

Digital Libraries: Topical Outline

- [Section 1. Foundations](#)
 - [Early visions](#), [definitions](#), [samples/examples](#), [resources/references](#), [projects](#)
- [Section 2. Search, Retrieval, Resource Discovery](#)
 - [Information storage and retrieval](#), [Boolean vs. natural language](#), [search engine tutorial](#)
 - Indexing: Phrases, Thesauri ([on web](#)), Concepts
 - [Federated search](#) and harvesting, [OAI](#), Crawlers/[spiders](#)
 - [Integrating links](#) and ratings, [fusion](#)
- [Section 3. Multimedia, Representations](#)
 - Text/audio/image/video/graphics/animation
 - Capture, Digitization, Compression
 - Standards, Interchange: [JPEG](#), [MPEG](#)
 - Content-based retrieval, Playback (e.g., [Real](#)), QoS, [SMIL](#)
- [Section 4. Architectures](#)
 - Modular/componentized, Protocols
 - InfoBus ([Stanford](#), [Java](#)), Mediators, Wrappers ([TSIMMIS](#))
- [Section 5. Interfaces](#)
 - Workflow, Environments, Taxonomy of interface components, Visualization
 - Design, Usability testing
- [Section 6. Metadata](#)
 - Ontologies, [RDF](#)
 - [MARC](#), [Dublin Core](#), [IMS](#)
 - Mappings, [Crosswalks](#)
- [Section 7. Electronic Publishing, SGML, XML](#)
 - Authoring, Presenting, Rendering, [Document Object Model \(DOM\)](#)
 - Dual-publishing, Styles ([XSL](#)), eBooks (e.g., [eBooks.com](#), [eBooks Central](#), [netLibrary](#))
 - Structure, Semi-structured information, Tagging/markup, Structure queries
- [Section 8. Database Issues](#)
 - Extending database technology
 - Structured and unstructured information
 - Multimedia databases, Link databases
 - Performance/replication/storage, e.g., [Internet2 Distributed Storage Infrastructure \(I2-DSI\)](#)
- [Section 9. Agents](#)
 - Distributed issues
 - Protocols, Negotiation
 - [Webbots](#) (automatic indexing)
- [Section 10. Commerce, Economics, Publishers](#)
 - Preservation and archives: [DLF page](#), [PADI \(AU\) page](#), [Besser on moving images](#)

- Terms and conditions, Open collections, Self-archiving
 - Economic models, [Micropayments](#)
- [Section 11. Intellectual Property Rights, Security](#)
 - Legal issues, e.g., [Gladney on digital preservation archiving and copyright](#)
 - Copyright, Rights management
- [Section 12. Social Issues](#)
 - Cooperation and collaboration, Ratings, Annotation ([PICS](#))
 - Educational applications ([NSDL](#)), [Digital divide](#)
 - Museums ([AMICO](#)), Cultural heritage, International concerns
 - Organizational acceptance/issues, Personalization

(c) 2000, 2001 Edward A. Fox, all rights reserved

Introduction to Digital Libraries:

- [Definitions](#): Some of the attempts made by various people to define a digital library.
- [Sample DLs](#): Illustrations of what is or may not be a digital library
- [Foundations](#): Introductory material related to digital libraries...
- [Scenarios and Perspectives](#): Various scenarios and perspectives that arise in a Digital Library context.

[\[Main\]](#) [\[Contents\]](#)

Please send comments/suggestions to [Ed Fox](#). (c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Definitions :

- "The new digital libraries will have features not possible in traditional libraries, thereby extending the concept of library far beyond physical boundaries. They will provide innovative resources and services. One example is the ability to interact with information: rather than presenting a reader with a table of numbers, digital libraries allow users to choose from a variety of ways to view and work with the numbers, including graphical representations that they can explore. With the extensive use of hypertext links to interconnect information, digital libraries enable users to find related digital materials on a particular topic."
([2001 PITAC Report](#), "Digital Libraries: Universal Access to Human Knowledge", p. 3)
- "Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities."
([Digital Library Federation](#))
- "Digital libraries are complex data/information/knowledge (hereafter information) systems that help: satisfy the information needs of users (societies), provide information services (scenarios), organize information in usable ways (structures), manage the location of information (spaces), and communicate information with users and their agents (streams)."
(Edward A. Fox, July 1999, according to 5S Framework)
- "Digital library work occurs in the context of a complex design space shaped by four dimensions: community, technology, services and content"
(Gary Marchionini and Edward A. Fox, "Progress toward digital libraries: augmentation through integration", pp. 219-225, guest editors' introduction to "Progress Toward Digital Libraries", eds. Gary Marchionini and Edward A. Fox, Special Issue, *Information Processing & Management*, 35(3), May 1999.)
- "The field of digital libraries deals with augmenting human civilization through the application of digital technology to the information problems addressed by institutions such as libraries, archives, museums, schools, publishers and other information agencies. Work on digital libraries focuses on integrating services and better serving human needs, through holistic treatment irrespective of interface, location, time, language and system. Although substantial collections may be created solely for the use of individuals, we consider sharable resources one of the defining characteristics of libraries. Libraries connect people and information; digital libraries amplify and augment these connections."
(Gary Marchionini and Edward A. Fox, "Progress toward digital libraries: augmentation through integration", *Information Processing & Management*, 35(3):219-225, May 1999.)
- For a thoughtful discussion of definitions, approaches, and community perspectives on "digital libraries" see "What are digital libraries? Competing visions" by Christine L. Borgman, pp. 227-

244, in "Progress Toward Digital Libraries", eds. Gary Marchionini and Edward A. Fox, Special Issue, *Information Processing & Management*, 35(3), May 1999.

- "The Digital Library is:
 - The collection of services
 - And the collection of information objects
 - That support users in dealing with information objects
 - And the organization and presentation of those objects
 - Available directly or indirectly
 - Via electronic/digital means."

([The Scope of the Digital Library](#), Draft Prepared by Barry M. Leiner for the DLib Working Group on Digital Library Metrics, 1998)

- "Digital library is a concept that has different meanings in different communities. To the engineering and computer science community, digital library is a metaphor for the new kinds of distributed data base services that manage unstructured multimedia data. To the political and business communities, the term represents a new marketplace for the world's information resources and services. To futurist communities, digital libraries represent the manifestation of Wells' World Brain. The perspective taken here is rooted in an information science tradition." ([Research and Development in Digital Libraries by Gary Marchionini, 1998](#))
- "an organized data base of digital information objects in varying formats maintained to provide unmediated ease of access to a user community, with these further characteristics:
 - an overall access tool (e.g. a catalog) provides search and retrieval capability over the entire data base;
 - organized technical procedures exist through which the library management adds objects to the data base and removes them according to a coherent and accessible collections policy."
 (Peter Graham, Rutgers University Libraries, 1997)
- "Digital libraries are a set of electronic resources and associated technical capabilities for creating, searching, and using information. In this sense they are an extension and enhancement of information storage and retrieval systems that manipulate digital data in any medium (text, images, sounds; static or dynamic images) and exist in distributed networks. The content of digital libraries includes data, metadata that describe various aspects of the data (e.g., representation, creator, owner, reproduction rights), and metadata that consist of links or relationships to other data or metadata, whether internal or external to the digital library." ([1996 UCLA-NSF Social Aspects of Digital Libraries Workshop](#))
- Digital libraries are constructed -- collected and organized -- by a community of users, and their functional capabilities support the information needs and uses of that community. They are a component of communities in which individuals and groups interact with each other, using data, information, and knowledge resources and systems. In this sense they are an extension, enhancement, and integration of a variety of information institutions as physical places where resources are selected, collected, organized, preserved, and accessed in support of a user community. These information institutions include, among others, libraries, museums, archives, and schools, but digital libraries also extend and serve other community settings, including classrooms, offices, laboratories, homes, and public spaces." ([1996 UCLA-NSF Social Aspects of Digital Libraries Workshop](#))

- "Systems providing a community of users with coherent access to a large, organized repository of information and knowledge."
([Clifford Lynch](#), 1995)
 - "systems providing a community of users with coherent access to a large, organized repository of information and knowledge. This organization of information is 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"
(adapted from [Interoperability, Scaling, and the Digital Libraries Research Agenda, report of the 1995 IITA DL Workshop](#))
 - "A library that has been extended and enhanced by the application of digital technology. Important aspects of the digital library that may be extended and enhanced include :
 - Collections of the library
 - Organization and management of the collections
 - Access of the library items and the processing of the information contained in the items
 - Communication of information about the items "(Terry Smith, UCSB, 1995)
 - "The generic name for federated structures that provide humans both intellectual and physical access to the huge and growing worldwide networks of information encoded in multimedia digital formats."
([The University of Michigan Digital Library: This Is Not Your Father's Library](#), [Bill Birmingham](#), 1994)
 - "A digital library is a distributed technology environment which dramatically reduces barriers to the creation, dissemination, manipulation, storage, integration, and reuse of information by individuals and groups."
([Edward A. Fox](#), editor, [Source Book on Digital Libraries](#), 1993, pg. 65)
 - "A digital library is a machine readable representation of materials which might be found in a university library together with organizing information intended to help users find specific information. A digital library service is an assemblage of digital computing, storage, and communicate machinery together with the software needed to reprise, emulate, and extend the services provided by conventional libraries based on paper and other material means of collecting, storing, cataloging, finding, and disseminating information."
([Edward A. Fox](#), editor, [Source Book on Digital Libraries](#), 1993, pg. 65)
-

Digital Library related terms/glossary

(by Peter Graham, Rutgers University Libraries, 1997):

- digital archive: a digital library which is intended to be maintained for a long time, i.e. periods longer than individual human lives and certainly longer than individual technological epochs. (Sometimes formerly also "digital research library.")
- digital preservation: preservation of artifactual information by digitizing its image (e.g. scanning a manuscript page, digitally photographing a vase, or converting a cylinder recording to digital form).
- electronic preservation: preservation of information that is in digital (that is, electronic) form, i.e. the techniques associated with refreshing, migration and assurance of integrity.

Digital Preservation techniques:

- Refresh: to copy digital information from one long-term storage medium to another of the same type, with no change whatsoever in the bit stream (e.g. from a decaying 800 bpi tape to a new 800 bpi tape, or from an older 5 1/4" floppy to a new 5 1/4" floppy).
- "Modified refreshing" is the copying to another medium of a similar enough type that no change is made in the bit pattern that is of concern to the application and operating system using the data, e.g. from an 800 bpi tape to a 1600 bpi tape or to a "square", cartridge, tape; or from a 5 1/4" floppy disk to a 3 1/2" floppy disk.
- Migrate: to copy data, or convert data, from one technology to another, whether hardware or software, preserving the essential characteristics of the data; generally forward in time. (At the moment, it is recognized, this final qualifier begs many questions.) Examples: conversion of XyWrite w/p files to Microsoft Word; conversion of ClarisWorks v3 spreadsheet files to Microsoft Excel v4 files; conversion of binary tape images of survey research

multi-punched tab cards to a data base format; copying an 800 bpi tape file to a sequential disk file; converting a DOS FoxPro data base to a Visual Basic database for Windows 95; converting a PICT image to a TIFF image; converting a ClarisWorks for Windows v4 w/p file to a Macintosh ClarisWorks v4 file.

Examples can be given, as here, for cases known to be required; the longer term preservation problem is to prepare for forward migrations when the future technologies are unknown.

- Emulate: (find and use better Comp SCI terms here, probably) in hardware terms, the creation of software for a computer that reproduces in all essential characteristics (as defined by the problem to be solved) the performance of another computer of a different design. Computers may emulate earlier computers in order to provide backward compatibility, or may emulate a future computer in order to provide a software development environment while the newer computer is still being fabricated.

In software preservation terms, the creation of software that analyzes the software environment of

a document such that it can provide a user interface to the document that substantially reproduces the essential characteristics of the document as it was created by its originating software.

