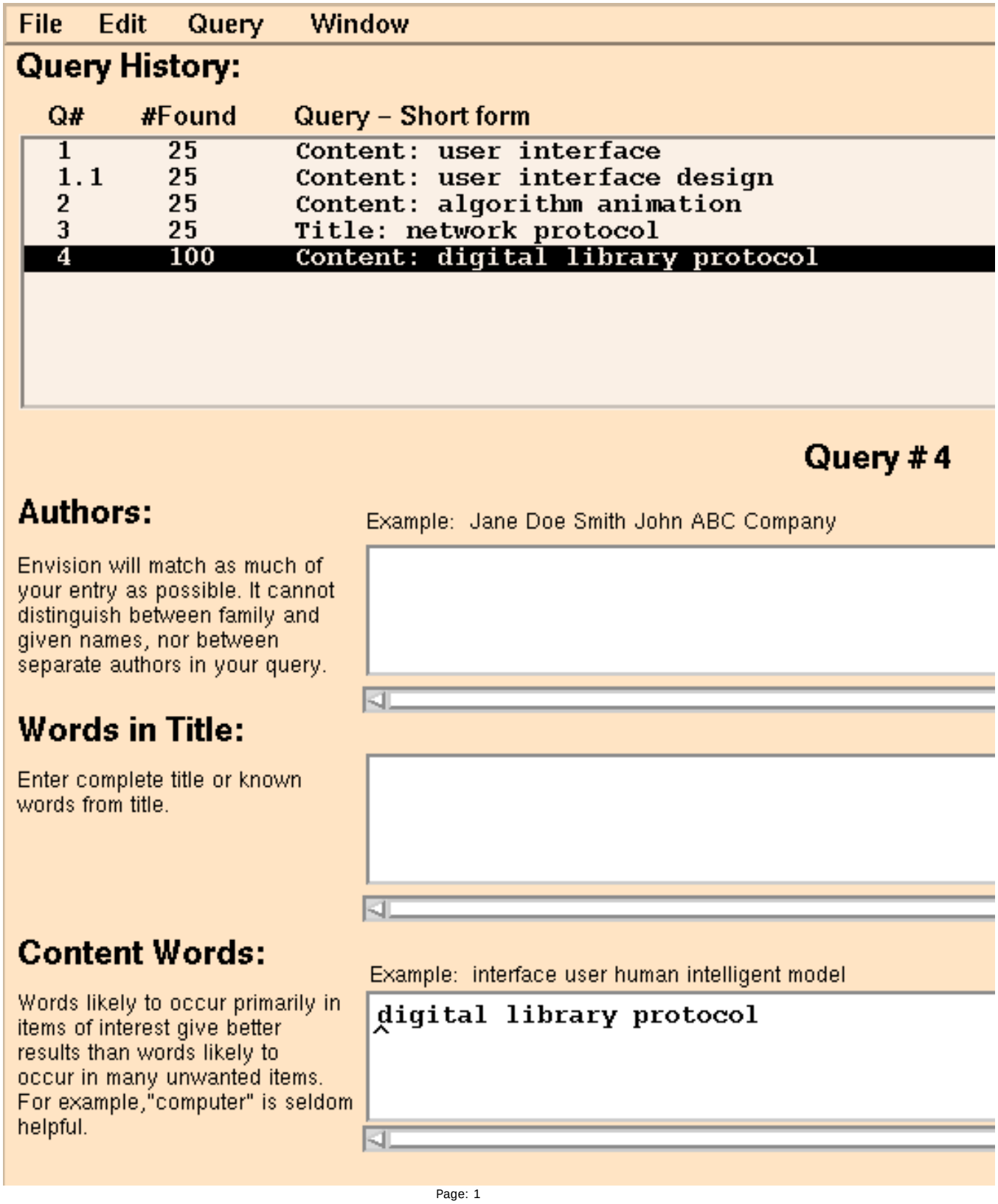
The system architecture is a combination of various elements:

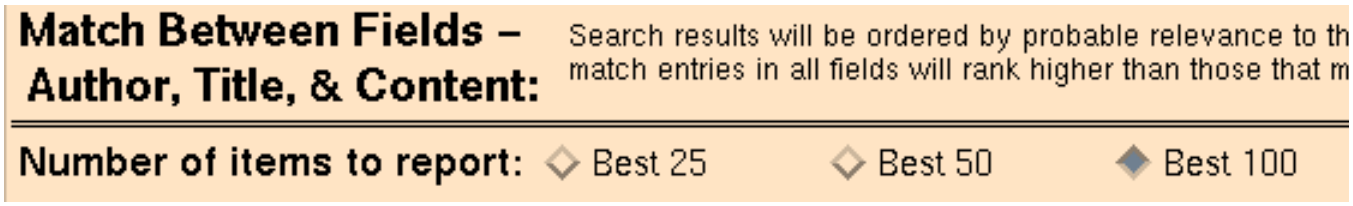


Envision - Query Screens

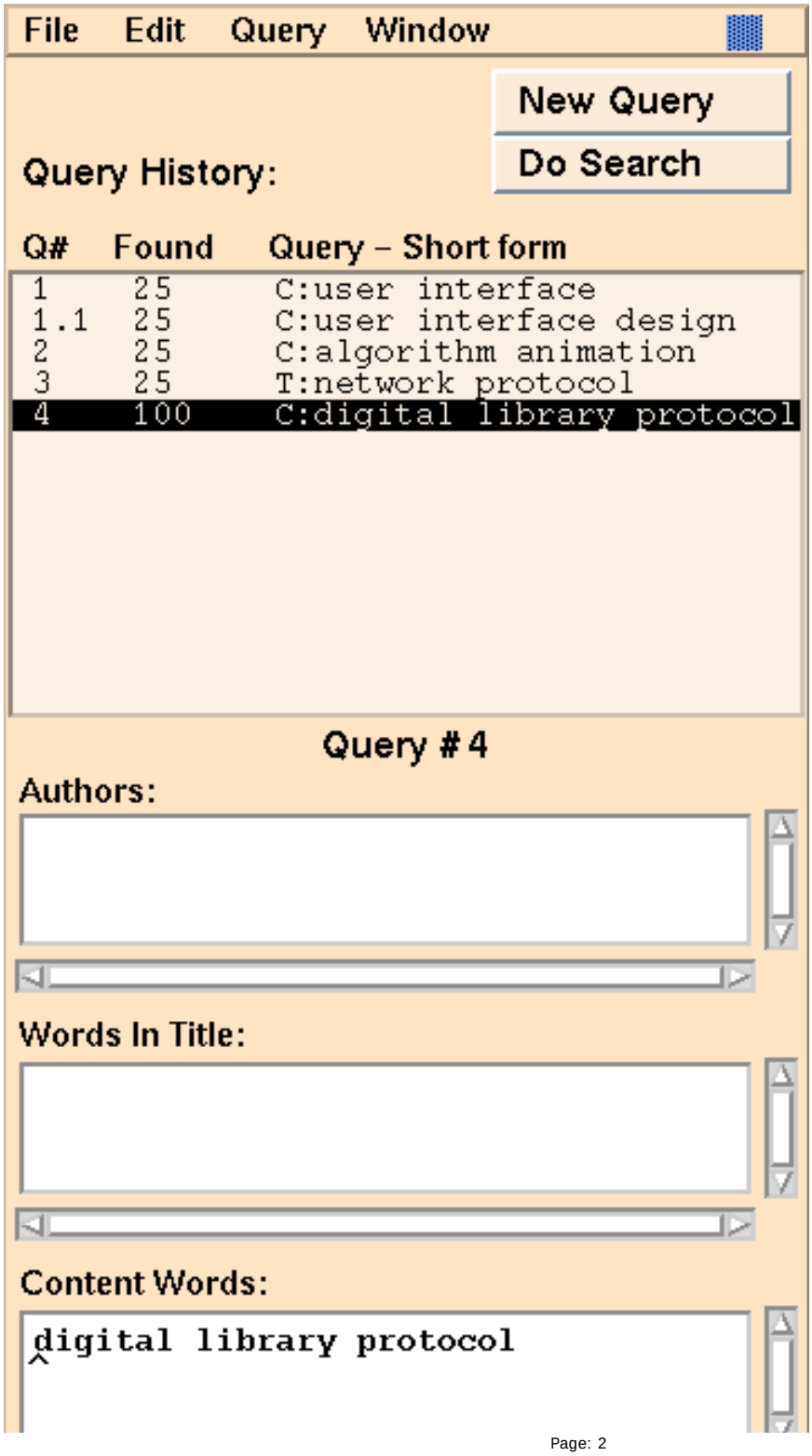
The interface includes:

- Big query:





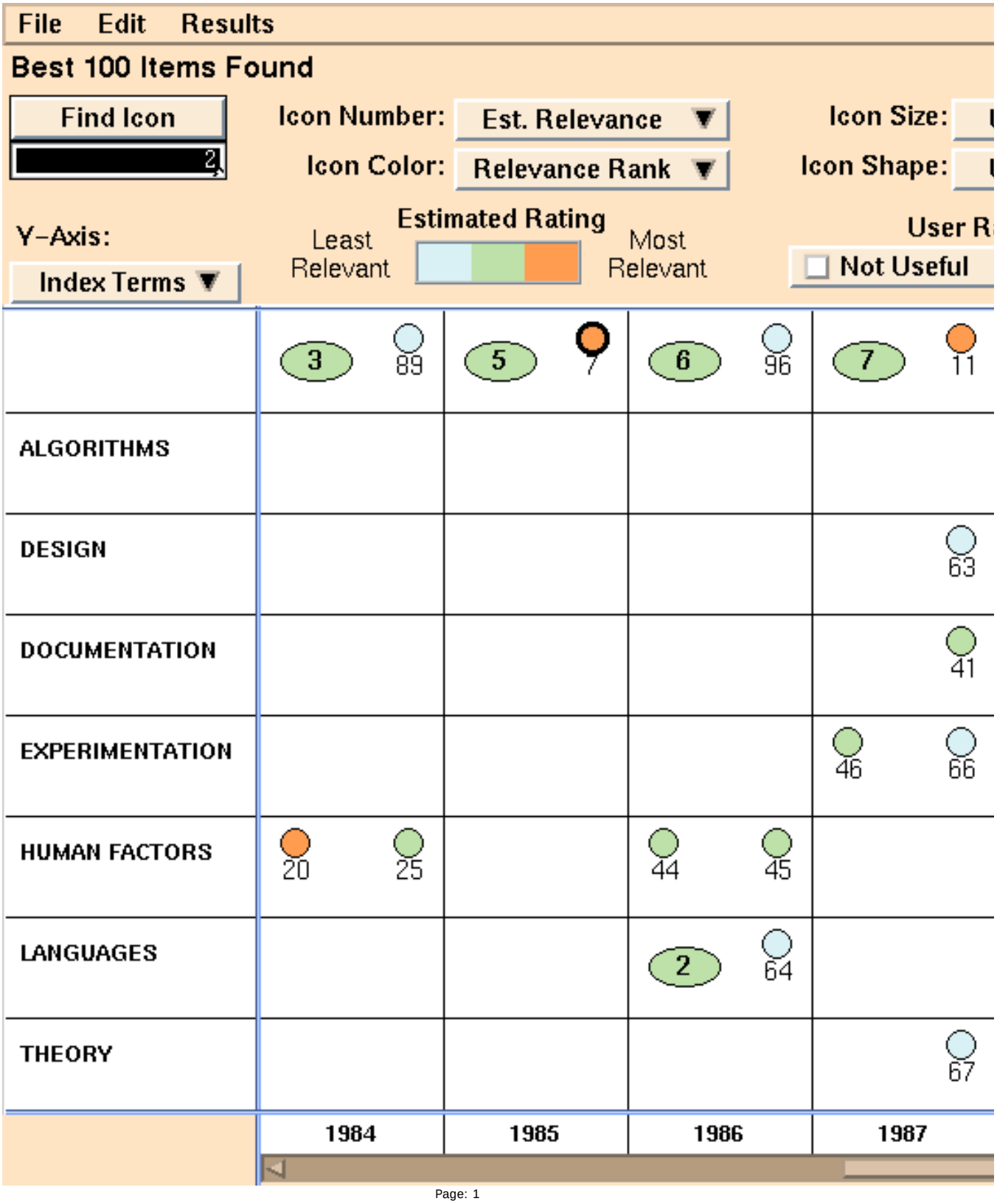
- Small query 1:

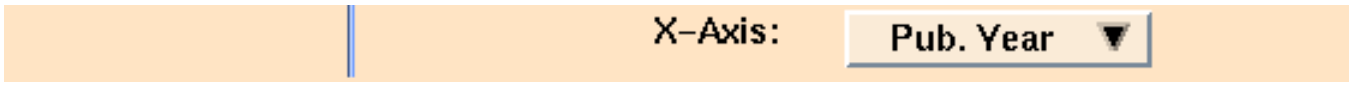


Envision - Results Screens

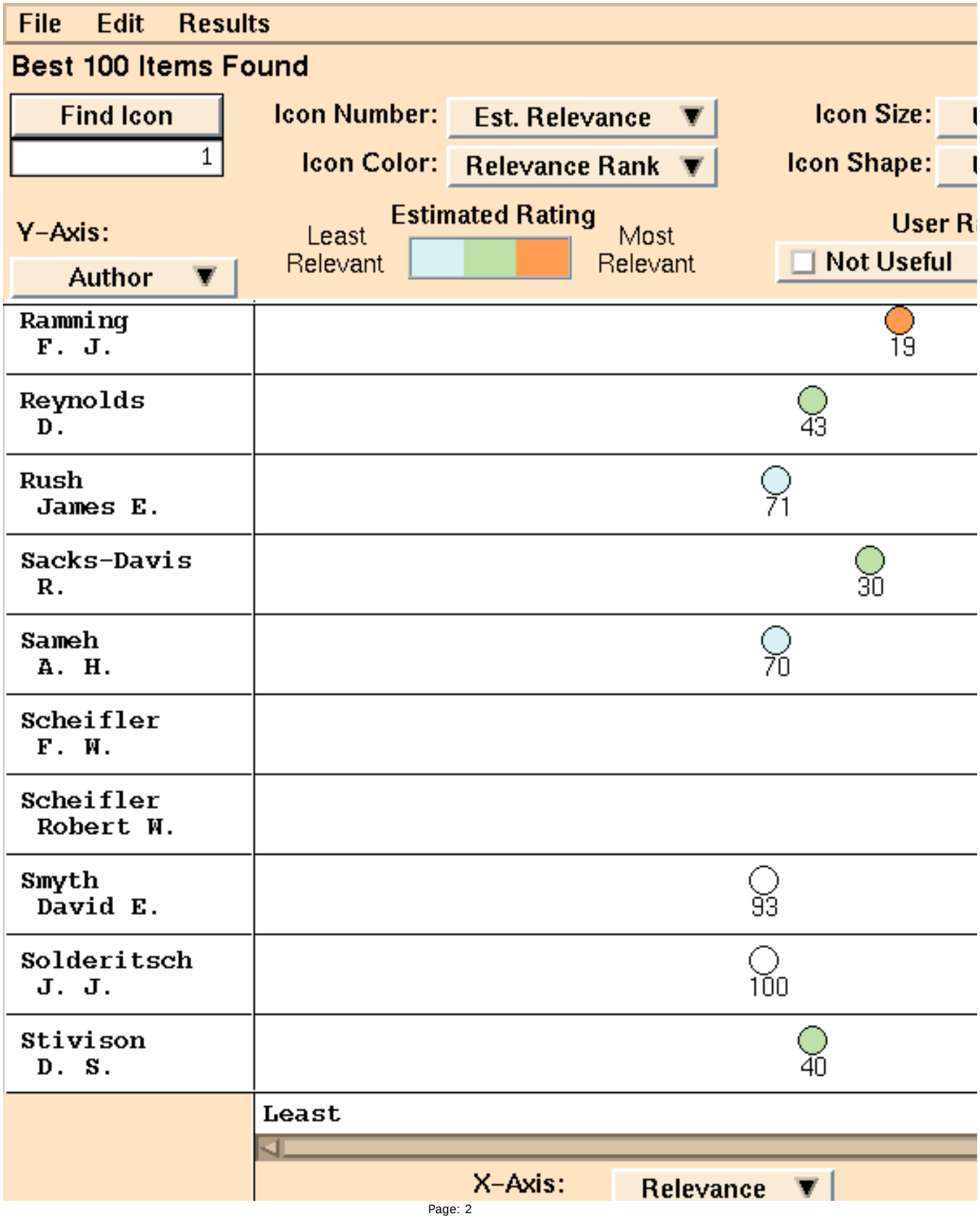
The interface includes:

- Graphic View 1:





- Graphic View 2:



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,

ACM Digital Library Collection at Virginia Tech

This archive was created by the digital library project at Virginia Tech, in cooperation with ACM and IBM. You can search and browse the documents in the Computer Science literature.

Search the ACM Digital Library Collection

- The **Fielded Search Form** allows you to perform a search on several fields of the bibliographic data,
- **Or** simply enter one or several words into the box below to search author, title, and abstract:

  [ ]
- **Or** browse articles from the ACM Digital Library Collection.

Further Information

- Virginia Tech Digital Library project page.
- Virginia Tech Computing Center research department.
- Cruise a list of documents related to Networked Computer Science Technical Reports Library (NCSTRL) - *Check out how to **participate** in NCSTRL, how to **download and install** the software, more technical details of the **Dienst** protocol and architecture (the technology behind NCSTRL), **links** to other sources of computer science technical reports, and more.*

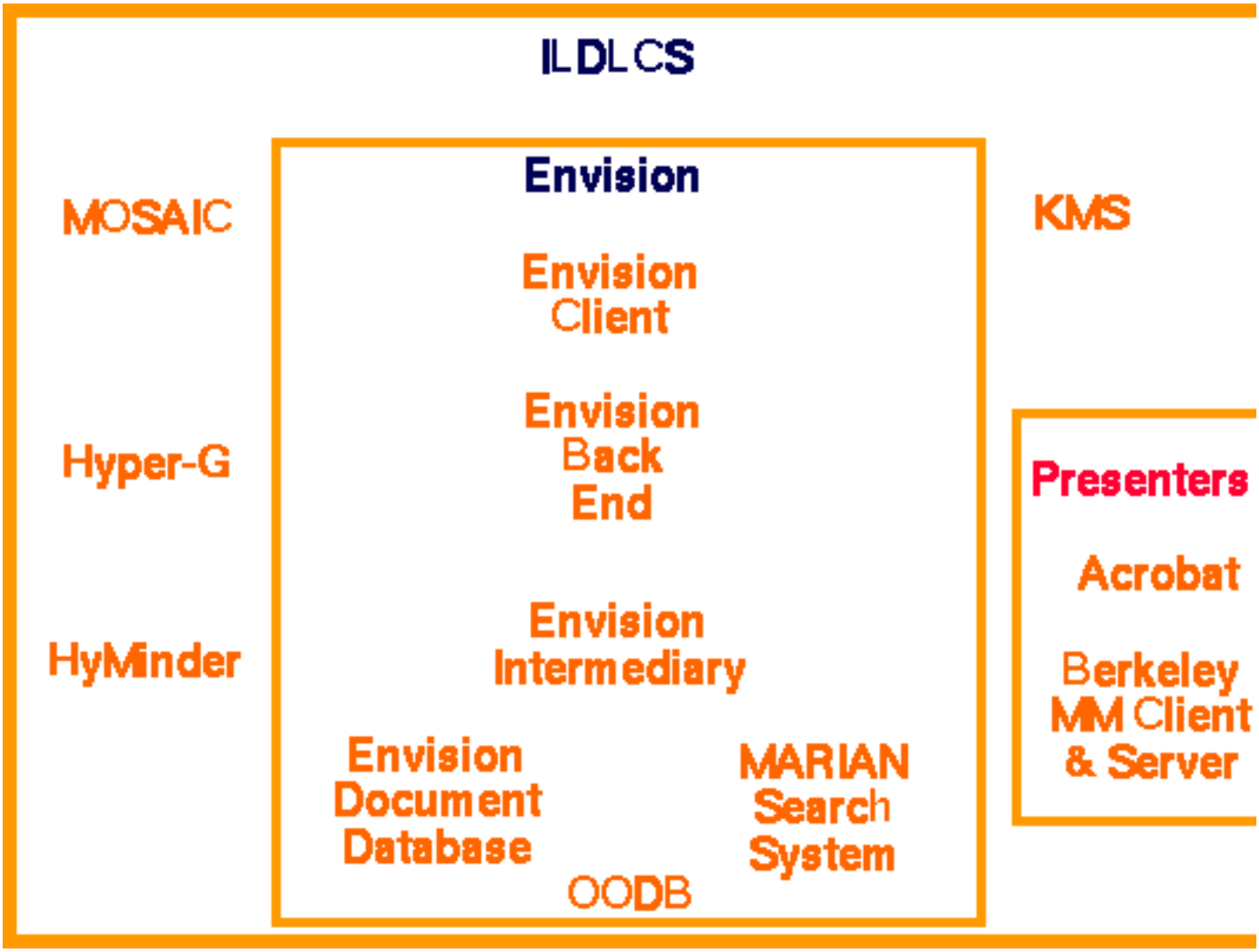
In case of problems send your comments to ACM digital library repository at Virginia Tech.

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

CS5604 - Information Storage and Retrieval Fall 1995 - Table of Contents

- **Assignments**
- **Calendar - Blacksburg** and **No. VA**
- **Computers and Tools**
- **Course Format**
- **Course Notes / Overheads**
- **Debates**
- **Department and Class Policies**
- **FAQ - Frequently Asked Questions**
- **Glossary (in process)**
- **Koofers (old quizes)**
- **News / Announcements (updated 951211@9:40pm)**
- **Photos of Class - Blacksburg** and **No. VA**
- **Quizes**
- **Readings and References**
- **Review**
- **Status of Student Progress - Quizes, Exercises** and **Other Work**
- **Searching Class Notes Online with Harvest**
- **Summaries of Articles**
- **Summaries of Class Sessions**
- **Syllabus - Blacksburg**
- **Syllabus - No.VA**
- **Trips**
- **Units/Course Outline - Blacksburg** and **No. VA**
- **WWW Link Sets: Instructor's - CS4624: Multimedia, Hypertext and Information Access - WWW Virtual Library (URLs organized by subject)**

Pointers to Previous Years' Materials

- **Fall 1994 and before**

Please send comments and suggestions to: fox@fox.cs.vt.edu

Units of CS5604 - Blacksburg

There are 11 units in this course, each with a 2-letter ID that symbolizes the main topical area considered. Each unit will be covered in 1-2 weeks of class time, has a set of associated readings, and has some lab or homework exercises that must be completed. Either 10 or 15 points, depending on the time and difficulty of the unit, will be given when mastery is demonstrated by a quiz grade of at least 90%.

- **DL:** *Digital Libraries*
  FOXE88d, SAMU91a, DONG87
  8/22, 8/24 (10 points)
- **IR:** *Information Storage & Retrieval*
  Ch1, SALT86a, Ch2
  8/29, 8/31 (10 points)
- **IF:** *Inverted Files / Boolean Systems*
  Ch3, Ch12, Ch15, SALT83d
  9/5, 9/7 [library], 9/12 (15 points)
- **SS:** *String Searching*
  Ch5, Ch10
  9/14, 9/19 (10 points)
- **RR:** *Ranking / Relevance Feedback*
  Ch14, SALT75b, Ch11
  9/21, 9/26, 9/28 [lab] (15 points)
- **CL:** *Clustering*
  Ch16
  10/3, 10/5 [lab] (10 points)
- **IN:** *Indexing / Document Analysis*
  Ch7, Ch8
  10/10 [lab], 10/12 (10 points)
- **SD:** *SGML / Document Translation*
  COOM87, MAMR87
  10/17, 10/19 (10 points)
- **HT:** *Hypertext*
  NIEL90a, HAAN92
  10/24, 10/26, 10/30 [MBONE] (15 points)
- **MM:** *Multimedia*
  FOXE91b, WALL91, PHIL91a, GREE92
  11/2 [lab], 11/7 [videotapes], 11/9 [lab / demos in Information Access Lab, McB 110], 11/14 (15 points)
- **KB:** *Knowledge-Based Information Retrieval*
  FURN87, LEBO88, MALO87a
  11/16 [lab], 11/28, 11/30 [lab] (15 points)

CS5604, Unit DL

Edward A. Fox
Department of Computer Science
Virginia Tech, Blacksburg VA 24061-0106

Abstract:

The vision of digital libraries, expressed so eloquently by Vannevar Bush in 1945 when he described Memex [1], provides a rallying point for work in information storage and retrieval. With our course theme of digital libraries, we begin by trying to figure out what they could and should be, their advantages and disadvantages, what might be in them and how they might function.

Unit Highlights

Lecture/Motivating Question:
What are the potential benefits, costs, and disadvantages relating to digital libraries?

Discussion:
Small groups of 2-6 people should meet during the week, drawing on the readings and other knowledge, to argue the pros and cons of digital libraries. Each group will send in an email summary to the instructor.

Computer Exercise:
Since netlib is one of the oldest and most widely used electronic mail-based service that provides digital library services, students will try it out to gain experience with the use of indexes, query facilities, and semi-interactive archives.

Introduction

On July 20-21, 1992, the National Science Foundation sponsored a workshop on digital libraries, prompted by an earlier proposal prepared by Lesk, Fox, and McGill, that called for a National Electronic Library for Science, Engineering, and Technology. NSF will fund a good deal of R&Din this area in the 1990s, helping bring to fruition the dreams of such visionaries as Vannevar Bush and J.C.R. Licklider.

At the workshop, David Hartzband of DEC recounted experiences of a major multinational corporation involved in office and factory automation. They found that technology alone is not enough, that social and anthropological knowledge is also needed to effect change. Thus, in this unit, which introduces the course and our theme of digital libraries, there will be readings and discussion about the legal issues and essential characteristics of digital media and digital libraries.

To give concreteness to the idea of digital libraries, you will use computer networks to access *netlib*, an electronic archive for numerical and mathematical software and related information. This will also illustrate how semi-interactive querying can be carried out using electronic mail (or xnetlib).

Since during this course we will be making use of the rapidly expanding digital library that is being developed in Project Envision, it is important to understand the background to that effort. The article on ACM Press

Database and Electronic Products describes earlier work toward an ACM digital library, relates it to products and services for ACM members and other users, and discusses some of the financial and pragmatic aspects of such an archive.

This unit paves the way for discussions of technology, methodology, theory, commercial and research systems for IS&R. It sets the stage for detailed discussions, and introduces the key theme of digital libraries, that was part of legislation introduced by Senator Gore in 1992, and re-introduced in Congress in 1993.

Objectives

From the *Course Objectives* a key point is to prepare students to discuss and explain the main issues relating to developing digital libraries and related services. Toward that end, this unit will deal with digital library efforts by ACM and by experts in numerical software, and will explore the legal and conceptual issues relating to the fundamental properties of digital media.