- Document: (use sense that Apple began to use, with Macintosh; anything manipulated by an application; find their definition and build on it. Note Dublin Core [and other] use of "document like object").
- Authenticate: of users, to verify that network users are in fact who they identify themselves to be; of documents, to validate the integrity of a document with respect to its original authorized creation.
- Authentication: (of a resource--i.e. of data, not people)
- Authenticity: (of a resource--i.e. of data, not people)
- Integrity: synonym of authenticity (of a resource--i.e. of data, not people)

[\[Main\]](#) [\[Introduction\]](#) [\[Contents\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Sample DLs: Illustrations of what is or may not be a digital library

Sites to consider that demonstrate DL functions

- [Amazon](#)
- [Pricewatch](#)
- [PriceScan](#)
- [Internet Movie Database](#)
- [Web Characterization Repository](#)
- [ETDs](#)
and the experimental [ETD union collection](#)
- [NCSTRL](#) (being redone)

Sites related to NSDL

- [NSDL](#)
- [SMETE.ORG](#)
- [DLESE](#)
- [iLumina](#)
- [CSTC](#)
and a [test version](#)
- [ResearchIndex](#)
- [CITIDEL](#)
- [ENC](#)
- [MERLOT](#)
- [Mathwright](#)
- [MathDL](#)

[\[Main\]](#) [\[Contents\]](#) [\[Foundations\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 2001, Edward A. Fox

Foundations (see Lesk Ch. 1, 8):

- [As We May Think](#) by Vannevar Bush - the visionary article that helped motivate early work on digital libraries, hypertext and information retrieval
 - 1996 UCLA workshop (focusing on user perspectives):
 - [Introduction](#)
 - [information life cycle](#)
 - [Artists](#)
 - [Business Records as Artifacts](#)
 - [Health-Information Systems](#)
 - 1995 IITA workshop: [Definitions and Roles of Digital Libraries](#)
 - [Digital Libraries: Issues and Architectures](#)
 - [Digital Library: Gross Structure and Requirements: Report from a March 1994 Workshop.](#)
-

Pedagogy:

We recommend that the above items be skimmed to obtain a general background regarding digital library research, development, and practice. Please also read chapters 1 and 8 of Dr. Lesk's book.

[\[Main\]](#) [\[Contents\]](#) [\[Introduction\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2000, Edward A. Fox, Rajat Gupta

Defining Scenarios & Perspectives:

[1995 IITA Workshop](#):

- [Publishing](#)
 - [Commercial](#)
 - [Library](#)
 - [Internet](#)
 - [Multimedia](#)
-

Pedagogy:

We recommend that the scenarios given be examined, especially for the group in which the reader fits.

[\[Main\]](#) [\[Contents\]](#) [\[Introduction\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Resources:

- [Projects](#)
 - [People](#)
 - [Countries and regions](#)
 - [Centers, sites and organizations](#)
-

[\[Main\]](#) [\[Contents\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. fox, Rajat Gupta

Projects:

DLI-2

- [DLI-2 home page at NSF](#)
- [DLI-2 projects funded from 1998-2000 submissions](#)
- D-Lib Magazine articles on DLI-2 by NSF etc.:
 - [FY 1999 Awards - S. Griffin](#)
 - [Commentary on DLI-2 - M. Lesk](#)
 - [NSF/JISC Int'l Initiative - N. Wiseman, C. Rusbridge, S. Griffin](#)
- [Selected abstracts of IIS awards \(including some DLI-2\)](#)
- Calls:
 - [NSF9863 - Digital Libraries Initiative - Phase 2 \(February 20, 1998\)](#)
 - [Addendum - Special Emphasis: Planning Testbeds and Applications for Undergraduate Education within the Digital Libraries Initiative - Phase 2](#)
 - [NSF996 - International Digital Libraries Collaborative Research \(November 9, 1998\)](#)

DLI-1

- DLI-1 home page at [NSF](#) and older one at [U. Illinois](#)
- [DLI-1 publications](#)
- [Carnegie Mellon University](#)
- [Stanford University](#)
- [University of California at Berkeley](#)
- [University of California at Santa Barbara](#)
- [University of Illinois](#)
- [University of Michigan](#)

[Library of Congress](#) and its [American Memory Project](#)

Los Alamos and U. Ghent, SFX: [paper](#) and articles in D-Lib Magazine: parts [1](#), [2](#), [3](#); [OpenURL Framework](#) (and [NISO standard effort](#))

[NARA](#) - National Archives and Records Administration

NASA [JSC Digital Image Collection](#)

NSDL (National Science, Mathematics, Engineering, and Technology Education Digital Library)

Related Sites and Projects:

- [Under Construction NSDL site](#)
 - [DLI-2 Planning Testbeds and Applications for Undergraduate Education](#)
 - [SMETE-Lib Study - NSF Science Mathematics, Engineering and Technology Education Digital Library reports](#)
 - [Funded Projects](#)
 - SMETE Information Portal: <http://www.smete.org>
 - [NEEDS - National Engineering Delivery System](#)
 - [Project Kaleidoscope](#)
 - Geoscience: [Call](#); [DLESE](#) (Digital Library for Earth System Education); [Windows to the Universe](#)
 - [ODU project](#) (including buckets)
 - U. Texas Austin: [Technology for Education 2000](#); [Virtual Multimedia Exams in Physical Anthropology](#); [High Res X-ray CT \(Computed Tomography\) Facility](#)
 - [Computer Science Teaching Center \(CSTC\)](#)
-

Selected International Efforts

Australia: [National Library DL Initiatives](#)

[Bibliotheca universalis](#): (G7)

[British Library DL Programme](#)

[CIDL](#) - Canadian Initiative on Digital Libraries

Electronic Theses and Dissertations Initiative: [NDLTD project](#), [Collection](#), [Submission Instructions](#)

[ERCIM](#): [DL initiative](#) (DELOS)

International Digital Libraries Association: [IDLA home page](#)

International Fed. of Library Associations and Institutions - [IFLA](#): [page pointing to DL info](#)

Japan:

- [Workshops - DLnet](#)
- National Museum of Ethnology - [MINPAKU](#): [Virtual Tour](#)
- [Kobe U.](#): [Digital Library Search](#), [TITAN Search using WWW](#)
- [Tokyo Inst. of Technology](#): [Library](#)
- [Kyoto U.](#): [Digital Library](#)
- [NAIST](#): [Digital Library](#)
- [ULIS](#): [Digital Library](#), [Multilingual HTML](#), [Multilingual folk tales](#)
- [University of Tsukuba](#): [Digital Library](#)

[MeDOC](#): (German Online Computer Science Library)

NSF-EU Working Groups and Meetings: [home page](#)

Singapore Network: [SINGAREN](#)

[UK Electronic Library Programme](#) including a project on preservation: [New Cedars Project: CURL Exemplars in Digital Archives](#) and a 13M record searchable OPAC called [COPAC](#); [Centre for DL Research](#) (U. Southampton); De Montfort U. former [International Institute for Electronic Library Research](#)

Selected Publisher / Information-Distributor Projects:

- [ACM DL](#)
 - [ProQuest \(UMI\)](#) and its [Digital Dissertations](#)
 - [Elsevier Science \(ScienceDirect, ...\)](#)
 - [IDEAL](#) (International Digital Electronic Access Library)
 - [IEEE-CS DL](#)
 - [OCLC](#) FirstSearch Electronic Collections Online
 - [Springer's Forum for Science](#) (The LINK Online Libraries)
-

Industrial Projects:

- [NEC: ResearchIndex \(CiteSeer\)](#)
 - [OCLC Research Projects](#)
-

Virginia Tech Projects:

- [Interactive Courseware on Digital Libraries](#) (this site itself is a part of it)
 - **Interactive Learning with a Digital Library in CS** <http://ei.cs.vt.edu/>
 - Interactive Learning with a Digital Library in CS arch <http://ei.cs.vt.edu/~cs5604/Adv/Adv-ILDLCS.html>
 - Courseware <http://ei.cs.vt.edu/courses.html>
 - [Project Overview](#) (for FIE'96, in PDF)
 - [Project Interim Report](#), Oct. 1996, PDF
 - [Project Report for NSF EI PI Meeting](#), Nov. 1996, PDF
 - **Envision (CS literature)** <http://ei.cs.vt.edu/~cs5604/Adv/Adv-Envision.html>
 - Envision report <http://ei.cs.vt.edu/papers/ENVreport/final.html>
 - **CODER** <http://ei.cs.vt.edu/~cs5604/Adv/Adv-CODER.html>
 - **MARIAN**
 - [home page](#)
 - system <http://opac3.cc.vt.edu/htbin/marian>
 - old overview <http://ei.cs.vt.edu/~cs5604/Adv/Adv-MARIAN.html>
 - [CSTC - Computer Science Teaching Center](#) and related effort
 - [CRIM - Curriculum Resources Interactive Multimedia](#)
 - [W3C Web Characterization Repository](#) (of logs, traces, tools, papers)
 - Virginia Tech DL Superstorage Research, using [VT-PetaPlex-1](#), a [PetaPlex](#) system from [Knowledge Systems Inc.](#) with at least 100 processors and 2.5 terabytes
-

Approaches to DL:

- Build upon existing electronic materials
 - Netlib (numerical analysis) <http://www.netlib.org/> and its search: http://www.netlib.org/utk/misc/netlib_query.html
- Build upon publishers collections
 - AAAS - Science Online <http://www.aaas.org/>

- ACM DL <http://www.acm.org/dl/>
- ACS (Chemistry) - Online <http://www.acs.org/>
 - CORE Overview <http://ei.cs.vt.edu/~cs5604/DL/DL2.html>
 - D-Lib Magazine, Dec. 1995, Making a Digital Library, Chemistry Online Retrieval Experiment <http://www.dlib.org/dlib/december95/briefings/12core.html>
 - CORE at OCLC <http://www.oclc.org/research/publications/arr/1994/part2/xscepter.htm>
- Elsevier
 - ScienceDirect <http://www.elsevier.nl/>
 - TULIP (material science & engineering) homepage <http://www.elsevier.nl/inca/homepage/about/resproj/tulip.shtml>
 - With universities + OCLC
- [Highwire Press](#)
- [IEEE](#)
- [IEEE-CS DL](#)
- [JSTOR](#)
- Commercial services and systems
 - IBM <http://www.software.ibm.com/is/dig-lib/>
 - Version 2 <http://www.software.ibm.com/is/dig-lib/v2factsheet/>
 - collection treasury <http://www.software.ibm.com/is/dig-lib/treasury/>
 - images - QBIC <http://www.qbic.almaden.ibm.com/>
 - news archive <http://www.software.ibm.com/is/dig-lib/newsarchive/>
- Enhance WWW (hypertext):
 - HyperWave <http://www.hyperwave.de/>
 - HyperWave [information server](#)
 - HyperWave author <http://www2.iicm.edu/hyperwave/author>
 - HyperWave author features <http://www2.iicm.edu/hyperwave/author/features.html>
 - HyperWave author specs <http://www2.iicm.edu/hyperwave/author/specifications.html>
 - Harmony <http://www2.iicm.edu/harmony>
 - Harmony screens <http://ei.cs.vt.edu/~cs5604/Adv/Adv-Harmony.html>
 - Amsterdam model <http://ei.cs.vt.edu/~mm/gifs/Amsterdam-hm.html>
- Community network multimedia history
 - BEV <http://www.bev.net>
 - BEV History <http://history.bev.net/bevhist/>
 - Timeline <http://history.bev.net/bevhist/historyBase/mainTimeline.html>
 - [Screen for Spring 1992](#)
 - [Screen for Article](#)
- Discipline - Greek Literature <http://www.perseus.tufts.edu/>
 - Evaluation - [article in TOIS](#)
- Discipline - Computer Science

- Technical reports
 - [WATERS](#) - through 1995
 - CSTR <http://WWW.CNRI.Reston.VA.US/home/cstr.html>
 - NCSTRL <http://www.ncstrl.org/>
 - Search results, Search results abstract
 - Doc. thumbnails, Doc. page 1
 - CoRR: <http://xxx.lanl.gov/archive/cs/intro.html>
- Ptrs
 - DLs for CS <http://fox.cs.vt.edu/DLCS.html>
 - Results page, document page from search
- Genre - ETDs - electronic theses and dissertations
 - Virginia Tech <http://etd.vt.edu/>
 - Submission form <http://scholar.lib.vt.edu/ETD-db/ETD-submit/login>
 - Approval form <http://etd.vt.edu/submit/ETDapp09-00.pdf>
 - Letter to students <http://etd.vt.edu/guidelines/>
 - Standards <http://etd.vt.edu/help/multimedia.html>
 - Collection <http://www.theses.org>
 - [Federated Search](#)
 - Project - Networked Digital Library of Theses and Dissertations <http://www.ndltd.org>
 - Brief description <http://www.ndltd.org/info/descr.htm>
 - D-Lib Magazine Overview September 1996
<http://www.dlib.org/dlib/september96/theses/09fox.html>
 - D-Lib Magazine Update September 1997
<http://www.dlib.org/dlib/september97/theses/09fox.html>
 - D-Lib Magazine Federated Search September 1998
<http://www.dlib.org/dlib/september98/powell/09powell.html>
 - FIPSE (US Dept. of Education) funding of 1996-1999 project
 - proposal abstract <http://www.ndltd.org/support/fipseabs.htm>
 - proposal full-text <http://www.ndltd.org/support/fipse10.pdf>
 - project final report ([PDF](#))

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta



DIGITAL LIBRARIES INITIATIVE

a community of
researchers and
agencies working
together to bring the
world's knowledge
to your desktop

Digital Libraries Initiative
Phase 2 HOME

Digital Libraries Initiative
Phase 1 (1994-1998)

Search

Highlight

[Presentations from the DLI2/IMLS Principal Investigators Meeting, Roanoke, VA, June 28, 2001](#)

[DLI2 LIST- a mailing list archival site for the digital libraries community](#)

[PITAC renewed; three new reports available](#)

Features

[DLI2 Funded Projects](#)

[DLI2 International Projects](#)

[Special Projects Program](#)

[Funded Workshops](#)

Related Information

[Glossary](#)

[News](#)

[Events](#)

[Recent Articles](#)

[Reports](#)

[Publications](#)

[National SMETE Digital Library](#)

[DL Resources](#)

[D-Lib Magazine](#)

[e-Culture Newsletter](#)

Sponsoring Agencies and Programs

National Science Foundation ([NSF](#))

[Digital Libraries Initiative](#)

Defense Advanced Research Projects Agency ([DARPA](#))

[Information Technology Office](#)

National Library of Medicine ([NLM](#))

[Extramural Programs](#)

Library of Congress ([LOC](#))

[Digital Library Initiatives](#)

National Endowment for the Humanities ([NEH](#))

[Digital Library Initiative](#)

National Aeronautics & Space Administration ([NASA](#))

In Partnership with

[National Archives and Records Administration](#) (NARA)

[Smithsonian Institution](#) (SI)

[Institute of Museum and Library Services](#) (IMLS) [IMLS PROJECTS](#)

[NSF Contact](#)

[Agency Contacts](#)

Digital Libraries Initiative Phase Two is a multiagency initiative which seeks to provide leadership in research fundamental to the development of the next generation of digital libraries, to advance the use and usability of globally distributed, networked information resources, and to encourage existing and new communities to focus on innovative applications areas.

Since digital libraries can serve as intellectual infrastructure, this Initiative looks to stimulate partnering arrangements necessary to create next-generation operational systems in such areas as education, engineering and design, earth and space sciences, biosciences, geography, economics, and the arts and humanities. It will address the digital libraries life cycle from information creation, access and use, to archiving and preservation.

DLI - Carnegie Mellon:

- [Home page - Infromedia](#)
- [Infromedia-II \(for DLI-2\)](#)
- [IEEE Computer article](#)
- [NetBill](#)

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Projects\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

DLI - Stanford:

- [Home Page](#)
- [IEEE Computer article](#)
- [testbed development](#)
- [info finding](#)
- [user interfaces](#)
- [DLITE \(task env\)](#)
- [SDLIP](#) (Simple DL Interop. Protocol) - also see [D-Lib Magazine article](#)
- [mediation infrastructure](#)

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Projects\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

DLI - Berkeley:

- [Home Page](#)
 - [IEEE Computer article](#)
 - [Tours](#)
 - [Collections](#)
 - [Source Code](#)
 - [Document-specific image decoders](#)
 - [GISviewer](#) (needs latest browser)
 - [Photos](#) and demos
 - [Context-based image queries](#)
 - [Blobworld](#)
 - [Image classification](#)
 - [UCB database management](#)
and the Open Source Berkeley DB: [Sleepycat Software](#)
 - [California Aerial Photos](#)
 - [United States Department of Agriculture PLANTS Photo Gallery](#)
-

Pedagogy:

We recommend that the reader study these materials as part of work to answer the following questions:

- MVD
 - How well does [MVD 0.9](#) work for you? Could you get the links on that page to work (use 2 windows of browser, one for the instructions, and one for testing)? What do you like most about it?
 - Did you use it on video or a PC or Mac with Netscape 4?
 - Did you work out Lens overlaying, such as OCR and then Magnify?
 - For the TableSort example, could you under Anno view the note?
 - Could you get the special behaviors to work: Biblio, where you Select a type of format, use the mouse to select an entry, use Edit and Copy to get a version in that format, and then paste elsewhere?
 - Could you get Doublespace in the View menu to work?
- Cheshire
 - Can you find interesting environmental documents using Cheshire II?
- TileBars

- What happens with TileBar search of "document" and "retrieval"?
- What happens with TileBar search of "fault" and "dam"?
- When is TileBar searching useful on a single document?
- Collections
 - What is the name of the DBMS used?
 - What is a database "schema"? How does it relate to "metadata"?
 - How many documents and how many images are in their collection?
 - How good is the OCRing? What research is underway to improve OCRing beyond that of ScanWorX and how well does it work? What is the main idea behind it?
 - How can you find the dams for a county?
 - How does the database table information for Almond dam relate to the page about it? To the OCR output about that page?
 - What is a VLURL? How do you construct it? Can you build one and show results for getting pictures of California wildflowers that have the string "rose" in their common names?
 - Display a distribution map for your favorite flower in California.
 - Can you tell the direction of flight from the aerial photos?
 - How do layers help with managing GIS information with the [GIS viewer](#)? Can you zoom in and out and pan around?

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Projects\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

DLI - Santa Barbara:

- [Home Page](#)
 - [IEEE Computer article](#)
 - [World Spatial Data](#)
 - [1994-1998 DLI-1 Project](#)
 - [H. Chen's work](#) (with "cool DL, Web, agent, visualization, and multilingual IR demos")
-

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Projects\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

DLI - Illinois:

- [Home Page](#)
- [IEEE Computer article](#)
- [Glossary](#)
- [SGML/XML Home Page](#), [SD Unit Notes in CS5604](#), [SoftQuad Products](#)
- Collections: [Publishers](#), [Software Companies](#)
- [Interspace](#)
- [Social Science Team Home Page](#)
- [DeLiver](#)
 - Before using DeLiver you should get one of the following 2 files and install it on your Windows 95/NT system. Be sure to have any version of Netscape closed after the download, when you do the install. These files are local to VT to save you the time of downloading as per the U. Ill. instructions. The Panorama versions each take about 1.9M for the install package but less than 1M for the C: drive installed version Netscape.
 - Explore the DeLiver pages, and try to answer the following questions.
 - What does the Help tell you about the system?
 - What is the coverage?
 - What are unusual services not provided by similar systems?
 - What is Panorama and what does it do to enhance WWW capabilities?
 - Can you use browsing to find the IEEE-CS articles (i.e., v. 29 n. 5) we looked at for this course?
 - Can you use searching to find the IEEE-CS articles we looked at for this course?
 - How does the presentation using WWW and Panorama differ from that you are familiar with (HTML, PDF)? What benefits are there from having Panorama?
 - What other interesting articles about digital libraries did you find?
 - Is the field specific searching of help?

Is the interface for DeLiver easy to understand? How could it be improved?

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Projects\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

University of Michigan Digital Library Activities

DLI General Information

- [Home Page](#)
- [IEEE Computer article](#)
- [Introduction](#)
- [Current Status](#)
- [Technologies](#)
- [Agents, Ontologies](#)

Campus Strategy

- Partnership of
 - [University Library](#)
 - [Information Technology Division](#)
 - [School of Information](#)
- combine: R&D; technology infrastructure; content access & user services; outreach
- shift to 21st century library model
 - user-centric, collaborative teams, global reach
 - distributed collections, heterogeneous access protocols, just-in-time information delivery
 - mixed funding models, value = access + services
- [Gateway Registry](#)
- [Electronic Reserve Shelf](#)
- [Knowledge Navigation Center](#): develop and support teaching and learning projects
- Questions:
 - How does the infrastructure at U. Michigan compare to that at your university?
 - How does this strategy relate to previous services of libraries?

Projects

- [JSTOR](#): Journal Storage: over 1.2M pages
- [Making of America](#): with Cornell - 5K volumes, [D-Lib article](#): scanning, OCR, SGML encoding, tif2gif, interface
- [DLPS Image Services](#): see also V. 5 N. 8 Oct. 1996 [Information Technology Digest](#)
- [Humanities Text Initiative](#)
- [Papryology](#)
- [Middle English Compendium Demo](#)
- [American Verse](#)

- [Collaboratories](#)
- Questions:
 - Which of these projects do you find most interesting? Why?
 - Which of these projects should your university become involved in?

Technical Approaches

- [see especially 1996 Ann Arbor Conf. on Electronic Records R & D](#)
 - Problem scenarios (see bullet list under **The Importance of Digital Preservation**)
 - Research questions (see **The 10 Research Questions**)
 - Research results: possible, requires changes and new types of efforts (see bullet list under **Research Projects and Results**)
 - [International Council on Archives](#): see **Guide for Managing Electronic Records from an Archival Perspective**, survey, literature review
- [Advanced Interfaces](#)
- [Ontology - Concept Descriptions](#) and [May 1997 slides](#)
- [Learning Agents](#)
- [SGML creation and delivery](#)
 - enormous collection: 2M pages
 - [flowchart](#)
 - [SGML resources](#)
- [Leveraging rich document formats](#)
 - patterns of use
 - ease of changing delivery: new standards (HTML), new rendering/packaging
 - collection management
 - Panorama, XML support by W3C
- Questions:
 - Will the agent and ontology approach work? Soon? For production DLs?
 - What is the support needed for establishing a digital library following the UMDL approach? Training?
 - What interfaces for DLs will be usable?

Last updated 6/23/2001.

The Library of Congress

[SEARCH THE CATALOG](#) | [SEARCH OUR WEB SITE](#) | [ABOUT OUR SITE](#)
[NATIONAL BOOK FESTIVAL](#) | [GIVING](#) | [JOBS](#) | [TODAY IN HISTORY](#)



Above: the interior dome of the Main Reading Room at the Library of Congress
For an [online tour of the Jefferson Building](#), click on the dome.

101 INDEPENDENCE AVENUE, S.E.
WASHINGTON, D.C. 20540
(202) 707-5000

COMMENTS: lcweb@loc.gov
[Please Read Our Legal Notices](#)

[COLLECTIONS & SERVICES](#) | [AMERICAN MEMORY](#) | [COPYRIGHT OFFICE](#) | [THE LIBRARY TODAY](#)
[THOMAS](#) | [AMERICA'S LIBRARY](#) | [EXHIBITIONS](#) | [HELP & FAQs](#)

Digital Library Network (DLnet)

[in Japanese](#)

Welcome to Digital Library Network Homepage. Here, we present programs and records of the series of workshops on Digital Libraries at University of Library and Information Science, Tsukuba Science City, Japan.

Digital Library Network (DLnet) was proposed at the First Workshop on Digital Libraries on August 31, 1994, to provide free-access forum on Digital Libraries. We welcome your comments and questions to DLnet.

English Pages

- [DL Workshop Program](#)
- [Titles and Abstracts from Digital Libraries \(ISSN 1340-7287\)](#)
- [ISDL'99: International Symposium on Digital Libraries 1999](#)
- [ISDL'97: International Symposium on Research, Development & Practice in Digital Libraries 1997](#)
- [International Symposium on Digital Libraries 1995](#)
(August 22 - 25, 1995 at ULIS)

Japanese Pages

- [DLnet HomePage](#)
- [International Symposium on Digital Libraries 1995](#)
(August 22 - 25, 1995 at ULIS)
- [DL Workshop](#)
- [Digital Libraries \(ISSN 1340-7287\)](#)

Others

- [DLnet Gopher Server](#)
- [DLnet Anonymous-FTP Server](#)
- [Roadmap Lessons \(In Japanese\)](#)
- [JAPAN/MARC Search experiment at ULIS](#)
- [Multilingual-HTML Browser Project](#)

- [Japanese Oldtales: A Multilingual E-text Collection](#)
- [Dublin Core Reference Description in Japanese](#)
- [University of Library and Information Science homepage](#)
- [University Library, ULIS](#)

[DL HomePage](#)

sugimoto@ulis.ac.jp

People:

[Rob Akscyn](#) of [Knowledge Systems Incorporated](#) with its [PetaPlex Project](#)

[Caroline Arms](#) of [Library of Congress](#)

[William Arms](#), at [Cornell CS](#), formerly at [CNRI](#)

[Dan Atkins](#) [University of Michigan, DLI-1 Digital Library Project](#) Director.

[Howard Besser](#) of [School of Information Management and Systems at Berkeley](#)

[Bill Birmingham](#): [University of Michigan, DLI-1 Digital Library Project](#) Researcher.