Other, specific objectives include being able to:

1. define and explain *digital library*, and describe a vision of its future in terms of possible scenarios for two different groups;

2. list three key challenges facing builders of digital libraries;

3. prepare a draft proposal that could be submitted to ACM Press Database and Electronic Products;

4. discuss six characteristics of digital media that are likely to lead to significant changes in future law about those media, and why;

5. obtain mathematical software and information from netlib; and

6. describe the main services, advantages, disadvantages, limitations, and costs of netlib.

Suggested Procedure

There are three main types of effort required. First, the readings (see the second from last section) should be carefully studied, keeping unit objectives in mind (see the third section). Second, students should prepare for and engage in the group discussions, drawing upon the readings and other resources (see next subsection). Third, students should use electronic mail (or xnetlib) to carry out the Exercises relating to netlib (see subsection after debate discussion).

Debate Topics

The class will break into groups. Each will be assigned 3 questions. Think about your group's debate topics, since people involved in that discussion group must discuss them (either pro or con). After your group has a discussion on its 3 topics totalling about 45 minutes, work together to send an email summary to the instructor, that has the name of each person in the group, and gives the consensus viewpoint on each of your 3 topics. Do a careful job in writing, with different people playing different roles (see discussion under Course Format). Rotate the roles for each different topic.

1. Technology will cause the demise of print publishers, because they will be unable to prevent theft and widespread distribution of electronic forms of their publications.

2. People won't take computers to the beach or put them on their night tables or spread them out to read the comics, since books and other print forms are much cheaper, friendlier, lighter, easier to use in a

variety of lighting situations, and cover a larger surface area.

3. ACM should allow electronic submissions for all of its publications to be made according to authors' wishes, and should ignore issues of standardization.

4. Electronic publication should be funded by a system of subscriptions, so users are *encouraged* to make use of published materials that are covered by any of their subscriptions.

5. ACM should not call for proposals but should instead carry out an ambitious electronic publishing effort in-house, that should rapidly break even financially.

6. There are too many electronic publishing forms and standards, and too little user-oriented access software, to motivate people to buy into the new technology.

7. Copyright is a useless concept for digital libraries because publishers will control and charge for access to and use of digital works.

8. Given the federal government's role in NSFNET and the NREN, ubiquitous networking seems imminent, and means that there will be national networked access to future digital libraries. One result will be serious problems with international copyright violation due to the ease of transmission.

9. Derivative works will be commonplace, and primary publishers will suffer greatly from those who have shaped and combined prior work in such a way as to abandon giving credit to original sources.

10. Copyright law cannot hope to deal with classifying works by media type, given that multimedia publications will become widely used.

11. Storage hierarchies, with personal, departmental, campus, state, regional, national, and international levels, will operate with multiple copies of each work available, in places chosen to optimize performance.

12. With highly linked collections of materials like hypertexts and their search trails, whose raw materials originated from numerous sources, copyright protection and remuneration should go to the new editor, and not to the original authors.

Computer Exercise

In this exercise you will send email to netlib, following the instructions in the article [2]. See /u1/README/netlib on fox.cs.vt.edu (or, click here with Mosaic or Netscape) for current instructions, which involve sending mail to the netlib@ornl.gov system. Alternatively, you can run the xnetlib program (which in most CS dept. machines is located in /usr/local/X11R5/bin). Another alternative is to click here with Mosaic or Netscape; xnetlib is no longer supported now that WWW browsers can access information directly.

Please send a copy of all results you receive to the instructor for review.

1. get a copy of the index, and then identify works that came from Virginia Tech.;

2. repeat task 1 above by doing a direct search for the same items;

3. find works by John Dennis (but not by other people named John or Dennis); and

4. find the originators of HOMPACK

Here are some hints:

1. Some people use their first initial instead of their first name.
2. Software systems may have only one originator.
3. Software systems at universities often are partially developed by the students of a faculty member who is responsible for the system.
4. Drs. Watson and Ribbens in CS at Virginia Tech work on numerical analysis problems; their names and research interests could be found among the departmental web pages.

Comments on Readings

Note: For information on all articles for the course, click here with Mosaic or Netscape.

Note: A new report on digital libraries and related research opportunities may be of interest. This includes definitions and use scenarios. Use a WWW browser to access Interoperability, Scaling, and the Digital Libraries Research Agenda.

FOXE88

This article describes the early work on ACM Press Database and Electronic Products, and plans for the future. The research aspects of this program have been carried forward into Project Envision, and ACM Headquarters and the Publications Board, along with the Electronic Publishing Volunteer Advisory Committee, are coordinating work on a plan that will include an electronic archive and electronic submissions.

Paragraph 1 gives a snapshot of the status, which is amplified at the end of the *Introduction*. The earlier part of the *Introduction* describes in general terms what technological and related advances have made digital libraries possible.

The last 3 paragraphs of *Vision* are important, dealing with collection building, standards, and the main classes of services. The *Challenges* section calls for more vision and ideas (such as those given in the first paragraph of the *Opportunities* section), then for focused R&D, and finally for work on economic, social and legal issues. Funding will be needed from ACM SIG's (paragraph 2) and from partnerships (paragraph 3).

The *Organization* and *Acknowledgments* sections are not relevant. However, the *Proposals* section gives important guidance regarding what to look for in developing information products, services, or even multimedia information packages.

SAMU91a

This article gives valuable insight into copyright and legal matters, but is especially helpful in pointing out important characteristics of digital media. The six characteristics should be carefully studies and pondered. The last one, on nonlinearity, should be re-read after completion of Unit 8 on Hypertext. Note that Ms. Samuelson's husband is Robert Glusko, who has been very active in R&Drelating to hypertext.

The second section, on *Replication*, explains an influential court case and its implications, and discusses several clever schemes for generating revenue in connection with electronic publishing.

Key to our course are the remaining sections. *Transmission and Multiple Use* are essential parts of digital libraries, but there are serious dangers of piracy. *Plasticity* is one of the key added values of digital media, but protecting authors' rights will demand careful balancing of this benefit with the need for extending copyright protection to changes. *Equivalence* issues relate to multimedia, but are not clearly explained here. *Compactness* is not really the theme of the next section - rather it is about storage cost-effectiveness and storage hierarchies, an idea which dates back to such discussions as [3]. Issues of *nonlinearity* will be dealt with later in the course, but are previewed in an interesting way here.

DONG87

The *netlib* system was one of the first to provide electronic mail-based access to archives, and serves the scientific computing community. This short article describes how the system works and can be used (though details and addresses have changed!), summarizes the contents of the archive, briefly explains the server, gives advantages and disadvantages, and closes with a list of needed future enhancements and opportunities.

Summary of Key Concepts

1. Digital libraries should carry over the benefits of current libraries, and add value in terms of speedy, ubiquitous, and flexible access, as well as through support of searching, browsing and linking.

2. Not only must we consider the tools and the content of digital libraries, but we also must be concerned with electronic media and their characteristics; the demands of applications; and questions relating to economic, social and legal policies.

3. Building digital libraries includes many types of work including:

    - cataloging and interconnection of existing and emerging resources;
    - developing new, integrated, electronic collections;
    - advancing research in IS&R, hypertext, navigation, computer supported cooperative work (CSCW), education; and
    - convincing others regarding solutions to social, economic, and legal issues.

References

1 V. Bush. As we may think. *Atlantic Monthly*, 176:101-108, July 1945.

2 Jack J. Dongarra and Eric Grosse. Distribution of mathematical software via electronic mail. *Communications of the Association for Computing Machinery*, 30(5):403-407, May 1987.

3 J. C. R. Licklider. *Libraries of the Future*. The MIT Press, Cambridge, MA, 1965.

Next Up Previous

---

fox@cs.vt.edu
Fri Sep 1 14:12:34 EDT 1995

CS4624: Multimedia, Hypertext, Information Access
Table of Contents (Spring 1996)

Please read: News / Announcements
(Most recent update is for 960503.)
DEADLINE FOR TURNING IN WORK WAS MIDNIGHT !

Outline

Student Information

- **Home Pages for Tuesday** and **Thursday**
- **Progress Report**

Activities

- **Lab Sessions**
- **Lecture Notes**
- **Projects**
- **Quizes, Final (ONLY FOR ENROLLED STUDENTS!)** and Pre-test, Post-test
- **Trips and Special Events**

Information Sources and Access Points

- **Class Inventory**
- **Computers and Tools**
- **Figures**
- **Figures with Captions**
- **Glossary**
- **Index**
- **Koofers (old quizes, final)**
- **Link Sets**
- **Tutorial on Multimedia (1994)**
- **Readings and References**
- **Searching All CS Class Notes Online with Harvest**

Syllabus (all together, as of 2/24/96, plus Outline) or in sections:

- **Calendar**
- **Department and Class Policies**
- **Instructor and GTA**
- **Syllabus Details: Grading, etc.**

Fall 1995 Version

Comment Form --- Email Submission

Note: "4984" is the number used 1995-96 but "4624" is the number to be used thereafter.

Virtual Computer History Museum

Our group will design a prototype virtual museum of computer history on the World Wide Web. We envision various methods of accessing the 'exhibits' within the museum, such as:

- **a text-based chronology of events**
- **a graphical timeline of events**
- **a searching mechanism**
- **an image gallery**
- **pre-defined tours based on specific times or people**
- **and links to other computer-related museums**

The Virtual Computer History Museum Group consists of:

- Charles Atwood (atwoodc@vt.edu)
- Jon-Erik V. Lido (jlido@vt.edu)
- Mike Marston (marston@csugrad.cs.vt.edu)
- Andy Wagliardo (awagliar@core-dump.com)

Much assistance, guidance, support, and direction was provided by

- Dr. Edward A. Fox (fox@vt.edu)
- Dr. J.A.N. Lee (janlee@vtopus.cs.vt.edu)

of the Computer Science Department at Virginia Polytechnic Institute and State University (Virginia Tech).

Meeting place: McBryde 110, Derring 2069

Meeting time: Monday 2:00pm - 3:00pm, Thursday 12:15pm - 1:45pm

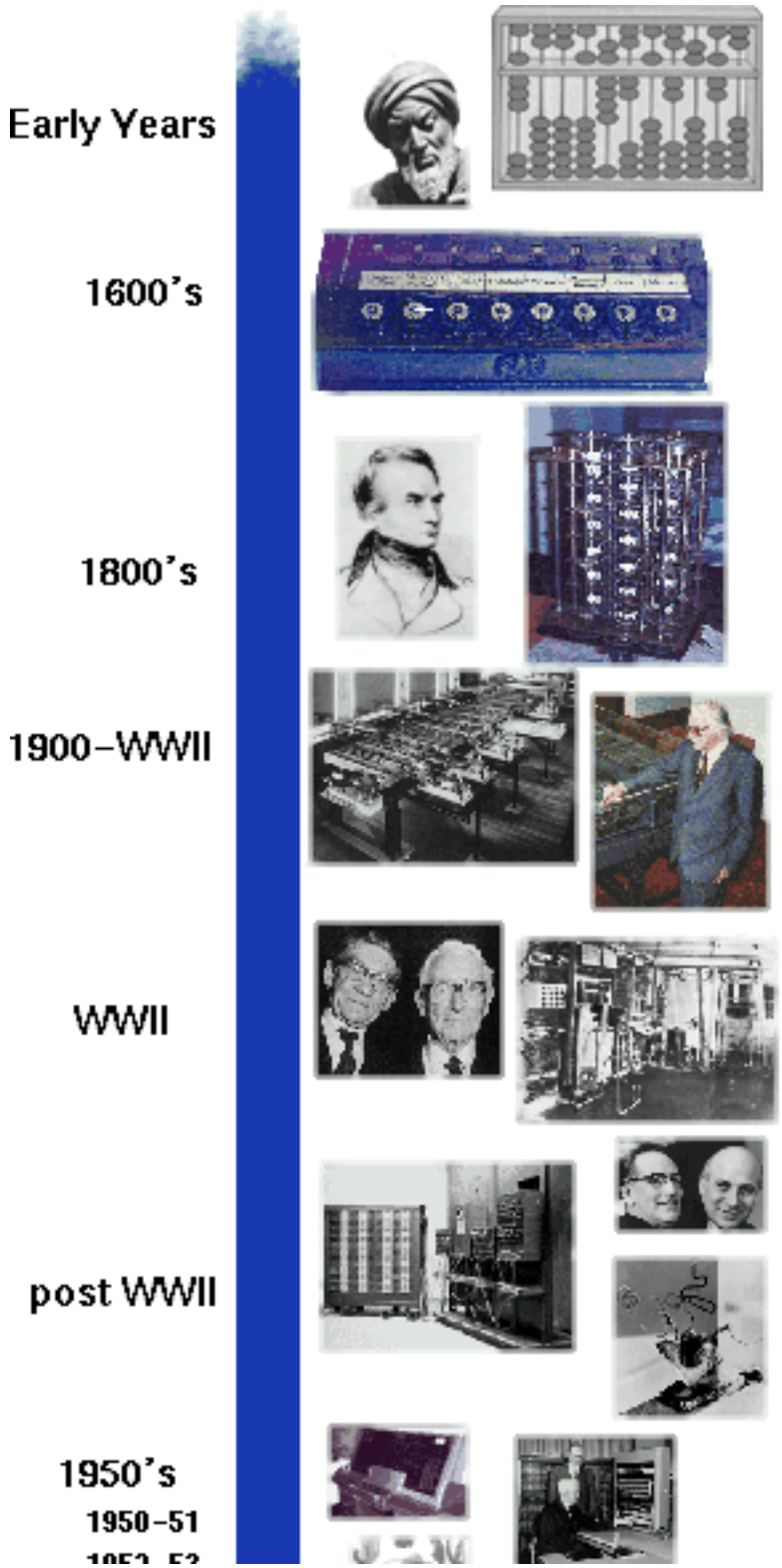
Contact: Andy Wagliardo ((540)232-3627 museum@mail.core-dump.com)

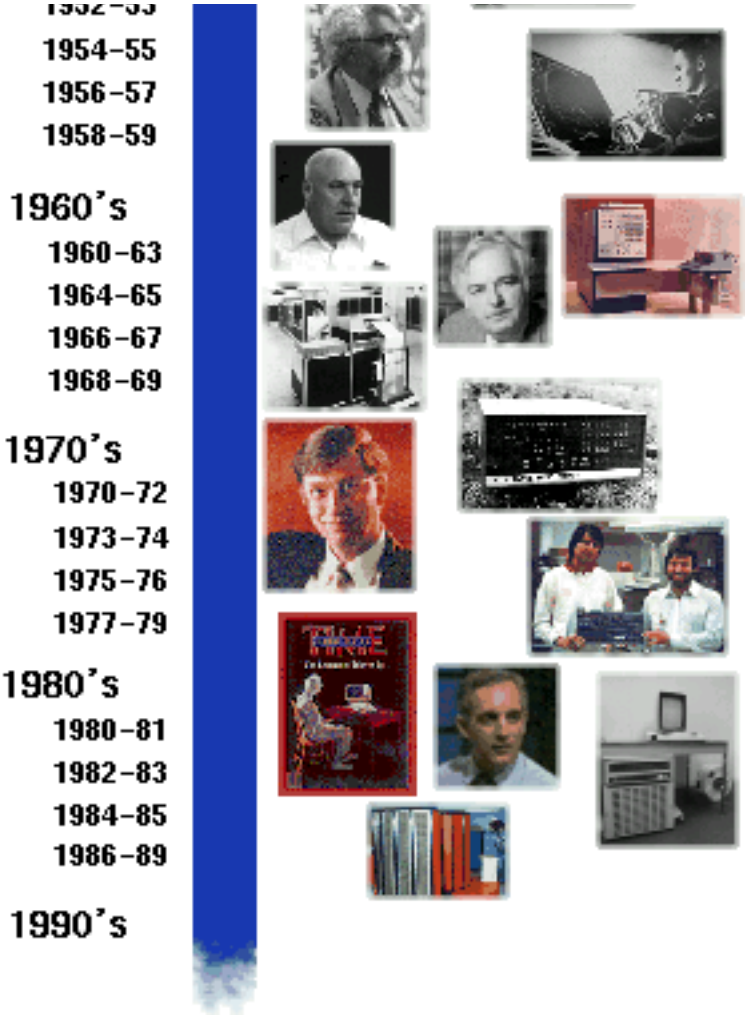
Mailing Address:
Andy Wagliardo
202 Barringer Hall
Blacksburg, Virginia 24060-0001

WWW Address:
http://ei.cs.vt.edu/CSNotes-classes/VirtualComputerHistoryMuseum

Timeline of Events in Computer History

Click anywhere on the imagemap to display the specified timeframe in computer history





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.

The market for information
A NetBill scenario
NetBill design
NetBill architecture

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.

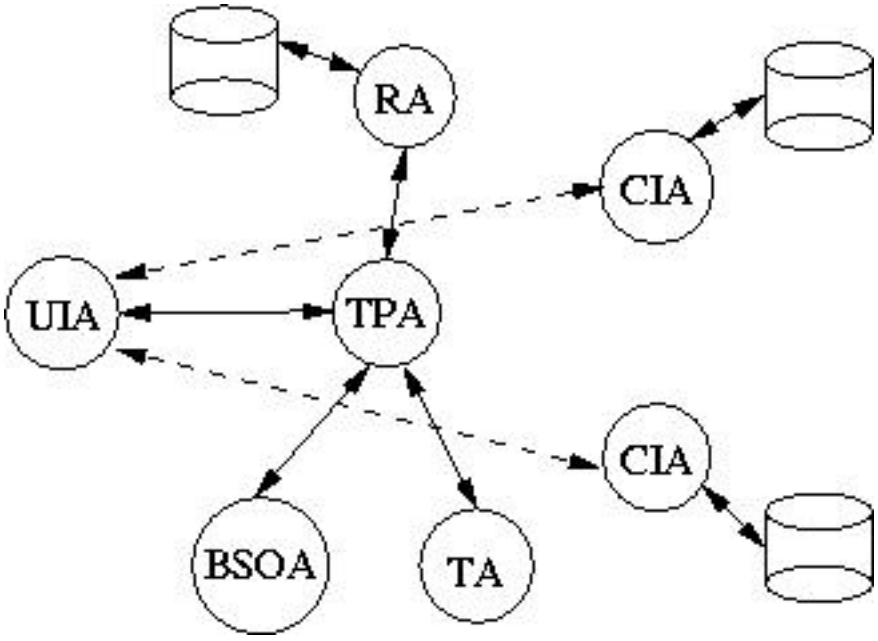


Figure 1: Schematic view of the UMDL architecture showing the agents in it and the communications between them. The solid lines represent the first part of the search, where the UIA looks for applicable collections. The dashed lines are the second part, when the UIA queries the collections that were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

(a subset of) them back to the UIA -- but, more often, what is returned does not satisfy the constraints of the query task. The TPA possesses procedural knowledge about how to handle such cases, which could in turn involve enlisting the help of other agents in the network, such as agents that can broaden/narrow topics of search (Broad System of Ordering Agents (BSOAs)) and agents that can provide synonymous terms (Thesaurus Agents (TAs)). Moreover, for a candidate collection, the TPA could contact the Collection Interface Agent (CIA) associated with it, to probe more deeply into its capabilities, content, price, or availability. Ultimately, the TPA formulates a candidate set of collections for the user to contact, and returns the set to the UIA. The UIA might then proceed to contact some of these CIAs directly and ask for the particular documents, before returning the final answer to the user.

- Task Planner Agent
  - TPA architecture
  - TPA language
  - TPA knowledge

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

Jose M. Vidal
jmvidal@umich.edu
Mon Mar 11 15:41:27 EST 1996

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.

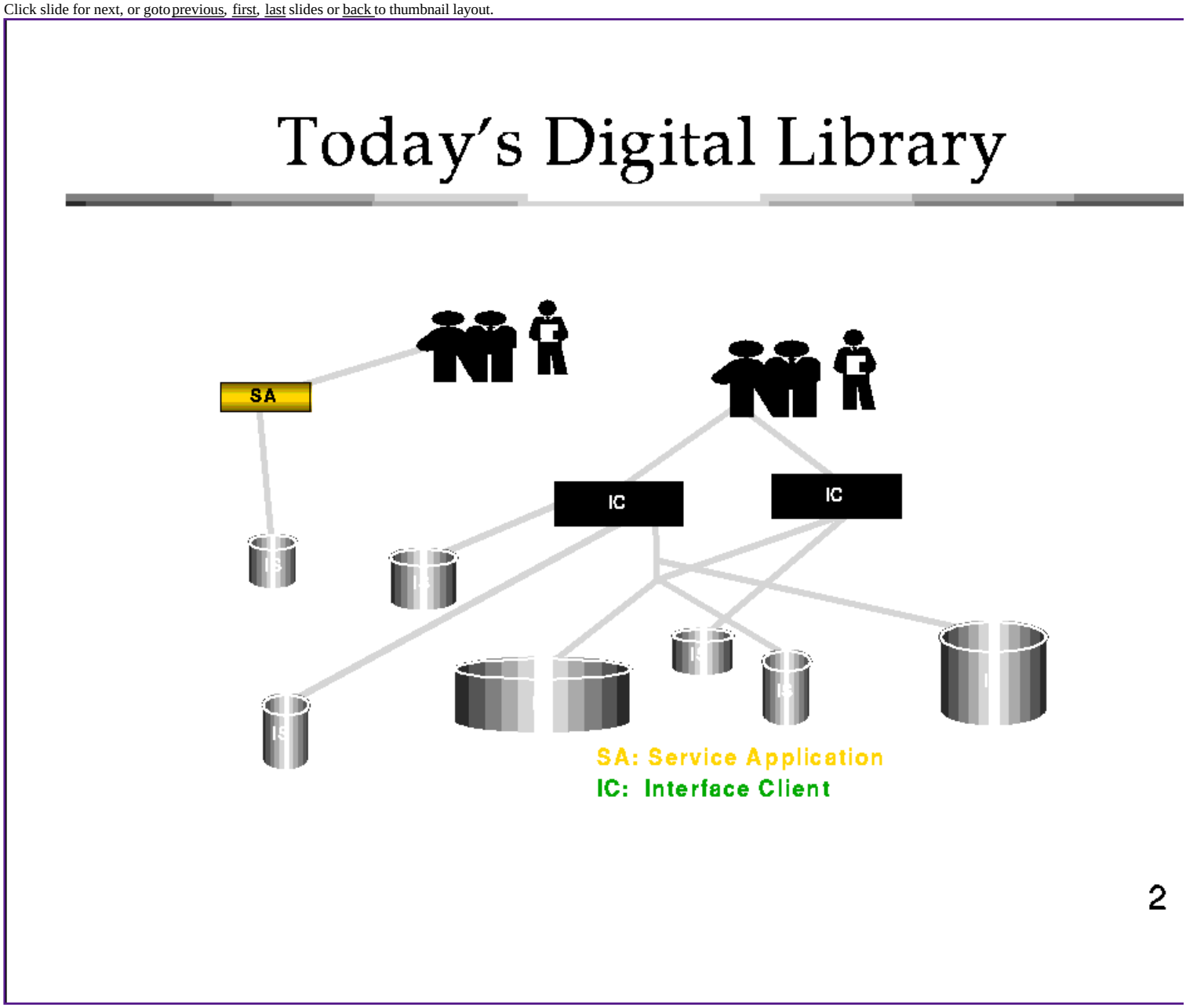
- ILU -- Our installation and examples.
- Python -- Our installation of Python programming language
- ObjectCenter -- Discussion of ObjectCenter
- COS -- Digital Library Testbed Common Object Services
- CVS -- Our use of CVS in the testbed.
- Various Manuals (CVS, Python, ILU)

- 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.

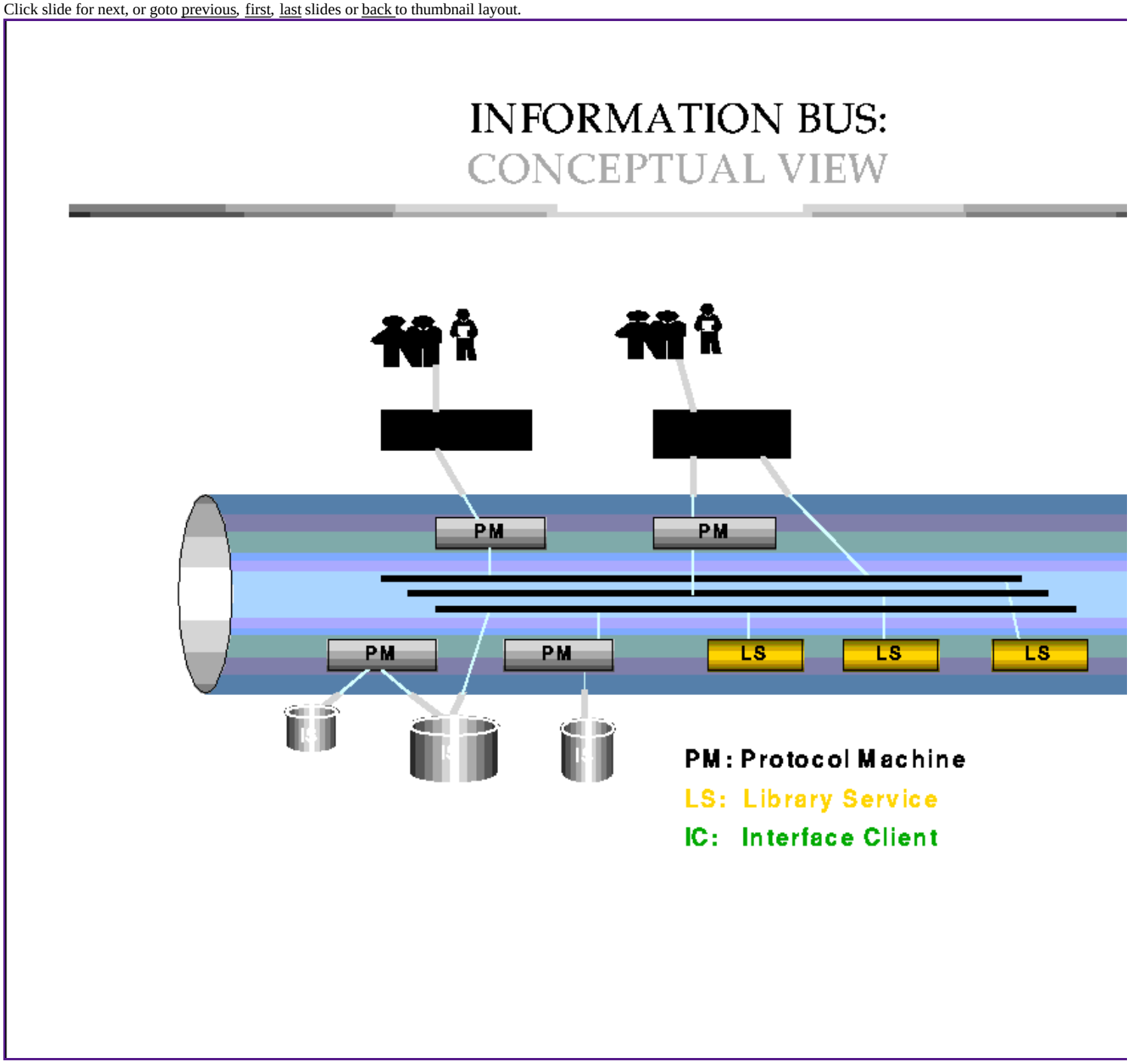


Click slide for next, or goto previous, or back to thumbnail layout.