[Chris Borgman](#) of [Information Studies at UCLA](#)

[Hsinchun Chen](#) Head of the [AI Lab of U. Arizona](#) and director of new [DLI-2 project](#)

[Stephan Fischer](#) - working on multimedia and metadata

[Edward A. Fox](#) Director of the [Digital Libraries Research Group](#) at Virginia Tech.

[Beverlee French](#), University Librarian and Executive Director, Interim [California Digital Library](#).

[Rick Furuta](#) of [CS at Texas A&M Univ.](#)

[Hector Garcia-Molina](#) in the [Stanford DB Group](#)

[Henry Gladney](#) retired from [IBM Almaden Research Laboratory](#)

[Dan Greenstein](#), Director of the [Digital Library Federation](#)

[Stephen Griffin](#), Program Director of the [Digital Libraries Initiative](#), in [NSF' IIS program](#)

[Robert Kahn](#) of [CNRI](#)

[Judith Klavans](#) of [Digital Libraries Projects at Columbia](#)

[Carl Lagoze](#) of [DL Research Group](#) of [CS at Cornell Univ.](#)

[John Leggett](#) of [CS at Texas A&M Univ.](#)

[Michael Lesk](#) Director of [NSF' IIS program](#) that runs the [Digital Libraries Initiative](#)

- [Images: Quantity is not always Quality - U. KY talk](#)
- [digital libraries](#)
- [library preservation](#)
- [information retrieval](#)
- [networking, etc.](#)
- [Projections for Making Money on the Web](#)

[Richard Lucier](#), College Librarian, Dartmouth. See his D-Lib [article on CDL](#)

[Clifford Lynch](#) Director of [CNI](#)

[Gary Marchionini](#)

- Previously at [U. Md.](#)
with its [DL Home Page](#)
- Now at [U. NC Chapel Hill School of Information and Library Science](#)
- [Encyclopedia article draft](#)
- [CACM April 1995 article](#)

[Michael Mauldin](#) ([home page](#), [Lycos](#), [CMU School of Computer Science](#))

[Bruce Schatz](#) Principal Investigator of [University of Illinois at Urbana-Champaign, DLI Project](#)

[Robin Sewell](#), co-PI with Hsinchun Chen (see above) on U. of Arizona DLI-2 project

[Marvin Sirbu](#) of [CMU Engineering and Public Policy](#)

- [publications available online](#)

[Terry Smith](#) from [Geography](#), Director of [Alexandria project](#) at [U. CA Santa Barbara](#)

[Howard D. Wactlar](#), Principal Investigator of the [Informedia Digital Video Library](#), [CMU School of CS](#)

Donald Waters of [The Andrew W. Mellon Foundation](#)

[Stuart Weibel](#) of [OCLC Office of Research](#)

[Robert Wilensky](#) Principal Investigator of [Berkeley DLI Project](#)

Note: for an extensive list of people involved in digital libraries, see the [Author Index](#) of D-Lib Magazine.

Note: for a list of some of the key people in the digital libraries field, see the report on this from a Delphi Study at http://www.coe.missouri.edu/~is334/projects/Delphi_DL/StatementAnalysis.htm: "By consensus, those identified in the rounds of the Delphi as the top ten (10) include: William Arms, Christine Borgman, Hector Garcia-Molina, Edward A. Fox, Carl Lagoze, Michael Lesk, Richard Lucier, Clifford Lynch, Gary Marchionini, Bruce Schatz, and Terence R. Smith."

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Henry Gladney:

- Access Control Articles in D-Lib Magazine:
Gladney et al., Safeguarding Digital Library Contents and Users:
 - [Assuring Convenient Security and Data Quality](#),
 - [Document Access Control](#)
 - [Digital Images of Treasured Antiquities](#)
 - [A Note on Universal Unique Identifiers](#)
 - [Storing, Sending, Showing, and Honoring Usage Terms and Conditions](#)
- [Gladney et al. report on DL requirements and architecture \(PostScript\)](#)

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[People\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Michael Mauldin

- [Michael Loren Mauldin](#), alias "Fuzzy," has many hats.
- He is Chief Scientist at [Lycos](#), Inc., the Internet Search Engine he created.
- He is also Managing Director of Virtual Personalities, Inc., a company dedicated to creating Self-Animated Computer Generated Human Characters.
- Finally, he is Adjunct Research Computer Scientist at Language Technology Institute of the School of Computer Science at Carnegie Mellon University.

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[People\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Countries & Regions:

(Chapter 11, page 245, "Books, Bytes and Bucks", Michael Lesk)

- **United States of America:** In the US, NSF, NASA and ARPA have funded six important Digital Library efforts, called the DLI (Digital Libraries Initiative). These programs each involve a large consortium of cooperating institutions but the six main ones are : University of California at Berkeley, University of Santa Barbara, University of Michigan, Carnegie Mellon University, Stanford University, and the University of Illinois.
 - University of California at Berkeley: Image content queries along with Xerox PARC, database extraction from documents, multivalent documents, NLP. Headed by Robert Wilensky.
 - University of Michigan: Scalability and Education. They are also investigating the use of agent architectures for Digital Libraries and trying to merge DLI with their other digital library efforts such as JSTOR and TULIP. Headed by Dan Atkins.
 - University of Illinois: Concentrating of using scientific journals as their base collection with diversity in both documents as well as publishers, making the transition process from SGML to HTML smoother, defining semantic spaces. Headed by Bruce Schatz.
 - Stanford University: concentration is on the infrastructure development such as bas networking and databases to support digital libraries. Also concerned with interoperability between different digital library projects. Headed by Hector Garcia-Molina.
 - University of California at Santa Barbara: spatial indexing and retrieval , image processing. Headed by Terry Smith.
 - Carnegie Mellon University: digital video, image analysis, speech recognition, face recognition, natural language understanding. Headed by Michael Mauldin and Marvin Sirbu.

Other than DLI, many research projects are underway at some other universities such as Virginia Tech and Texas A&M. In the near future, extensive funds are expected to be allocated for Digital Libraries.

The Library of Congress, under James Billington is digitizing 5 million of its items in a massive \$60 million effort. Other universities involved in related projects are Georgia Tech, Cornell, MIT, University of Tennessee, Washington and California and Virginia Tech (known for the Envision system of Ed Fox). Other limited efforts include University of Virginia, University of Georgia and Columbia University.

- **United Kingdom:** Though efforts are still limited to penny-pockets, 20 million pounds have been

set aside from digital library projects. The program originally called FIGIT, now known as E-LIB funded 35 projects. Work includes cataloging of archives, digitization of documents and data sharing. Some of the more notable efforts are : Digitizing the Burney collection of pre-1800 newspapers and scanning of Batley News, the Canterbury Tales project that involves scanning all pre-1500 manuscripts and some of the similar projects. However, the most notable is the Electronic Beowulf project which is a US/UK collaboration between Kevin Kiernan (University of Kentucky), Paul Szarmach (Western Michigan University) and the British Library.

- **France:** Work includes some scanning of old manuscripts with the most notable being the Tresor de la Langue Francaise project at the University of Nancy. The French, along with the Japanese are also leaders in the Group 7 project which is a museum project. Other efforts are INIST and FOUDRE (1989 to 1992) followed by EDIL and ELITE.
- **The EU:** The European Union funds a large number of international efforts in digital libraries. (Please see page 255 of Michal Lesk's book for details)
- **Japan:** Japan is involved in some digitization and cataloguing efforts and has a \$50M project on. They are also working on modern document delivery and OCR.
- **Australia:** Australia has recently made a modest effort to enter into digital library research. They are planning some digitization projects with a \$10M (Australian) digitization project on the anvil. They are also interested in digitizing Aborigine scriptures and paintings.
- **Elsewhere:** Many other countries are involved in digital library research on much smaller scales. Notable amongst them are Canada, Singapore, Korea and China.

NOTE 1: For detailed information on any of the above please refer to Dr. Lesk's book (recommended as supplement text for this course).

NOTE 2: See also the table pointing to various national digital libraries from April 1998 CACM [online pages](#)

See also [DLI2 International Digital Libraries Projects](#)

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Centers, sites and organisations:

Some major Digital Library centers and research programs, separately described:

- [Carnegie Mellon University](#)
 - [CNRI](#)
 - [Library of Congress](#)
 - [Stanford University](#)
 - [University of California at Berkeley](#)
 - [University of California at Santa Barbara](#)
 - [University of Illinois](#)
 - [University of Michigan](#)
 - [Texas A&M](#)
 - [Virginia Tech](#)
-

Selected other sites:

[ACM DL](#) : Tap into the ACM Digital Library, a vast resource of bibliographic information, citations, and full-text articles.

IEEE-CS [Digital Library](#)

IBM

- [IBM DL Home page](#)
- [IBM Renaissance Consortium Panel](#) and [workshop](#)
- [images - QBIC](#)

[National Library of Medicine](#)

[Digital Library Research Program](#) at

[Lister Hill National Center for Biomedical Communications,](#)

[National Institutes of Health](#)

[OCLC](#) (OCLC is a nonprofit, membership, library computer service and research organization dedicated to the public purposes of furthering access to the world's information and reducing

information costs).

- Research <http://www.oclc.org/research/>;

SiteSearch <http://www.oclc.org/oclc/menu/site.htm>

Xerox

- [DL Interfaces Home Page](#)
- [Scientific American article](#)
- [Scatter/Gather examples](#)
- Questions:
 - Compare
 - What are the various interfaces built? How do they compare? What is the best use of each?
 - Scatter/gather
 - Explain clustering, relate it to scatter/gather.
 - What are special problems with large category systems and how can they be solved?

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

CNRI:

- home page (site map) http://www.cnri.reston.va.us/site_map.html
- Architecture
 - Kahn-Wilensky Framework for Distributed Digital Object Services_
<http://WWW.CNRI.Reston.VA.US/home/cstr/arch/k-w.html>
 - key architectural issues
(1996)<http://WWW.CNRI.Reston.VA.US/home/cstr/arch/slides.html>
 - architecture for information in digital libraries
<http://www.dlib.org/dlib/february97/cnri/02arms1.html>
 - Digital Object Architecture Project <http://www.cnri.reston.va.us/doa.html>
- Handle System (<http://www.handle.net/>) and Digital Object Identifier System
(<http://www.doi.org/>)
- CS-TR Computer Science Technical Reports <http://www.cnri.reston.va.us/cstr.html>

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Centers\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Library of Congress:

- American Memory <http://lcweb2.loc.gov/>
 - Call/Awards about American Memory <http://lcweb2.loc.gov/ammem/award/>
 - Sponsors and Contributors to the National Digital Library Program
<http://lcweb2.loc.gov/ammem/sponsors.html>
-

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Centers\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Search, retrieval, resource discovery:

Searching - LoC

- [LoC Home Page](#)
- [Z39.50 maintenance agency; part 1](#)
- [The WWW Virtual Library arranged by LoC standards](#)
- [Understanding and Comparing Web Search Tools](#)
- [Matrix of WWW Indices: A comparison of Internet indexing tools](#)

Federated search

- [UIUC Federation Across Heter. DBs](#)
- [STARTS](#)
- [INFOSEEK patent](#)
- [TSIMMIS](#)
- [Virginia Tech Federated Search Demonstration for NDLTD \(theses, dissertations\)](#)
- [Emerge \(NCSA component architecture\)](#)

CyberStacks (WWW, Classification, Catalogs, Reviews/Clearinghouses)

- [Home Page](#)
- [Net Projects](#)
- [Alphabetical topics vs. LC ranges](#)
- [Call for contributions](#)
- Question: Which efforts are far along? What demonstrations can you find that are the most informative / explanatory? How well does the Library of Congress classification system fit for WWW resources?
- Related work: [OCLC's Scorpion Project](#); [DDC](#); [Mantis](#); [CORC](#)

Columbia

- [D-Lib Article on Images/Video](#)
- [WebSeek Home Page](#)

Database Groups

Filtering

- [Defn](#) from U. Md. [Information Filtering Project](#)
- [Paracel automated genomic sequence and text analysis systems](#)
- What is *information filtering*? How does it differ from information retrieval?

Cross-Language Information Retrieval Resources

- [Eurospider](#) and [ISN LASE Search demo](#)
- [Readware](#)
- [Mundial](#) - English and Spanish Demo
- Questions:
 - What languages are covered?
 - How well are phrases handled?