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.

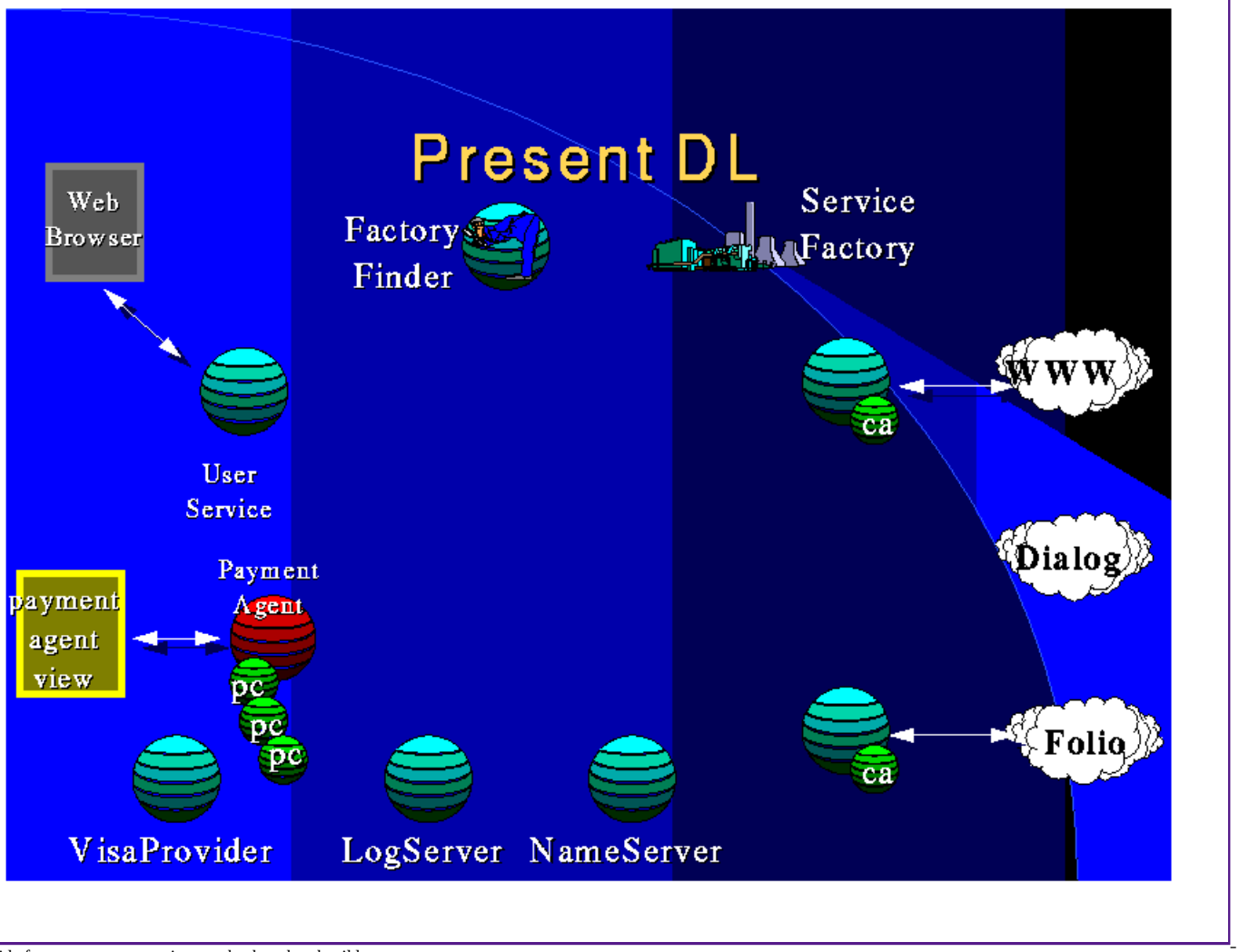


Click slide for next, or goto previous, or back to thumbnail layout.

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.

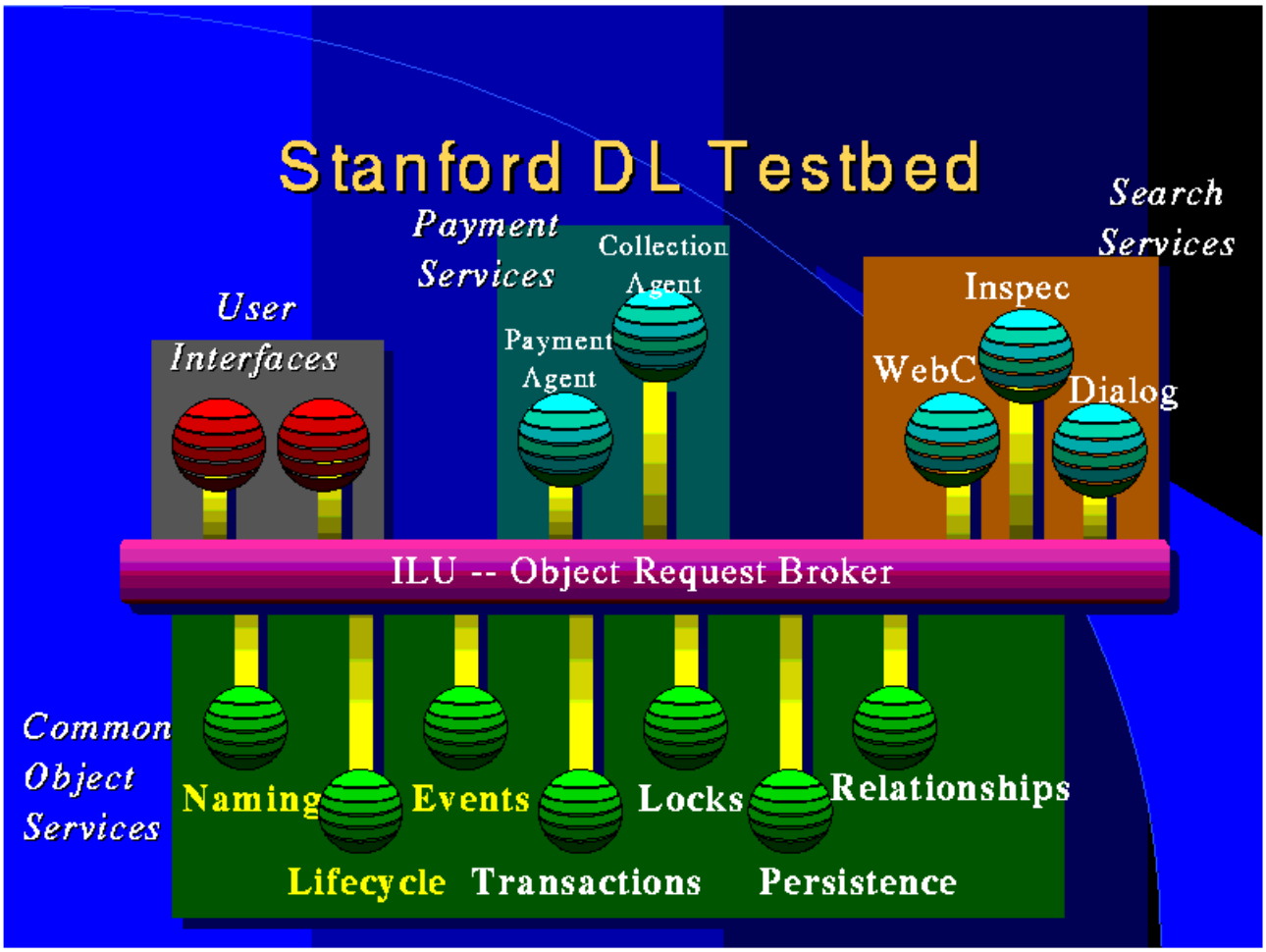


Click slide for next, or goto previous, or back to thumbnail layout.

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

COS - Common Object Services: Slide 23 of 49.

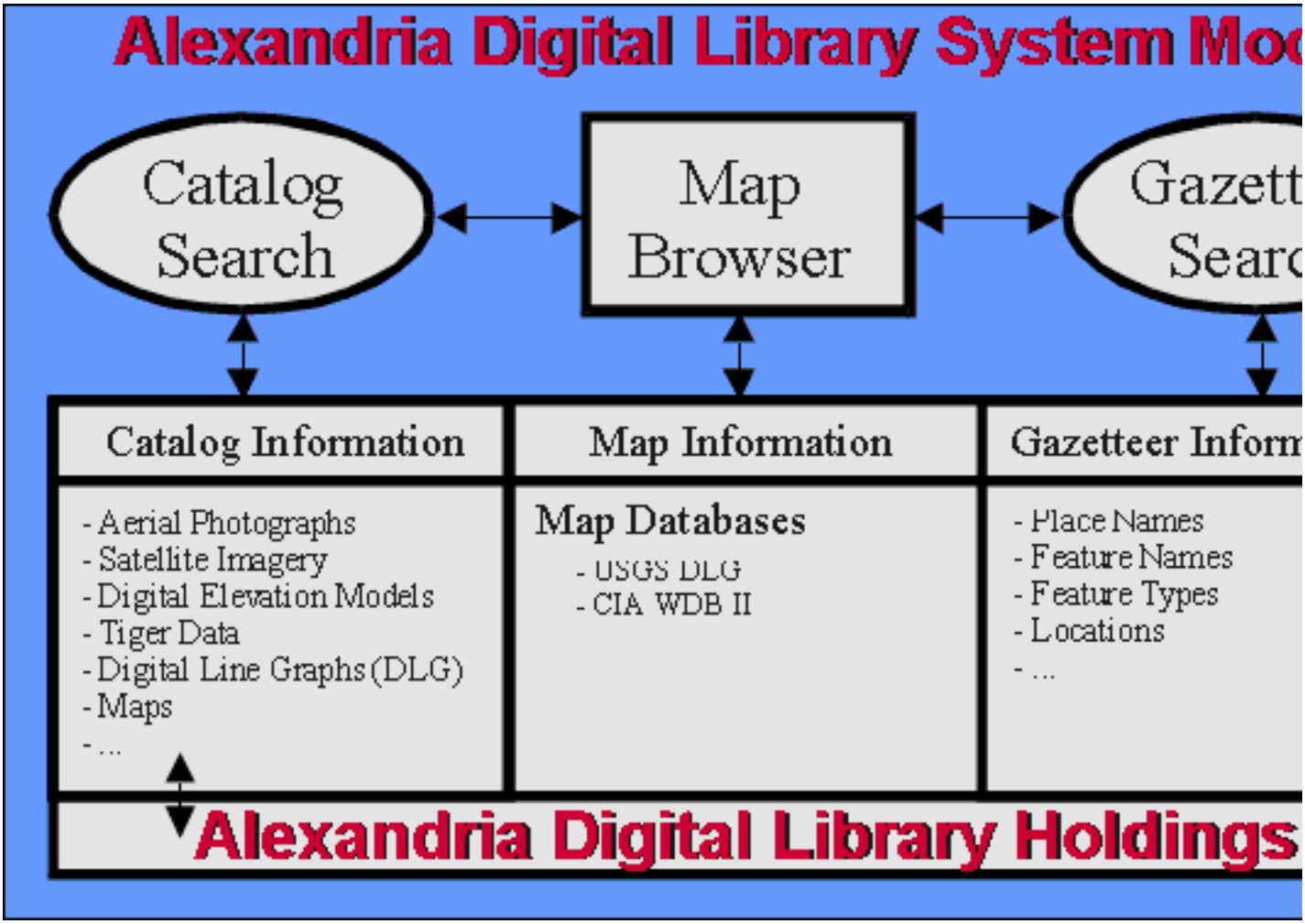
Click slide for next, or goto previous, first, last slides or back to thumbnail layout.