Stanford DL info finding projects

[Berkeley documents and queries](#) (please study carefully, answering questions)

UCSB spatial indexing and retrieval

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

About ETD Federated Search

Federated Searcher allows users to perform parallel queries across several dozen search sites provided by participants of the Electronic Theses and Dissertations Project. Each site is described using a specially designed XML markup language called *SearchDB*. A Java-based federated search server maps queries to each site you select by using the XML description as a submission template. It submits each query and collects results as each site replies. Currently, each result set is presented as a separate document, although future plans include result set merging.

[Show me all ETD sites](#)

or

Find cataloged sites about

Search or Browse the Catalog

One of the many ways in which this service differs from other "metasearch" services is in its use of metadata for search sites. The first step to performing a federated search is to select the sites you would like to search. Each site has a local description that includes information about its particular specialty. So if you want to perform searches to help you decide where you should take your next vacation, you can search the catalog for **Computer Science** and then perform federated searches for things like **object oriented programming** or **Java** or **research results** against those sites most likely to index documents about computer science.

[All ETD sites currently included in the Federated Search](#)

Questions? Comments? etd@ndltd.org

[NDLTD](#)



emerge@ncsa.uiuc.edu

About EmERGE

EmERGE is an NCSA effort to develop middleware components of a new distributed search infrastructure which addresses the scale and heterogeneity of scientific data. Our components enable search services to interoperate across scientific domains by providing user-configurable tools for mapping between metadata schemas, performing search queries against multiple data sources, and performing query pre- and post-processing. Access to our search services is through platform-neutral standard and emerging-standard tools such as [Z39.50](#), [Open Archives](#), [XML](#), and [Java](#).

Here's a [slide show](#) with an overview of our research area and component architecture. And [here's one](#) which gives an overview of interoperability issues in distributed scientific information retrieval.

Collaborations

EmERGE is part of [NCSA's Data Mining and Visualization Division](#). Our components have been developed in collaboration with the [National Cancer Institute](#), the UIUC Digital Library Initiative and [CANIS](#), [NASA Project 30](#). We've also participated in panel discussions and advisory meetings with the [Committee for Institutional Cooperation](#) and the [UIUC Library Gateway](#) project.

EmERGE is currently helping to build the National Biological Digital Library in collaboration with the [University of Missouri](#), the [Missouri Botanical Garden](#), and the [Graduate School of Library and Information](#)

[Science](#) at UIUC. The NBDL is an NSF-supported effort to engage the education community in the development and use of federated plant science data collections.

Database Groups:

- [Garlic - IBM Almaden](#)
- [Penn.](#)
- [Stanford](#)
- [U. Md.](#)
- [UCB database management](#)
and the Open Source Berkeley DB: [Sleepycat Software](#)
- [Oracle](#)

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Oracle and Digital Libraries

[Oracle 9i Database](#)

In 9i the interMedia Text component is Oracle Text.

[Oracle interMedia](#)

Overview by Omar Alonso, Omar.Alonso@oracle.com, 650-607-3410:

Briefly, Oracle interMedia extends Oracle8i to manage rich content, including text, documents, image, audio, video, and geographic location, together with traditional business information.

interMedia is a standard feature of Oracle8i. It is included with every Oracle8i license, and provides content services to JDeveloper, Oracle Developer, iFS, WebDB, Oracle applications and Oracle partners.

Using interMedia services it is possible to . . .

- Use standard SQL to index and search text and documents stored in Oracle8i, in files and on the Web, including metadata associated with rich content, to provide retrieval capabilities fundamental to Web and other applications.
- Parse, index, and load rich content in Oracle8i and deploy to the Web with support for popular web page composition tools, web server technologies, and Web media formats (e.g., GIF, JPG, AU, WAV, MP3, QT, Real) delivered in either batch or streaming modes.
- Develop dynamic Web applications with rich media content using interMedia APIs for Java, C++ and PL/SQL.
- Tune Oracle8i based content repositories to achieve scalability and reliability superior to o.s. file based systems.

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 2000-2001, Edward A. Fox

Information Filtering Defined

A universally accepted definition of information filtering is, unfortunately, still lacking. So here is my personal definition, which I have used to build the Information Filtering Resources [web page](#). Generally, the goal of an information filtering system is to sort through large volumes of dynamically generated information and present to the user those which are likely to satisfy his or her information requirement.

In order to sharpen this definition, a distinction should be drawn between information collection and information filtering. In some domains (e.g. USENET News) the collection effort is minimal because the information comes to you. In other domains (e.g. the World Wide Web) the collection effort can be considerable because no mechanism exists to draw new information to the attention of a filtering system. The point to be made here, though, is that information collection is an interesting area in its own right, but I do not propose to include it in my definition of information filtering. In my view, the information filtering problem begins only after you have gained access to the new information.

Information filtering has been applied to a several domains using a variety of technical approaches. The original methods were manual alerting services that brought new information to the attention of users of research and special libraries. At the time this was referred to as Selective Dissemination of Information (SDI), a name which fell from favor about the time the Strategic Defense Initiative (SDI) was introduced in the United States :-). A few modern systems have adopted this remarkably descriptive name for the filtering process, however, and the interest in information filtering that has resulted from the present research thrusts in digital libraries arises at least in part from this tradition.

With the growth of the internet and other networked information, research in automatic filtering of networked information has exploded in recent years. Because of their low cost, large volume, and ease of recognizing new information, the most popular domains for research systems have been USENET News and electronic mail. The recent explosive growth of the World Wide Web has made this an interesting domain which has attracted some good research, although the information collection problem appears to make this a more difficult domain in which to conduct basic research on information filtering techniques. Another domain which has attracted considerable research interest is the annual Text REtrieval Conference (TREC) in which a standard text collection is used and a carefully controlled evaluation methodology is enforced. In TREC the information filtering task is referred to as "routing," adding somewhat to the confusion of terminology in this field. In fact, TREC recently adopted a special interest "filtering" track which adopts a different evaluation methodology but which conforms to the definition of filtering presented above. Commercial systems which filter newswire articles and other specialized information sources are becoming available as well. Filtering techniques will likely be applied to other domains such as images, sound and video in the future.

The distinction between information filtering and the more established field of information retrieval has proven to be the source of some confusion as well. Information retrieval broadly deals with the selection of information, and many of the features of information retrieval system design (e.g. representation,

Cross-Language Information Retrieval Resources

This page is designed as a resource for people conducting research in [cross-language information retrieval](#). It is intended to collect references to all information on information retrieval systems which can accept queries in one language and return documents in another. It is maintained by the [Digital Library Research Group](#) of the [College of Information Studies](#) at the University of Maryland. If you are aware of resources that are within the scope of this page but do not appear here, please [send mail to Doug Oard](#).

[December 1997 D-lib Magazine Article](#)

An introduction to cross-language information retrieval. A web page that was prepared for a [public lecture](#) here at Maryland provides another perspective on the topic.

[Conferences](#)

An excellent source of information. This page includes links to the full proceedings of most major cross-language information retrieval workshops as well as to a fairly complete list of upcoming conferences and workshops that include some treatment of cross-language information retrieval.

[Cross-Language Information Retrieval Papers and Project Descriptions](#)

Another excellent place to look for information. Here you will find descriptions of experimental work on cross-language text retrieval that may not have been presented at one of the major workshops

[Working Systems](#)

Here you will find links to experimental and commercial cross-language information retrieval systems that you can either obtain or use over the net. Some carry a fairly hefty price tag, others are free.

[Related Resource Pages](#)

Web pages which collect links to resources that may be of interest to cross-language information retrieval researchers. None of these pages are devoted solely to cross-language information retrieval.

Last modified: Fri Nov 24 21:40:29 2000

[Doug Oard](#) oard@glue.umd.edu



CS5604 - Information Storage and Retrieval

Fall 1996 - Table of Contents

- [Assignments](#)
- [Calendar](#)
- [Computers and Tools](#)
- [Course Format](#)
- [Course Notes / Overheads](#)
- [Department and Class Policies](#)
- [FAQ - Frequently Asked Questions](#)
- [Glossary \(in process\)](#)
- [Koofers \(old quizzes\)](#)
- [News / Announcements](#) (updated 961213@5am)
- [Photos of Class](#)
- **Projects:** [Initial Suggestions](#), [Groups](#), [Completed Projects](#)
- [Quizes](#)
- [Readings and References](#)
- [Review](#)
- [Searching ei.cs.vt.edu Online with Harvest](#)
- [Status](#)
- [Syllabus](#)
- [Trips](#)
- **WWW Link Sets:** [Instructor's - CS4624: Multimedia, Hypertext and Information Access - WWW Virtual Library \(URLs organized by subject\)](#)

Extended Boolean Queries and Retrieval

Problems with Boolean

- A AND B AND C AND D AND E --- if miss one
 - get nothing, instead of those with 4, or later those with 3, etc.
 - don't have an easy way to reformulate for all the combinations
- A OR B OR C OR D OR E --- if have several
 - counts just like if only have one
 - don't have an easy way to show that prefer more than one occurrence
- A NOT B --- eliminates even casual use of term B
- No ranking
 - so users must fuss with retrieved set size, structural reformulation
 - so users must scan entire retrieved set
- No weights on query terms
 - so users cannot give more importance to some terms --- retrieval:2 AND system:1
 - so users cannot give more importance to some clauses --- retrieval:1 AND (MMM OR Paice):2
- No weights on document terms
 - so indexers are forced to make strict binary decisions --- forcing fewer index terms and lower recall
 - so no use can be made of importance of a term in a document --- if occurs frequently
 - so no use can be made of importance of a term in the collection --- if occurs rarely

Fuzzy Set Theory

- Zadeh since 1965
- Studied here in EE
- Recently adopted in Japan: numerous patents: fuzzy controls, shower heads
- Start with notion of sets for : tall, small, large, bright, kind, ...
- Use range [0,1] instead of choice (0,1)
- Redefine AND as MIN
- Redefine OR as MAX
- Evaluate NOT B as $1 - \text{value}(B)$

Applying Fuzziness to IR

- If want Boolean laws to apply, must use MIN/MAX definitions.
- Can apply to automatic document indexing with term weight =

- 0, if term not present in document;
- $0.5 + 0.5 \cdot \text{TF}/\text{MAX-TF}$, if term is present in document;
- some reduced value, if a related term is present instead.
- Have no simple way to consider query term weights.
- Still have problems:
 - A AND B AND C AND D AND E --- only term with lowest value counts
 - A OR B OR C OR D OR E --- only term with highest value counts
 - Computational and space costs are higher than for Boolean.

MMM Model

- Idea: generalize MIN and MAX by redefining AND and OR as linear combination of them:
 - AND: $\text{Cand} * \text{MIN} + (1 - \text{Cand}) * \text{MAX}$
 - OR: $\text{Cor} * \text{MAX} + (1 - \text{Cor}) * \text{MIN}$
 - Good values seem to be Cand in $[0.5, 0.8]$ and Cor in $[0.2, 1]$.
- Problem: still only considers 2 terms (one with lowest weight, and one with highest weight) as opposed to all terms in query.

Paice Model

- Idea: consider all of the terms in the query.
- Idea: use a normalized geometric series, down-weighting the contribution of terms not close to the fuzzy set value (i.e., MIN for AND, MAX for OR).
- Formula has single coefficient, r , which works well as 1 for AND queries or 0.7 for OR queries.
- Sort document terms based on their weight:
 - in ascending order for AND queries;
 - in descending order for OR queries.
- Evaluate similarity for that document by dividing
 - SUM (for all query terms in $[1, n]$) of $r^{i-1} * d_i$
 - by the normalization value
 - SUM (for all query terms in $[1, n]$) of r^{i-1}

P-Norm Model

- Idea: consider all of the terms in the query.
- Idea: parameterize the strictness of each AND or OR operator with a p-value.
- Idea: have a general model, p-norm, that has as special cases the standard Boolean model (with fuzzy set interpretation --- when p is infinity) and the vector-space model (with inner-product similarity --- when p is one).
- Thus we get a spectrum of models with decreasing strictness, i.e., strict AND ... soft AND ...

vector ... soft OR ... strict OR:

- p-norm AND with $p=\infty$ behaves like strict Boolean AND (i.e., MIN)
- p-norm AND with p at moderate values softens the strictness of the AND
- p-norm AND with $p=1$ behaves like p-norm OR with $p=1$ and behaves like vector space model
- p-norm OR with p at moderate values softens the strictness of the OR
- p-norm OR with $p=\infty$ behaves like strict Boolean OR (i.e., MAX)
- Idea: use L-p family of norms to compute similarity by measuring:
 - distance from 0 point (i.e., none of query terms present) for OR;
 - $1 - \text{distance from 1 point}$ (i.e., all of query terms present) for AND.
- Idea: visualize all this with equi-similarity contours at fixed p -values.