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

Digital Libraries Webmaster
Webmaster@diglib.stanford.edu

Tutorial Table of Contents

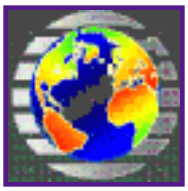
Prev Next



Conceptual model of the Alexandria Web interface

- Conventions
- Session / System Setup
- Map
- Gazetteer
- Catalog
- Overview of Current Holdings
- Walkthroughs (Example Sessions)
- Feedback
- Technical Reference
- Acknowledgements

1996 ANNUAL REPORT Table of Contents



Alexandria Digital Library

---

1 PROJECT SUMMARY

2 SIGNIFICANT EVENT

3 OVERVIEW OF PROGRESS IN LIBRARY DEVELOPMENT

4 CURRENT STATUS OF TESTBED DEVELOPMENT
4.1 Architecture of the Testbed System
4.2 Catalog The Component
4.2.1 Basic Metadata: USMARC and FGDC Standards
4.2.2 The ADL Gazetteer
4.2.3 Other Catalog Issues
4.3 Collections
4.4 The User Interface
4.4.1 User Interface Issues
4.4.2 User Interface Implementation
4.5 Image Processing
4.5.1 Browsing and Progressive Delivery
4.5.2 Texture-Based Retrieval
4.6 Parallel Processing
4.7 Computing Support for the Testbed: Hardware and Communications
4.7.1 Current Computing Equipment
4.7.2 Current Networking Support
4.7.3 Current Storage Support
4.7.4 Current "Other" Hardware and Software Support
4.7.5 Equipment and Facilities Needs

5 RESEARCH ACTIVITIES AND PROGRESS
5.1 Library Team
5.1.1 Metadata and Catalog Interoperability
5.1.2 Alexandria Atlas Subteam
5.2 Interface Design and Evaluation Team
5.3 Information Systems Team
5.4 Image Processing Team
5.5 Performance and Parallel Processing Team

6 COMPARISON OF ACTUAL AND PREVIOUSLY PLANNED R&D ACTIVITIES
6.1 Previously Planned R&D Activities
6.2 Actual Versus Agreed R&D Activities

7 PLANNING FOR AN OPERATIONAL DIGITAL LIBRARY

8 MANAGEMENT REPORT
8.1 Organizational Structure

8.2 Interactions and Levels of Activity
8.3 Personnel
8.3.1 Professional Staff
8.3.2 Faculty Investigators
8.3.3 Graduate Students Associated with the Project

9 EXTERNAL RELATIONS AND INTERACTIONS
9.1 Advisory Board
9.2 Third DLI meeting at UCSB
9.3 Alexandria Metadata Workshop
9.4 First Alexandria Design Review
9.5 Interactions with Old and New Partners
9.5.1 California Environmental Resources Evaluation System (CERES)
9.5.2 Central Imagery Office (CIO)/National Imagery and Mapping Agency(NIMA)
9.5.3 Central Intelligence Agency (CIA)
9.5.4 CIESIN
9.5.5 Excalibur/ConQuest
9.5.6 Defense Mapping Agency
9.5.7 Digital Equipment Corporation (DEC)
9.5.8 ERDAS
9.5.9 Environmental Systems Research Institute (ESRI)
9.5.10 Hughes
9.5.11 Library of Congress
9.5.12 Lockheed/Space Imaging
9.5.13 NASA
9.5.14 Oracle
9.5.15 O2
9.5.16 San Diego Supercomputer Center (SDSC)
9.5.17 Sierra Nevada Ecosystem Project (SNEP)
9.5.18 SPOT Image
9.5.19 The Analytic Science Corporation (TASK)
9.5.20 United States Geological Survey (USGS)
9.5.21 United States Navy, Stennis
9.5.22 United States Navy, San Diego (NRAD)
9.5.23 Earth Data Analysis Center (EDA), University of New Mexico
9.5.24 Utah State University/Mojave Database Cooperative
9.5.24 Xerox
9.6 Interoperability Agreements and Activities
9.6.1 Stanford DLI Project
9.6.2 Berkeley DLI Project
9.6.3 Illinois DLI Project
9.6.4 CMU DLI Project
9.7 Digital Libraries for School Students
9.8 Visits and Demonstrations
9.9 Talks and lectures
9.10 Professional interactions with DL Community
9.11 Publicity

10 EDUCATIONAL ACTIVITIES

11 PUBLICATIONS
11.1 Abstracts of Selected Articles
11.1.1 Testbed Research and Development
11.1.2 Information Systems Research
11.1.3 Image Processing Research
11.1.4 Parallel Processing Research

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

(National Center for Supercomputing Applications)

- Evaluation

Graduate School of Library & Information Science (*GSLIS*) + Sociology, Economics

- Technology

GSLIS + Computer Science (*CS*), NCSA

Principal Investigators

- Bill Mischo, UL, Testbed Lead
- Tim Cole, UL, testbed collection
- Joseph Hardin, NCSA, Infrastructure Lead
- Beth Frank, NCSA, infrastructure software
- Ann Bishop, GSLIS, Testbed Evaluation
- Bruce Schatz, GSLIS, Technology Research

- executive committee: UL, NCSA, GSLIS, CS

Primary Partners

- journal/magazine publishers:

IEEE Computer Society

Institute of Electrical and Electronics Engineers (IEEE)

American Society of Civil Engineers (ASCE)

American Society of Agricultural Engineers (ASAE)

American Physical Society (APS)

American Institute of Physics (AIP)

American Astronomical Association (AAS)

John Wiley & Sons

INSPEC, Compendex (Engineering Index)

- testbed: SoftQuad, EBT, OpenText , Hewlett-Packard
- infrastructure: OCLC, CNRI, Spyglass, Microsoft

DLI Collaborators

- Stanford interoperability experiment
- Michigan search interface, OpenText SGML
- CNRI/Cornell secure object store
- Carnegie-Mellon NetBill charging trial

???

- UC Santa Barbara GIS correlations
- UC Berkeley image processing
- Carnegie-Mellon network video

Illinois Project Groups

- Testbed

process, index, search, display SGML collection

- Infrastructure

multiview interface to Web distributed repositories

- Evaluation

usage and users of testbed with information context

- Technology

semantic retrieval, both manual & automatic

analysis environments with objects & semantics

Testbed Goals

Large Organized Collection

- SGML pipeline direct from the publishers (deposit)
- complete articles fully tagged and indexed (search)

Large Number of Users

- faculty/students around UIUC then Big Ten
- Internet interface with multiple views (display)

Careful Sociological Evaluation

- needs assessment, usability studies
- surveys, instrumentation

Testbed Status

Year 2 (Sep 95 ->)

- production pipeline for a few journals
- testing production database and components
    - Grainger Library (public terminals, usability studies)
    - Beckman Institute (physics), Computer Science Dept
- implement full Web version for deployment

Year 3: University of Illinois

Year 4: CIC Universities (midwest Big Ten)

goal is 100,000 documents & 100,000 users

Testbed Collection

- full SGML from Jan 1995 forward
- production on AIP (2000), APS, ASCE
- multi-year archive from IEEE, IEEE CS
- ISO 12083 DTD with figures, tables, equations
- publisher hands-on workshop Nov 16-17
- plans for publisher-maintained repositories
- problems with rendering scientific literature

Testbed Components

- Gathering: SGML from publishers
- Processing: normalize (federate) tags
- Indexing: store term/tag lists

- Control: VisBasic search interface
- Searching: OpenText fulltext engine
- Displaying: Panorama SGML viewer
- Fetching: Mosaic gets SGML/DTD files

federated SGML repository across the Net

Production Web Version

Client Interface

- multiple view interface (Java)
- session control across repositories

Network Gateway

- repository protocols (SQL, Z39.50)
- maintains state of search history

Server Search

- deposit canonicalized SGML documents
- index using DTDs for full-text retrieval

Multiple Views

- Different Levels of Search Interface
- Drag-and-Drop between views

- Integrates A& I (Abstracting and Indexing)
- Term Suggestion followed by Text Search
- Subject Thesaurus for coarse-grain suggest
- Concept Space for fine-grain suggest

- Visual Basic prototype, Java for multiplatform

New Web Servers

- HTTP servers evolve into Object Repositories
- NCSA Web server 2.0 released December

- Modular Steps towards Repositories
- Multiple Procotols (HTTP, Keep-Open)
- Security, Metadata Checking, Link Maintain

- Stateful Gateways support Distributed Sites
- Towards Sessions in Later Versions

Sociological Evaluation

different methodologies in granularity & scale

Needs Assessment

- ethnographic observations in libraries and labs
- focus groups and user interviews

Testbed Evaluation

- conceptual framework for evaluation
- planning for usability tests of pre-productions
- development of system instrumentation

Community for Social Studies of DLs

- Allerton Institute Oct 1995 on concepts & methods
- sessions at DL95 and DL96 on user research

Technology Research

towards the Interspace: Net correlation

- Scalable Semantic Retrieval

concept spaces and vocabulary switching

- Distributed Object Stores

secure object infrastructure (w/CNRI, Cornell)

- Analysis Environment Systems

correlation of information across repositories

Semantic Retrieval

automatic indexing of concepts

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

useful for interactive searching

- given a term, can suggest other terms
- merging concept spaces supports vocabulary switching

concepts require supercomputing

- concept space for INSPEC took 1 day on SGI Challenge
- co-occurrence matrix for 400K abstracts

Analysis Environment

- objects fine-grain manipulation
- navigation & grouping path recording
- retrieval & classification concept spaces
- correlations path matching via concept spaces

- prototype in Smalltalk, ObjectStore, ILU
- application in personal info, DLI, GIS

The 21st Century: Analysis

- Beyond Search to Analysis
- Cross-Correlating Information from many sources across the Net
- The Net solves problems

- Every community has its own special library
- Every community & every person does A& I !!

Go back to the Home Page

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

- useful for interactive searching
    - given a term, can suggest other terms
    - merging concept spaces supports vocabulary switching

- concepts require supercomputing
    - concept space for INSPEC took 1 day on SGI Challenge
    - co-occurrence matrix for 400K abstracts

Publishing Cycle

USER: request

LIBRARY: reference

INDEXER: classify

PUBLISHER: quality

AUTHOR: generate

- users are authors, computers are publishers
- every community has a repository
- a billion repositories on the Net !!

Vocabulary Switching

- fine-grained concept spaces

for every community and subcommunity

- user and collection modeling

choose domains for user and for search

- interactive vocabulary switching

intersect at common terms to suggest across domains

- supercomputers as time machines

personal computers same computations in 5-10 years

Switching Experiments

small-scale in molecular biology (JASIS)

- worms and flies
- 5000 documents generate each space
- 10 hours per space on a workstation
- “sperm” as connection term

large-scale in engineering (in progress)

- 3M abstracts from Compendex and Inspec
- 15 large domains of engineering (200K per space)
- 10 hours per space on a supercomputer
- “fluid dynamics” as connection term

Community Repositories

- User-driven Community Searching
  - choose topics (repositories) you know
  - choose topics you want to know
- vocabulary switch across domains
  - community specific term suggestion
- Interspace simulation with 1000 communities
  - Compendex partitioned by class codes
  - 3M abstracts and 1K spaces on Convex Exemplar

first crack in large-scale “semantic” retrieval

Computer-Assisted Indexing

- domain experts but classification amateurs
  - large community A& I is too general and too old
  - small community A& I is not consistent and much labor

- useful for interactive subject classification
    - automatic suggestions for potential classifications
    - domain expert culls list from “controlled” vocabulary

- semi-automatic support via concept spaces
  - concept dictionary of tag words from co-occurrence
  - tag frequency in documents determines classification

Building the Interspace

- every machine has its own information space
- every machine has its own concept space
- spaces for every user and every community

- search is matching selected objects
- relies on computer-assisted A& I

- analysis is merging community spaces
- vocabulary switch through graph intersect

The 21st Century: Analysis

- Beyond Search to Analysis
- Cross-Correlating Information from many sources across the Net
- The Net solves problems

- Every community has its own special library

[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

Introduction
Announcing the National Digital Library Competition

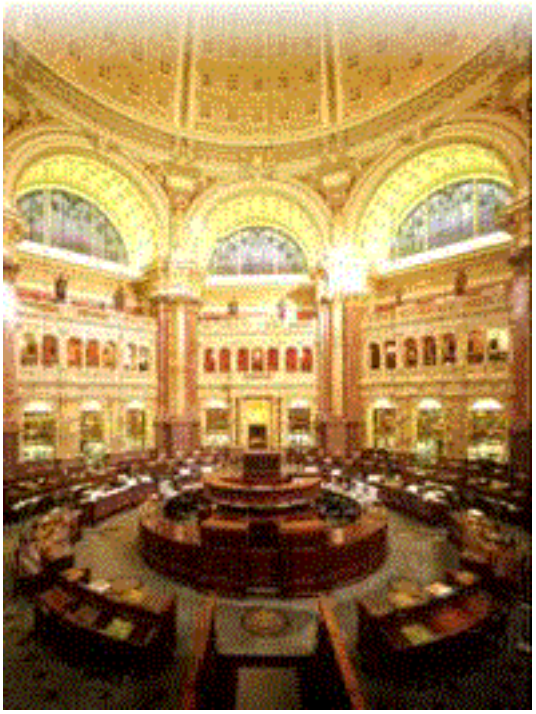
Summarized Project Guidelines

Awards

Application Process

Evaluation of Proposals

For More Information
Includes a recommended reading list



***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:

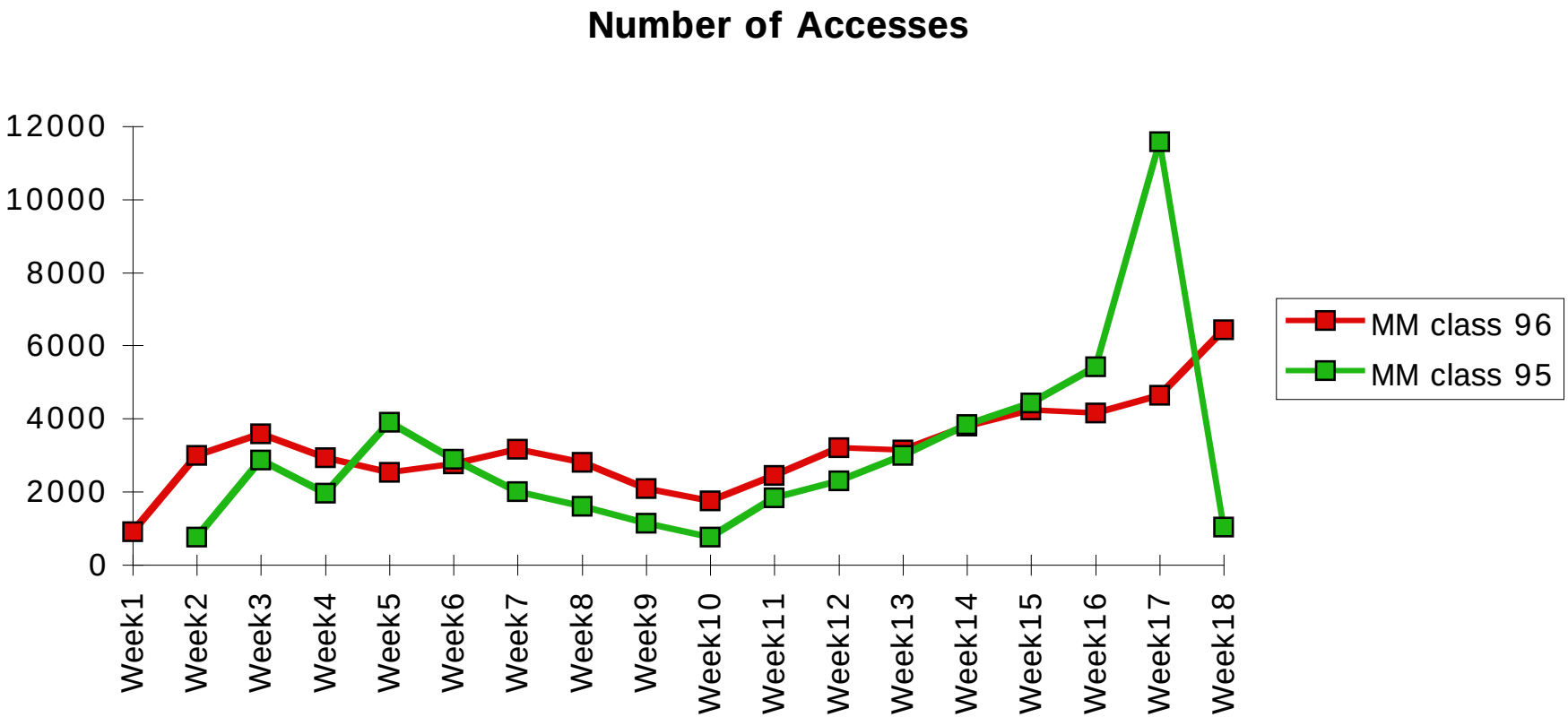
- Marc Abrams [URL: http://www.cs.vt.edu/vitae/Abrams.html]
  Computer Science Department
  Virginia Tech
  Blacksburg,VA 24061-0106 USA
  E-Mail:abrams@vt.edu

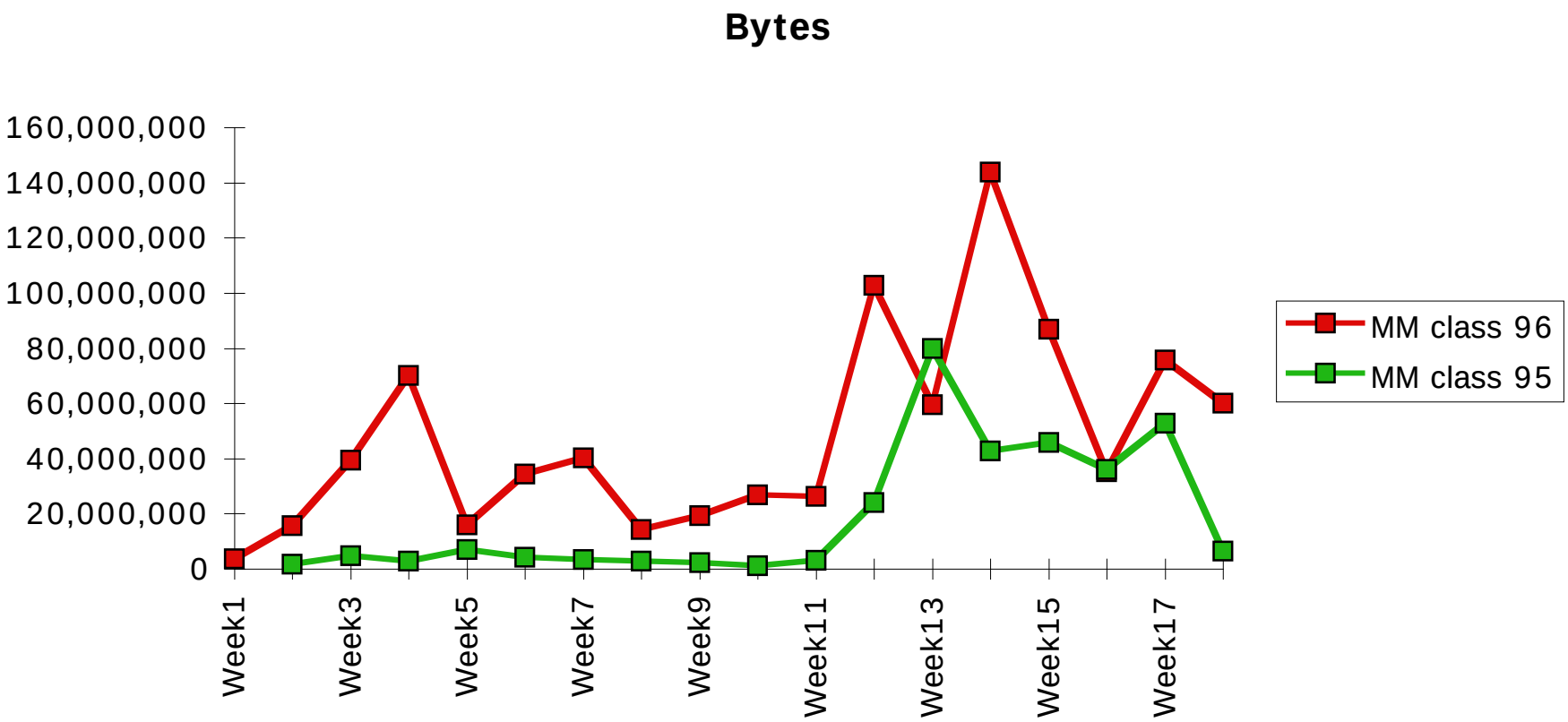
- Ghaleb Abdulla [URL: http://csgrad.cs.vt.edu/~abdulla/]
  Computer Science Department
  Virginia Tech
  Blacksburg,VA 24061-0106 USA
  E-Mail:abdulla@csgrad.cs.vt.edu

- 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:

- Z39.50 in a Nutshell - (An Introduction to Z39.50) by John A. Kunze & R. P. C. Rodgers. Written at National Library of Medicine, July 1995.
- The ANSI/NISO Z39.50 Protocol: Information Retrieval in the Information Infrastructure by William Moen (added here July 1995)
- IETF NIR document by Mark Needleman
- RFC1729: Using the Z39.50 Information Retrieval Protocol in the Internet Environment, by Clifford Lynch. December 16, 1994
- RFC1625: WAIS over Z39.50-1988, by M. St Pierre et al. June 1994
- Facilitating the Creation of Z39.50 Origins in the UK, by A.M.Addyman
- Z39.50 FAQ - very dated. A new Z39.50 FAQ is being developed.
- A list of Z39.50 available databases. This list is available as a Z39.50 database (at z3950.research.att.com) or via a WWW gateway.

There are several people/organizations working on Z39.50 -> WWW gateways.

- Prentiss Riddle is keeping a good page tracking WWW-to-Z39.50 Gateways. I recommend checking there rather than depending on this list!!
- The implementation being done by CNIDR is available for general use and free; this is part of the CNIDR software described above. The gateway is at cnidr.org.
- The Stanford gateway, with perl source code is supposed to be very useful and flexible.
- The Library of Congress has up a fairly complete set of resources at their gateway.
- The AT&T Library Network is also experimenting with gateways; to see this in action select here.

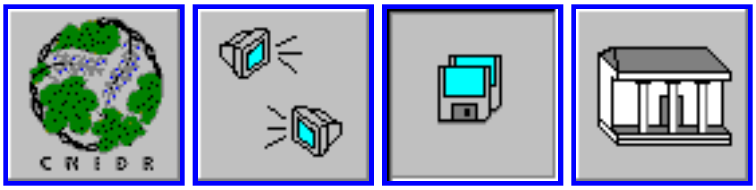
Profiles

Profiles are formal implementation agreements within the context of standard. There are two major profiles being worked on with the Open Systems Environment Implementors Workshop (OIW), the Government Information Locator Service (GILS), and the Wide Area Information Service (WAIS).

- The Office of Management and Budget, in concert with the Information Policy Committee of the Information Infrastructure Task Force, to promote the establishment of an agency-based Government Information Locator Service (GILS). GILS is intended to help the public locate and access public information throughout the U.S. government. Note GILS is based on Z39.50.
  The document is available via anonymous FTP in
  Microsoft Word for WIndows format, Word Perfect 5.0, Rich Text Format, and as ASCII text.
- The WAIS Profile of Z39.50 V2 specifies the required components of Z39.50-1994 for full WAIS functionality. Please send comments on the Profile to oiw-l@mozart.esl.com.
- A draft of the Geospatial Metadata Profile (GEO) is ready for review at the URL listed below. This profile is intended to be a guide to developers to support the attributes defined in the Content Standards for Digital Geospatial Metadata promoted by the U.S. Federal Geographic Data Committee.

Online views of Z39.50 related products/services

- Ameritech Library Services
- BookWhere for Windows - Z39.50 Client software
- Chemical Abstracts Service
- Data Research Associates, Inc
- Geac Computer Corporation, Ltd
- Library of Congress
- Online Computer Library Center, Inc (OCLC)
- The Research Libraries Group, Inc. (RLG)
- SIRSI Corporation
- VTLS, Inc.



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

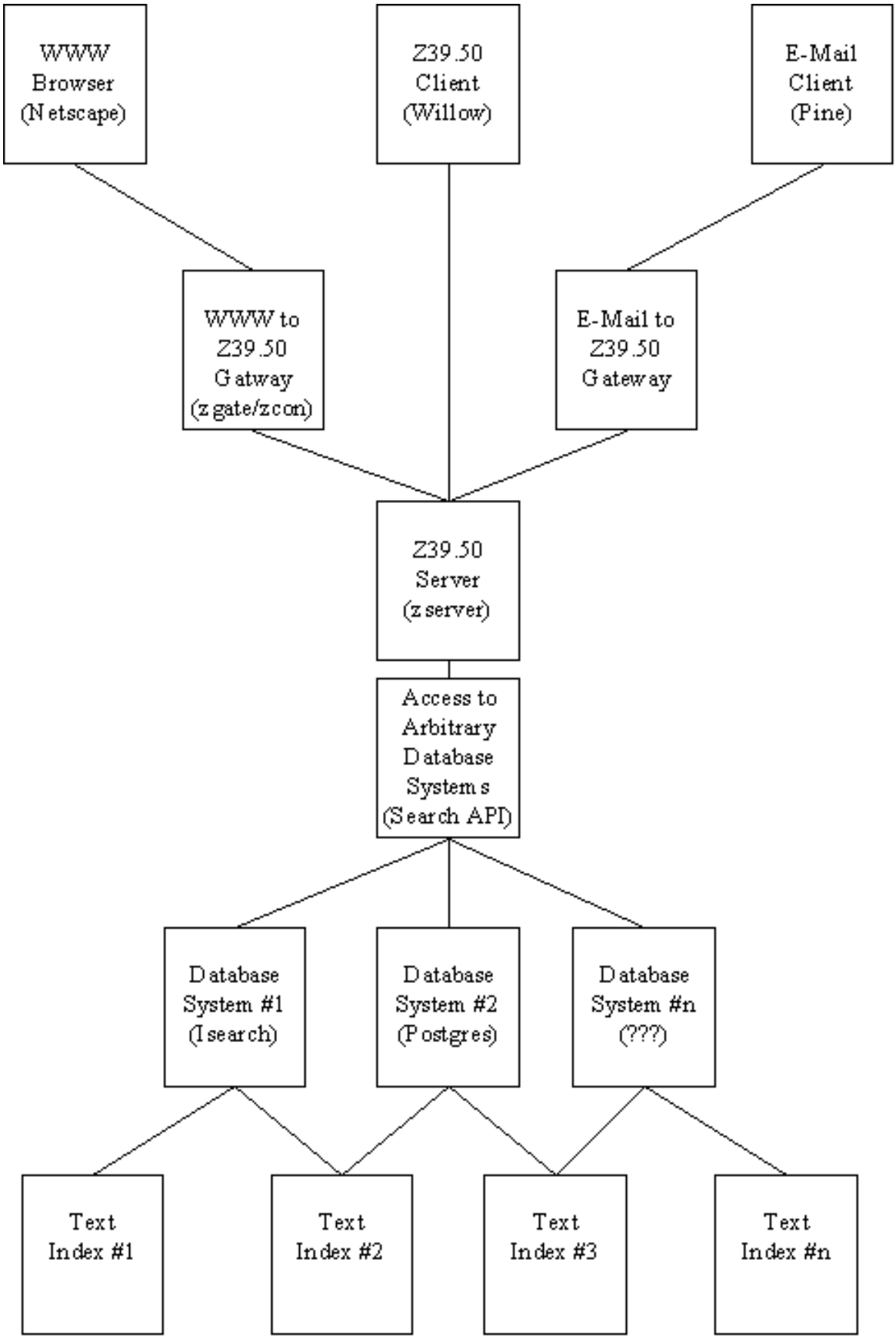
- **Help us to better serve you!**
- Diagram of Overall Architecture - Details available via Administrator's Guide below
- Demo of Stateful http to Z39.50 Gateway - Demonstrates access to various database systems
- Other systems using Isite
    - NASA Global Change Master Directory
        - Z39.50 Ranked 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