Comparison of Extended Boolean Models

- All seem to work best when AND is interpreted fairly strictly, and OR is interpreted less strictly.
- All are computationally more expensive than Boolean, but at the same time are more effective (i.e., precision at given recall level).
- Computational costs seem to be (in the general case): $\text{MMM} < \text{Paice} < \text{P-norm}$
- Effectiveness (i.e., precision at given recall level) seems to be: $\text{MMM} < \text{Paice} < \text{P-norm}$

Implementation Issues

- Need to parse and represent queries (with clause and term weights).
- One way to evaluate "similarity" for a document is to "walk" the query tree in a depth-first traversal --- can be done by recursive evaluation.
- Need to store document weights (unless assume binary weights, or compute at retrieval time based on postings or other statistics).
- Can first do standard Boolean processing and then use an extended Boolean model to prepare a ranking for those retrieved.
- However, to improve recall, should really retrieve all documents that have any of the query terms, and then compute "similarity" for those, to get a full ranking.

Web Thesaurus Compendium

Quick access:

[- thesauri alphabetical](#)

[- thesauri indexed by subject](#)

[- thesaurus-related literature and resources](#)



[- thesaurus-related research groups](#)



[- thesaurus-building tools and software](#)



See NewHoo's listings of reference works at [NewHoo Reference/Thesauri](#)



Have a look at [Links2Go Thesauri Key Resources](#)

Introduction

The thesauri and classification schemes in this collection are all available on the web with various search and browse facilities, and various degrees of hypertext linking. "Search" means you can enter a term and search for it directly; "browse" means you can look through alphabetical or hierarchical lists. Several of the systems allow a search in data collections to be launched directly after finding the desired search terms in the thesaurus.

The term "thesaurus" is used loosely here to refer to any structured collection of interrelated terms; often, but not necessarily, in a certain domain.



In addition to the thesauri/classifications themselves, I also plan to include links to [thesaurus-building tools and software](#), to a selection of [academic literature on thesauri](#), and to some [research groups](#) doing related work. Check back in a few weeks for first drafts.

This is an ongoing list, by no means complete. If you have a thesaurus or other links for inclusion, please contact me ([Barbara Lutes](#)) at lutes@darmstadt.gmd.de. For more information on work being done in the areas of multimedia information retrieval and digital libraries at our

OAI

- [Open Archives Initiative Home Page](#)
 - [OAI Pages at VT](#)
 - The Open Archives Initiative: Building a low-barrier interoperability framework. [9 page PDF version of article presented at JCDL 2001](#)
 - [Other Documents](#)
 - [Repository Explorer](#)
 - [Arc: Cross Archive Searching](#)
 - [Kepler](#)
 - [Survey of E-Prints archives](#)
-

[\[Main\]](#) [\[Contents\]](#) [\[Metadata\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 2001, Edward A. Fox

Multimedia, Representations:

The Basics:

- [text file formats](#)
- [graphic file formats](#)
- [hypermedia & multimedia](#)

ACM DL'97 Tutorial: [Multimedia Information and Systems](#)

[ACM SIG on Information Retrieval](#) ; [ACM SIG on Multimedia](#) ; [IEEE-CS TC on Multimedia Computing](#) ; [Computing Curricula 2001](#)

Digital Video

- [KRDL: Seamless Integration of Video Contents for Web-based Presentations over Different Devices](#)
- [KRDL: Video to SlideShow System \(ViSS\)](#)
- [CNN uses Quicktime for WWW daily news clips](#)
- [Digital Video Resources on Internet](#)

MHIA Courseware and Curricula

- [Curriculum Resources in Interactive Multimedia \(CRIM\) Home Page](#)
- [MHIA Home Page](#)
- [SIGIR 96 Workshop](#)
- [Drexel 96 Workshop](#)
- [IR Courses](#)
- [Multimedia Courses](#) (Dublin, Ireland)
- [MM 1996 Workshop](#)
- [Lisbon 1997 Workshop](#)
- Questions:
 - What is the need for education related to information? What jobs?
 - What subjects should be covered in such education programs?
 - How should those subjects be ordered into each specific program?

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Architectures:

Core topics include:

- [D-Lib article on architecture](#)
- [Other CNRI activities](#)
- **Naming**
 - [PURL](#)
 - [Handles](#)
- [Networks](#): online notes of Dr. Lesk

Other topics of general interest, that are being studied by the [D-Lib Metrics Group](#) include:

- **Distributed processing (client/server)**
- **Interoperability** (see [IITA workshop on Interoperability](#) and some of work at [Stanford](#), [EU](#), as well as the [Open Archives Initiative](#))
- **Performance**
- [Agent-Based Architecture](#), W. Birmingham, D-Lib Magazine, July 1995

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Interfaces:

[Stanford DL user interface projects](#)

Xerox Interfaces for Information Access

- [Home Page](#)
- [Scientific American article](#)
- [Cat-a-Cone figures](#)
- [Scatter/Gather examples](#)
- Questions:
 - Compare
 - What are the various interfaces built? How do they compare? What is the best use of each?
 - Scatter/gather
 - Explain clustering, relate it to scatter/gather.
 - What are special problems with large category systems and how can they be solved?

[Envision](#) project at Virginia Tech, [MARIAN](#) sequel

[Berkeley](#): TileBars, Multivalent documents

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

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.

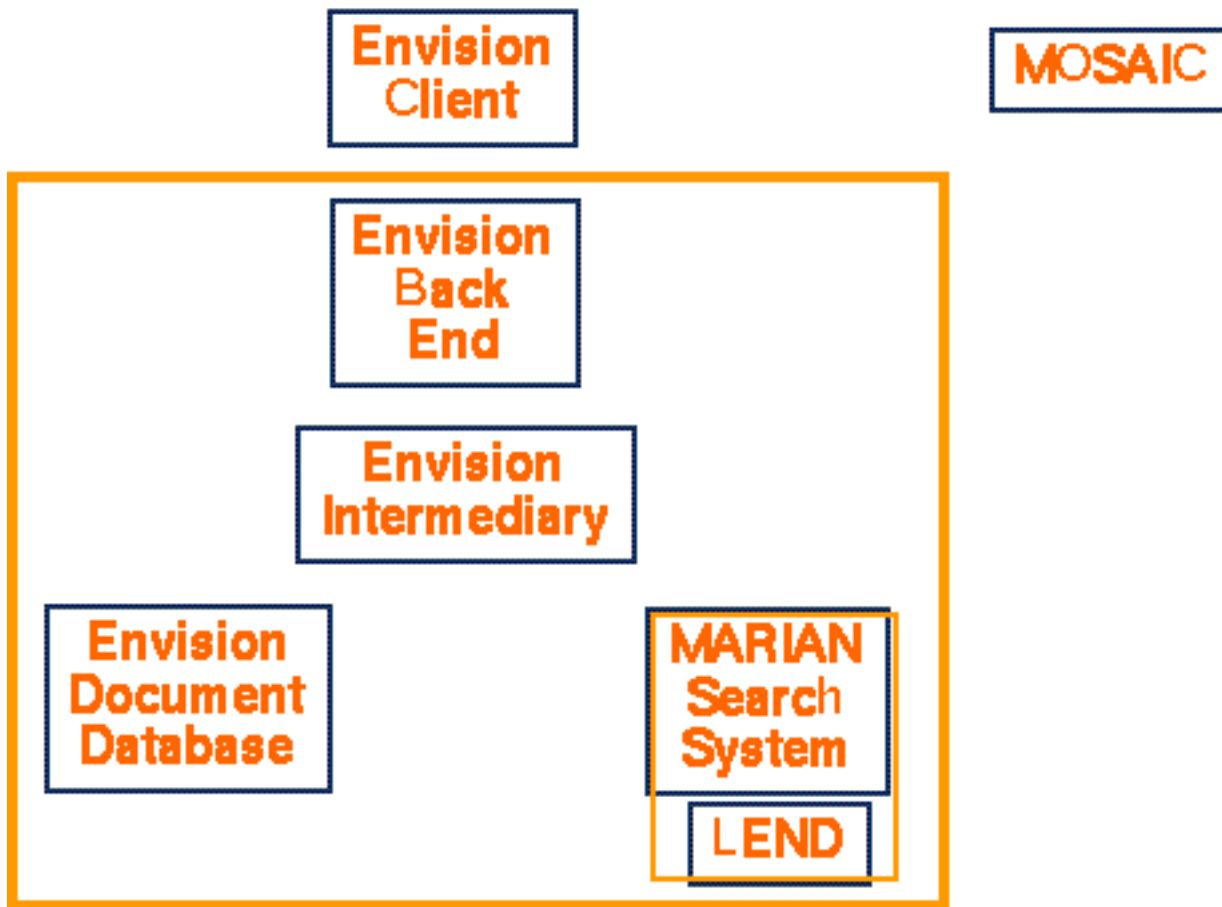
Efforts have concentrated on building an archive based upon SGML, developing an object-oriented database, applying the MARIAN retrieval system and WWW, and constructing a special search interface based upon user wishes.

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:

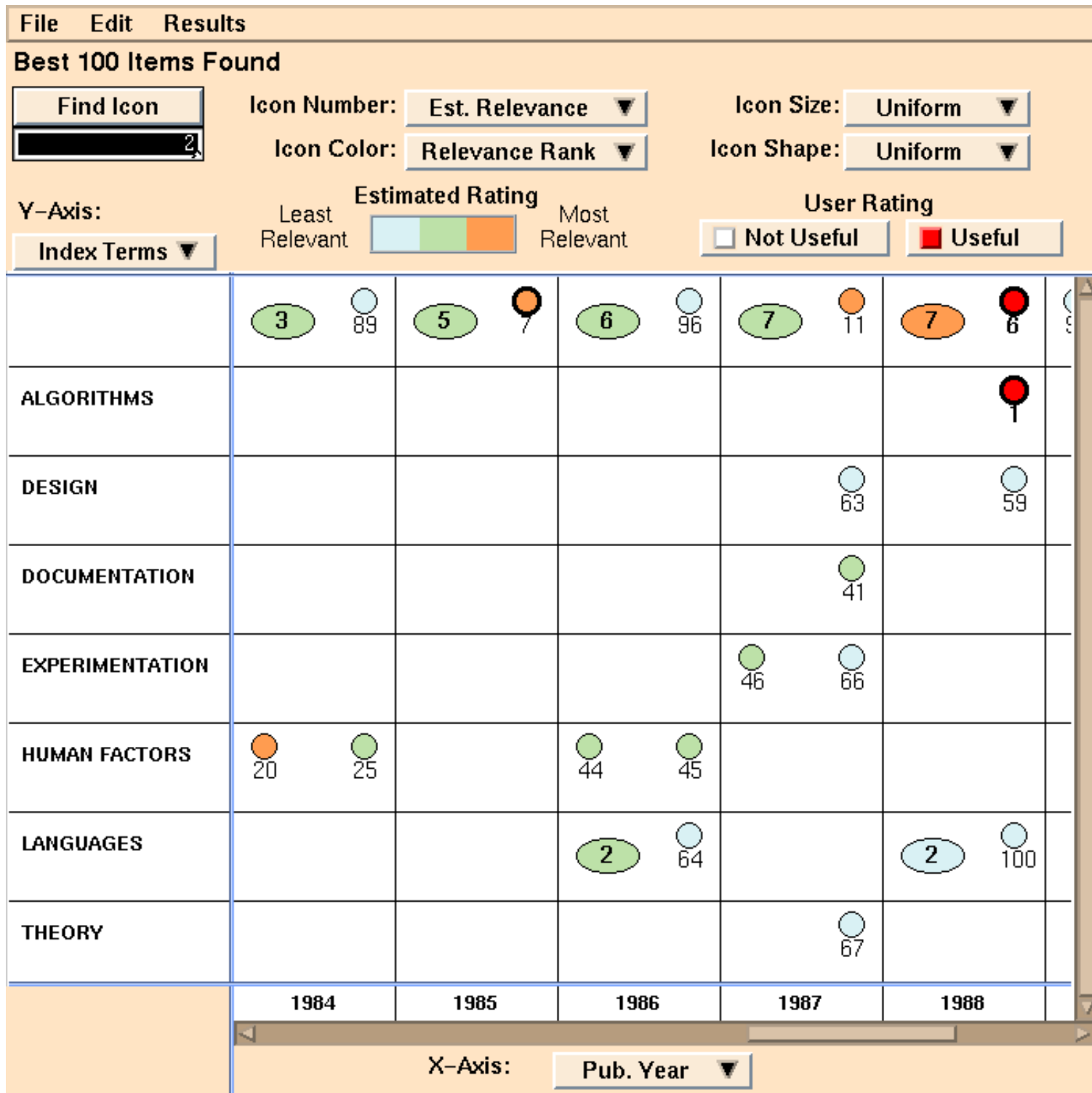
Envision



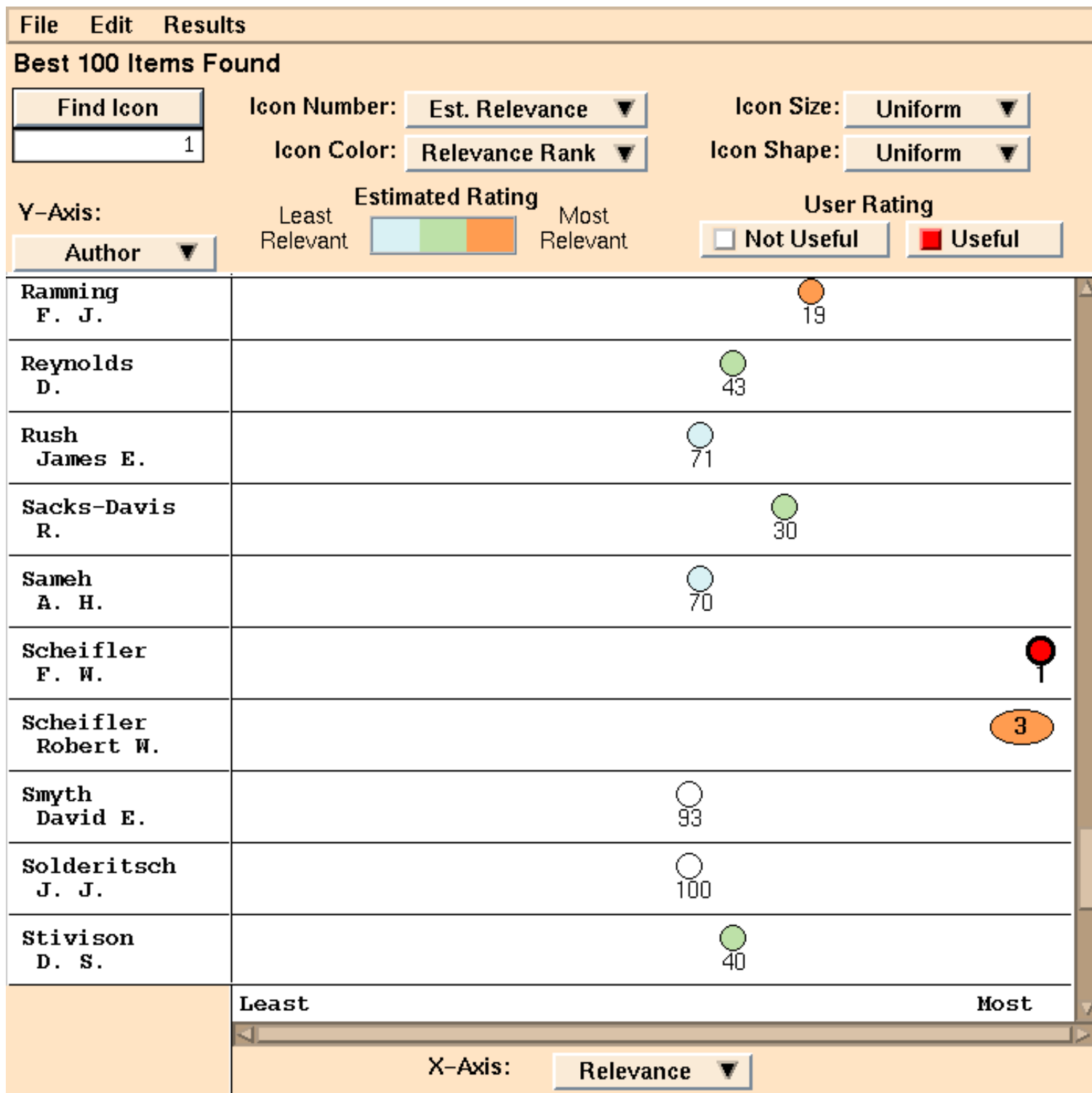
Envision - Results Screens

The interface includes:

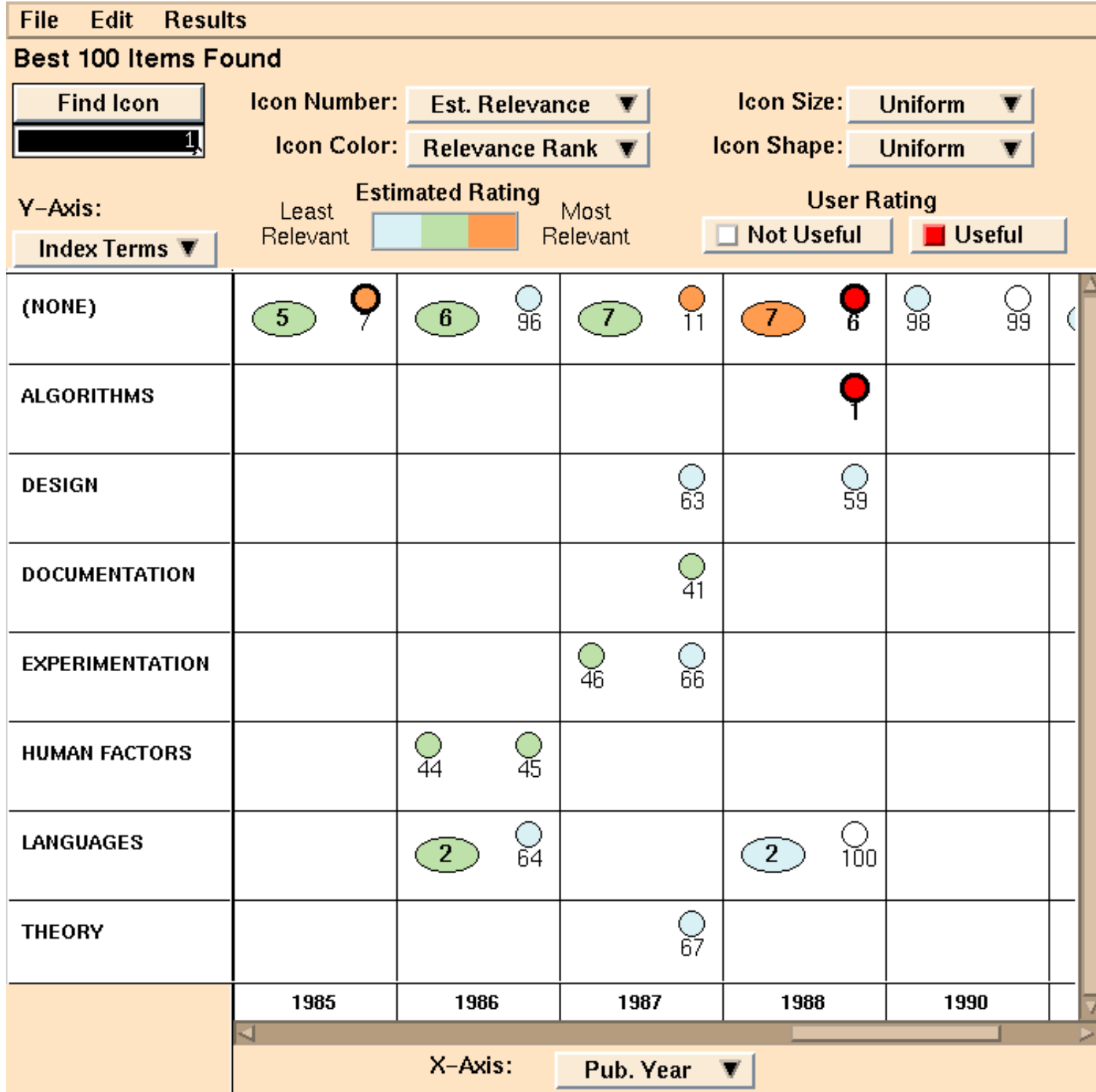
- Graphic View 1:



- Graphic View 2:



- Graphic View 3:



Metadata:

- [IMS Metadata](#)
- [Metadata: the Foundations of Resource Description](#)
- [OCLC/NCSA Metadata Workshop Report](#)
- [OAI](#)
- [RFC-1807](#)
- [TEI](#)
- [BASIS article](#)
- [D-Lib Working Group on Metadata](#)
- [STARTS](#)
- [Dublin Core Metadata Initiative](#) (and [NISO standards effort](#))
- [Alliance Metadata Standards Working Group at NCSA](#)

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001. Edward A. Fox, Rajat Gupta

Notes on Metadata and the Web

For an overview paper on related areas, read about the [Warwick Framework](#), a container architecture for aggregating metadata.

These notes are based on the articles that appear in the Oct./Nov. 1997 issue (v. 24 no. 1) of the *Bulletin of the American Society for Information Science* (ASIS). The issue title is *Organizing Internet Resources: Metadata and the Web*.

Some of the key topics considered are:

- Dublin Core, its evolution, its adaptations
- Cataloging, MARC, and their extension to Internet
- Automatic classification: Scorpion
- Naming: URL, URN, URI, URC, DOI

Useful Links by Topic - Alphabetical

The following links are either taken from the articles in the *Bulletin* issue or relate closely and fill in helpful information.

- [InterCat Project](#)- proof-of-concept database, made of records extracted from OCLC's WorldCat, demonstrating catalog services plus Web access to resources of the Internet
- [International Conf. on Principles and Future Development of AACR](#)- related papers, on Anglo-American Cataloging Rules, and their revision
- [Persistent URLs](#)- PURLs
- [Dublin Core Home Page](#)
- [Dublin Core Elements](#)
- [Dublin Core element Coverage](#) - proposed standard
- [Center for Electronic Text in the Humanities](#)
- [EAD \(Encoded Archival Description\): SGML for Archival Finding Aids - LoC](#)
- [EAD \(Encoded Archival Description\): SGML for Archival Finding Aids - Berkeley](#)
- [UC Berkeley Finding Aids](#)
- [Cataloging Internet Resources: Manual and Practical Guide, by Nancy B. Olson](#)
- [RDF Home Page](#)- Resource Description Framework, on metadata architecture on the Web
- [UKOLN Metadata Home Page](#)- summary of pubs, projects, metadata resources from UK and beyond, definitions
- [metadata element sets crosswalks](#)- mappings and relationships between various metadata sets, including Dublin Core
- [OCLC](#) and its [Research Department](#)

- [Stuart Weibel](#) - senior research scientist at OCLC, leader of Dublin Core efforts
- [Workshops on Metadata](#)
- [Dublin Core Workshop, 4th, official report](#) - held at National Library of Australia - and a [light-hearted account](#)
- [Resource Discovery project in Australia](#)
- [National Library of Australia PANDORA Project](#) (Preserving and Accessing Networked Documentary Resources of Australia)
- [In the Company of Strangers: Challenges and Opportunities in Metadata Implementation](#) paper by Maxine Brodie, policy level issues which impact on metadata implementation at the State Library of New South Wales, Sydney, Australia
- [Architecture for Access to Government Information](#) : report, Australia, 1996
- [ERIN - Environmental Resources Information Network](#), Australia - also runs a metadata listserv
- [Core Data Elements for Land and Geographic Directories in Australia and New Zealand](#)
- [Dataset Publishing - A Means to Motivate Metadata Entry](#), by S.D. Callahan, B.D. Johnson, and E.P. Shelley - Australian Resources, NPI Theory (choice behavior)
- [meta-searcher called HotOIL that accesses both HTTP and Z39.50 servers - demo](#) - translates user requests, merges results, displays summary
- [MetaWeb project](#) - develop and disseminate metadata tools
- [GEM](#) - educational resources - which calls for adding elements like Resource Needed, Standard, Audience, Pedagogy, Quality - see [elements](#)
- [NetFirst](#) - database/directory, cataloging of Internet (uses Dewey)
- [Canadian Information by Subject](#) - info on Canada in Internet (uses Dewey)
- [BUBL Information Service, Scotland, higher education, with subject tree](#) (uses Dewey)
- [Internet Public Library Youth Division](#) (uses Dewey)
- [Blue Web'n, by Pacific Bell, to organize Web sites for students, educators, ...](#) (uses Dewey)
- [Enhancing the indexing vocabulary of DDC by C.J. Godby](#)
- [Scorpion project at OCLC](#)

Acknowledgements

Thanks are given to the authors of the respective articles, from whose contributions the notes above are derived. All distortions of their content and intention are the fault of E. Fox, who apologizes for any misrepresentation inadvertently resulting from this attempt to summarize a valuable set of interesting articles.