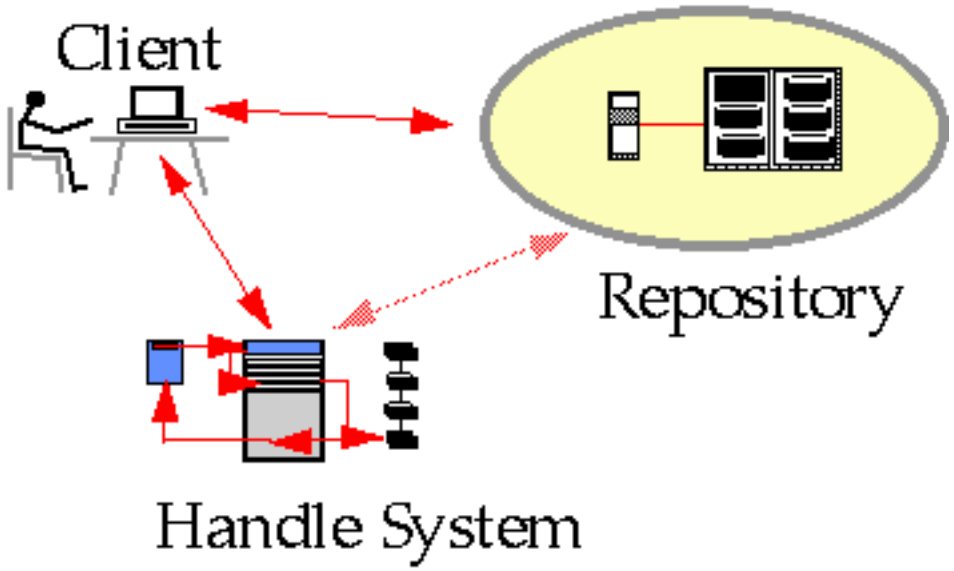
Digital Object Architecture Project

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

- Employment opportunities

- **Agents and ...**

  - Agents and the Knowledge Sharing Effort
  - Agents and security
  - Agents and learning
  - Agents and Ontologies.
  - Agents and Robots
  - Agents and artificial life
  - Agents and sex
  - Agent and humor.
  - Agents and virtual environments, muds, ...
  - Agents and other miscellaneous topics (e.g., IR)

- **Agents for ...**

  - Agents for Manufacturing
  - Agents for Commerce
  - Agents for human-computer interfaces
  - Agents for, on, and by the web.

AgentWeb is maintained at the UMBC Lab for Advanced Information Technology by **Tim Finin** (*finin@umbc.edu*).

Modified on Wednesday, 10-Jul-96 12:56:04 EDT -- 02980 hits since June 25, 1996

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

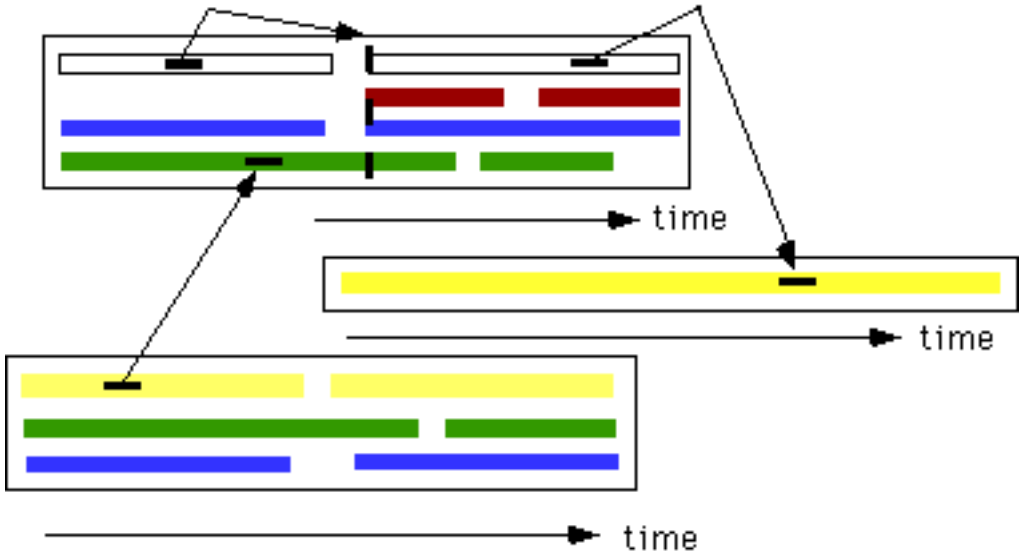
Other types of searches:
You may also combine words or phrases within a specified proximity, or locate segments such as sections where two words or phrases both occur.

Browse the *TEI Guidelines*.

- Bibliographic header of the TEI Guidelines
- Preface
- Acknowledgments
  - TEI Working Committees (1990-1993)
  - Advisory Board
  - Steering Committee Membership
- Changes from TEI P1 to TEI P3
- Part 1: Introduction
- Part 2: Core Tags and General Rules
- Part 3: Base Tag Sets
- Part 4: Additional Tag Sets
- Part 5: Auxiliary Document Types
- Part 6: Technical Topics
- Part 7: Alphabetical Reference List of Tags and Attributes
- Part 8: Reference Material

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



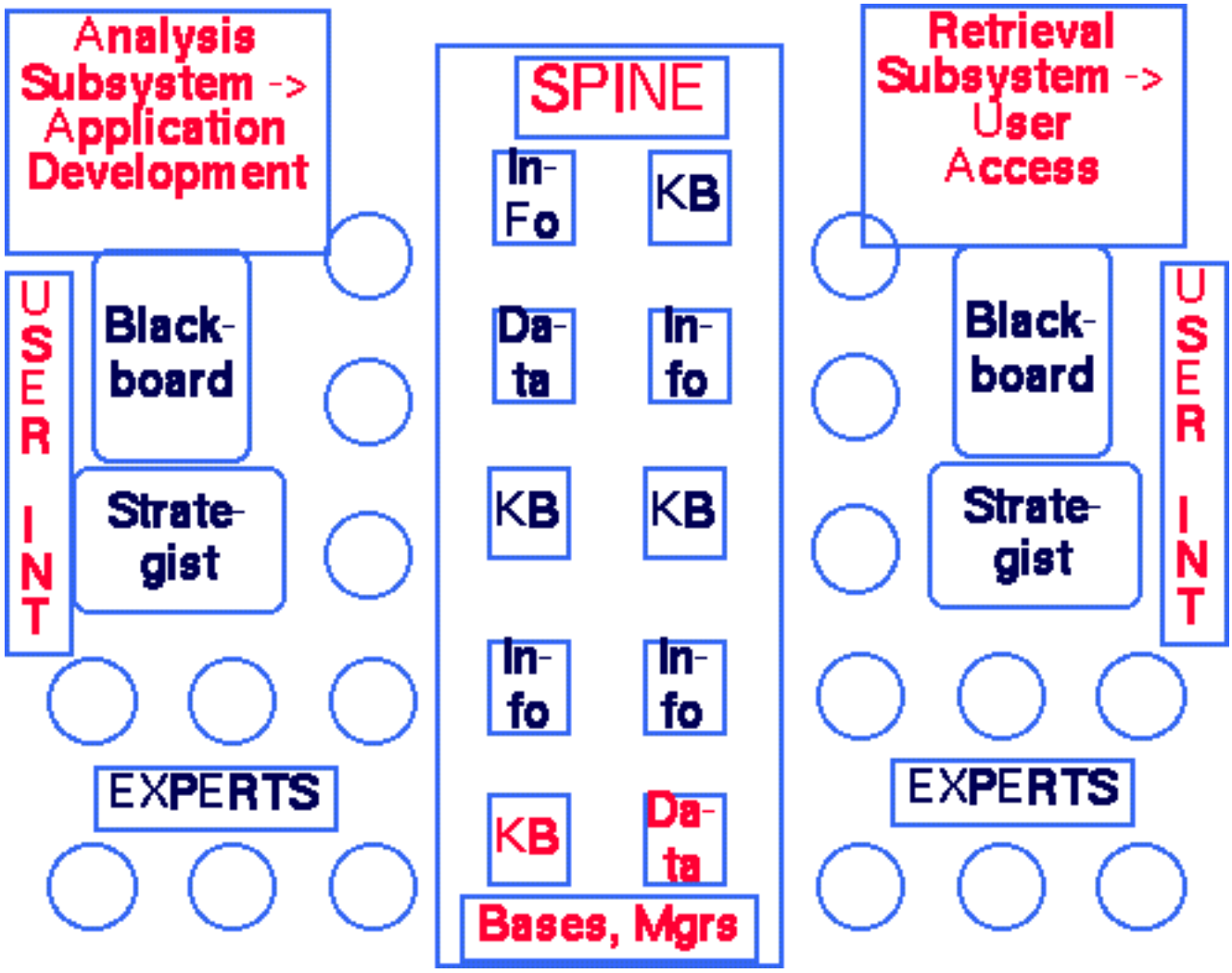
HYPERMEDIA COMPOSITES

[Adapted from HARD94 p. 53, Fig. 1c]

Models: Amsterdam - hypermedia

- Each of the 3 boxes represents a component or node that includes multimedia (time-based) information.
- When a component is presented, it *plays* for a period of time.
- The middle component is atomic, having only one media type. The black marker 2/3 of the way along is an anchor, which would allow one to *jump* directly to that point (in time / play).
- The other two components are composites, with several streams of multimedia information. The bottom one has three streams and six components, where components shown with the same color are played-back in sequence.
- The bottom component has one *from* anchor to a span in the first green-stream child-component of the top composite component.
- The top component is the most complex. About half-way through is a synchronization point, so one can jump there and all the streams will start then, at the same time. That point is reached from an anchor in the first white component. Other anchors point outside or are pointed to from outside.

COmposite Document Expert/extended/effective Retrieval (NSF: 1985-9)



Application Domains: 1985 -

- Electronic mail (AIList Digest, Collins English Dictionary)
- Navy messages (Naval Message Analyzer)
- Medical information on cardiology (400M)
- Campus Catalog - MARIAN (1G - underway)

Environment

- Communications: TCP/IP (supporting our own language for interprocess data/information/knowledge transfer)
- Operating Systems: versions of UNIX
- Programming Languages: C, C++, Prolog

MARIAN Mosaic Interface

To try out the MARIAN system you can use the MARIAN Mosaic interface which provides the top-ranked 30 items after searching against the Dec. 1993 version of the Virginia Tech catalog.

Overview

MARIAN (Multiple Access Retrieval of Information with ANnotations) is a system developed by the Virginia Tech Computing Center starting in 1991. It runs on a collection of NeXT computers, using one or more threads for each user session. A similarity value is computed for each field (e.g., title, author, subject). The combiner module computes an overall similarity that is used to rank the documents, so users see the top-ranked items only.

Using the NeXT interface, one can call for successive sets of 30 items. Also, one can request circulation data that is obtained from the VTLS computer by way of an expert system analysis module.

In addition there is a Gopher+ interface, one using the curses interface package and access using telnet.

In 1995 the MARIAN system should be opened for wider use, after the data is brought into synchronization with the current contents of the VTLS computer.

History

One precursor of MARIAN is the CODER system. Many of the ideas from it, the interprocess communication approach, and the English lexicon developed for it, are used in MARIAN.

The other precursor of MARIAN is the REVTOLC study --- Retrieval Experiment, Virginia Tech On Line Catalog. A pilot study was done with 300,000 records and 52 users. The full study is reported at length in the 1993 Ph.D. dissertation of Amjad Daoud. Students preferred our approach to VTLS. They also preferred ranked retrieval to standard Boolean retrieval. Details follow:

- Experiment with 500,000 records
- 4 methods compared
    - Boolean queries
    - Extended Boolean interpretation
    - Vector queries
    - Vector queries + probabilistic feedback
- Each person tried 2 methods, 2 queries/method
- 12 groups: 4 choices for method 1, 3 left for method 2
- 18 queries: taken from ones recorded in library
- Tested with 216 users, each spending 60-90 minutes

Welcome to

ARIAN

Searching the Virginia Tech Library Collection

Query

Quit MARIAN | Help

Preferences

*** NEWS ***

New format for document close-ups.

MARIAN is now using a new display format for individual items.
Use the "Comment" feature to tell us what you think of it.

Some more data is shown for each work in the libraray collection. More importantly, the information has been set up to be easier on the eye. The order has also been changed, for instance to bring all the authors (including illustrators, performers,and so forth) into a single place. We hope you find this organization sensible.

The short descriptions in the list of results from a search have been slightly cleaned up too.

MARIAN is a search system for library catalog data. The system you have reached searches the Virginia Tech Library collection, updated regularly.

This page and the WWW gateway to MARIAN are both under construction.

It is always better to use the buttons on MARIAN pages than to move back through your Web browser history pages and start from a page there. That's because the MARIAN gateway does not always interact well with pages cached in your Web browser. This page should be an exception, but even it may not be. If the gateway (or your browser) gets confused, reconnect to this URL and start again.

The MARIAN system is in the final phase of development. Things are pretty trustworthy, but the occassional failure is to be expected. As a rule of thumb, if your results look anything like what you expected, they are probably completely correct. If they are completely off-the-wall, or if nothing comes back, the system has probably dropped a ball. Try again later.