- Guest editors' intro. to Special Section, by Efthimis N. Efthimiadis and Allyson Carlyle
- Cataloging Internet Resources: Survey and Prospectus, by Erik Jul
- The Dublin Core: A Simple Content Description Model for Electronic Resources, by Stuart Weibel
- Uniform Resource Identifiers and the Effort to Bring "Bibliographic" Control" to the Web: An

Overview of Current Progress, by Ray Schwartz

- Options for Organizing Electronic Resources: The Coexistence of Metadata, by Sherry L. Vellucci
 - Metadata in Australia, by Carmel Maguire
 - GEM: Using Metadata to Enhance Internet Retrieval by K-12 Teachers, by Stuart Sutton and Sam G. Oh
 - From Book Classification to Knowledge Organization: Improving Internet Resource Description and Discovery, by Diane Vizine-Goetz
 - Scorpion Helps Catalog the Web, by Keith Shafer
-

Please follow the above mentioned links to find answers to the following questions:

- What is metadata?
- How many elements are in the Dublin Core?
- What are some new elements added for educators in GEM?
- Describe TEI briefly and explain how it relates to Dublin Core work.
- Explain *finding aid*.
- Describe EAD briefly and explain how it relates to cataloging archival collections.
- Where are their detailed instructions on how to catalog the internet?
- What is RDF?
- What is happening in UK re metadata?
- What mappings are there between metadata representations?
- What is the Resource Discovery project in Australia?
- What happened at the Australian metadata meeting?
- What is covered by the Dublin Core *coverage* element?
- What metadata is needed for geographic information?
- When you search on "digital library" with HotOIL, what refinements are suggested? What are the results of the default processing of your query and what sources were used? Can you find the abstract of a talk on archiving the Internet?
- What WWW search/browse services use Dewey?
- What systems are available to automatically catalog WWW pages?



Dublin Core Metadata Initiative

Making it easier to find information.

[ABOUT THE INITIATIVE](#)
[DOCUMENTS](#)
[GROUPS](#)
[RESOURCES](#)
[DCMI NEWS](#)
[TOOLS AND SOFTWARE](#)
[MEETINGS AND PRESENTATIONS](#)
[PROJECTS](#)

OVERVIEW

[About the Initiative](#)
[Contact](#)
[DCMI News](#)
[Documents](#)
[Meetings and Presentations](#)
[Projects](#)
[Resources](#)
[Tools and Software](#)
[Workshops](#)

READY REFERENCE

[DC Element Set](#)
[DC Qualifiers](#)
[FAQ](#)
[Mailing Lists](#)
[Translations](#)
[Usage Board](#)
[Usage Guide](#)

MIRRORS

[Australia](#)

(hosted by National Library of Australia)

[United Kingdom](#)

(hosted by UKOLN)



DC-2001: International Conference on Dublin Core and Metadata Applications

October 22 - 26, 2001 - Tokyo, Japan

The Dublin Core Metadata Initiative is an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models. DCMI's activities include consensus-driven working groups, global workshops, conferences, standards liaison, and educational efforts to promote widespread acceptance of metadata standards and practices.

New Document

[New Working Draft of a Library Application Profile is now available](#)

2001-08-09, This document proposes a possible application profile that clarifies the use of the Dublin Core Metadata Element Set in libraries and library-related applications and projects. It will be discussed at a meeting of the DC-Libraries Working Group in Boston, Massachusetts on 22 August in conjunction with the IFLA conference. [More Information...](#)

New Project

[ExLibris Special Collections Directory](#)

2001-08-04, The ExLibris Special Collections Directory seeks to provide a comprehensive directory of special collections libraries that support digitization and make content such as manuscripts, art images, and electronic texts available over the Internet.

General Announcements

New issue of the DCMI Update Newsletter now available

2001-08-03, The [DCMI Update newsletter, July 2001](#), is now available. The newsletter highlights news items, project and tool announcements that have occurred since the last issue.

General Announcements

Catalan translations of the Element set and Qualifiers are now available

2001-07-27, The [Biblioteca de Catalunya](#) has translated the [Dublin](#)

WORK IN PROGRESS

[Public Comment Needed](#)
[Status of Deliverables](#)

WORKING GROUPS

[Administrative Metadata](#)
[Agents](#)
[Architecture](#)
[Citation](#)
[Collection Description](#)
[Education](#)
[Government](#)
[Libraries](#)
[Registry](#)
[Standards](#)
[Tools](#)
[Type](#)
[User Guide](#)

INTEREST GROUPS

[Business](#)
[Collaboratory](#)
[Moving Pictures](#)
[Multiple Languages](#)

Electronic Publishing:

- [The SGML/XML Web Page](#)
- [CS5604 unit on SGML](#): check out the related course notes offered at Virginia Tech.
- [Elsevier](#)
[TULIP](#)

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Trouble reading the page? You are probably using a non-compliant browser. Consider upgrading, or click [here](#)



Document Formats Domain

The Extensible Stylesheet Language (XSL)

XSL is a language for expressing stylesheets. It consists of three parts: [XSL Transformations](#) (XSLT): a language for transforming XML documents, the [XML Path Language](#) (XPath), an expression language used by XSLT to access or refer to parts of an XML document. (XPath is also used by the [XML Linking](#) specification). The third part is XSL Formatting Objects: an XML vocabulary for specifying formatting semantics. An XSL stylesheet specifies the presentation of a class of XML documents by describing how an instance of the class is transformed into an XML document that uses the formatting vocabulary. For a more detailed explanation of how XSL works, see the [What Is XSL](#) page.

For background information on style sheets, see the [Web style sheets](#) resource page. XSL is developed by the W3C [XSL Working Group \(members only\)](#) whose [charter](#) is to develop the next version of XSL. XSL is part of W3C's [Style Activity](#), whose work is described in the Style [Activity Statement](#).

Specifications

- [XSLT 1.0](#) - [XPath 1.0](#) - [XSLT 1.1 \(WD\)](#) - [XSL 1.0 \(CR\)](#)

Mailing Lists

- [XSL-List](#), main list about XSL- The XSL-FO list at W3C. [Subscription information](#), [archive](#) - [XSL-FO](#): another mailing list on FOs.

Implementations

- XSLT: too many to list here. Check [xslt.com](#).- [XSL Formatter](#) (Win, free evaluation version)- [XEP](#) (Java, free evaluation version)- [FOP](#) (Java, open source)- [PassiveTeX](#) (TeX, open source)- [Unicorn FOs](#) (TeX, free Windows binaries)- [REXP](#) early implementation based on FOP- [jfor](#): FO to RTF converter (Java, Open Source)

Translations

- [XPath 1.0 \(German\)](#) - [XPath1.0 \(Russian\)](#) - [XPath1.0 \(Russian\) \(2\)](#) - [XSLT1.0 \(Japanese\)](#) - [XPath 1.0](#)



The **Internet2 Distributed Storage Infrastructure (I2-DSI)** is a replicated hosting service for Internet content and applications. The channels listed below are replicated across a distributed infrastructure consisting of servers with substantial processor and storage resources. Each user request is directed to the server closest to the requesting client in networking terms. The result is that network traffic is kept local and load is balanced among the distributed servers.

I2-DSI is a joint project of the University of Tennessee, Knoxville's [Innovative Computing Laboratory](#), the University of North Carolina at Chapel Hill's [School of Information and Library Science](#), and [Internet2](#). Contact Project Director [Micah Beck](#) (UTK) or Co-Leads [Bert J Dempsey](#) (UNC-CH) and [Terry Moore](#) (UTK).

The [online proceedings](#) of the **1999 Network Storage Symposium (NetStore '99)** are now available.

I2-DSI Channels

Servers operating in Hawaii, Indiana, North Carolina, South Dakota, Tennessee, Texas... and now Virginia Tech!

Problems accessing these channels? Contact [I2-DSI admin](#).

CPAN http://cpan.dsi.internet2.edu	The Comprehensive Perl Archive Network.
Docsouth http://docsouth.dsi.internet2.edu	Documenting the American South collections (UNC-CH AAL)
High MPEG http://highmpeg.dsi.internet2.edu	High bandwidth MPEG-1 videos for local streaming delivery.
MetaLab Linux http://linux.dsi.internet2.edu	A comprehensive Linux repository. (MetaLab)
Mandrake Linux http://linux-mandrake.dsi.internet2.edu	A Mandrake Linux repository from Virginia Tech
Mars http://mars.dsi.internet2.edu	The Mars'98 Polar Lander mission. (NASA JPL)
Netlib http://netlib.dsi.internet2.edu	Mathematical software, papers, and databases. (UTK ICL)

Commerce, Economics, Publishers:

NetBill

- [Home Page](#)
- [E-Commerce Resources](#)
- [Overview article on payment systems from IEEE Spectrum](#)
- Questions: How would this work with ETDs? What are the advantages and disadvantages relative to other approaches?

E-Commerce sites: [Yahoo](#), [Roger Clarke](#)

[The Economics of Digital Libraries by Robert M. Hayes](#)

[A Need For A Common Infrastructure: Digital Libraries and Electronic Commerce](#), Daniel Schutzer, D-Lib Magazine, April 1996

Commerce part of CS6604 lecture

- Workshop on Tech. of Terms and Conditions; Final Report to NSF - including Breakout Group Reports
- [Cornell CS 502: Computing Methods for Digital Libraries Lecture 25 Access Management Administration](#)
- [EC98, International IFIP Working Conference on Distributed Systems for Electronic Commerce](#), Hamburg, Germany, June 4-5, 1998

[Projections for Making Money on the Web](#) (Michael Lesk, Harvard Infrastructure Conference, 23-25 January 1997)

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Intellectual property rights, copyright laws and legal issues:

(Chapter 10, page 223, "Books, Bucks and Bytes", Michael Lesk)

- [Cyberspace Law for Non-Lawyers](#): This is an electronic course : a "real" course in the "real world" This site includes a discussion function which will allow you, if you are so inclined, to post your own comments and reactions to the individual messages that the instructors have mailed out.
- [Pamela Samuelson](#) and pointers based on her pages and recommendations
- [Electronic Commerce](#)
- [EC98, International IFIP Working Conference on Distributed Systems for Electronic Commerce](#), Hamburg, Germany, June 4-5, 1998
- [Stanford U. work on electronic commerce, legal pointers](#)
- Copyright law in Netherlands (in Dutch): [background home page](#), [page on intellectual property and copyright](#)

Other related references:

- Digital Copyright Protection - Peter Wayner - AP Professional - Boston, 1997
- Scholarly Publishing: The Electronic Frontier - ed. Robin P. Peek and Gregory B. Newby - The MIT Press, Cambridge, MA, 1996
- The Network Nation - Starr Roxanne Hiltz and Murray Turoff - The MIT Press, Cambridge, MA, 1994
- Ubiquitous Email ...

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta

Pamela Samuelson Plus Recommendations on Law and Digital Libraries

[Professor Pamela Samuelson](#) is one of the leading authorities on legal issues in the area of intellectual property rights (IPR). A new [MacArthur Fellow](#), a Fellow of the [Electronic Frontier Foundation](#), a Fellow of the [Cyberspace Law Institute](#), she is a Professor at the [University of California at Berkeley](#) with a joint appointment in the [School of Information Management and Systems](#) and the [School of Law](#).

For more information on this and related topics, see

- [Selected Papers by Pamela Samuelson](#)
- [Law 276: Cyberlaw](#) - by Pamela Samuelson, University CA, Berkeley
- [Infosys 296A: Future of the Information Society, Copyright & Community](#) - by Peter Lyman and Pamela Samuelson, University CA, Berkeley
- [Cyberspace Law for Non-Lawyers](#), which attracted over 20,000 subscribers, by [David Post](#), Temple U. School of Law; Lawrence Lessig, [Harvard Law School](#); [Eugene Volokh](#), [UCLA School of Law](#)
- [Crash Course in Copyright](#) from UT system, including the [Digital Library](#)
- [Copyright Management Center](#) of IUPUI, directed by [Kenneth Crews](#)
- [The ILTguide to Copyright](#) at Columbia, for educators
- [Copyright Law Materials](#) at Cornell Legal Info. Institute
- [Copyright & Fair Use](#) site of Stanford University Libraries
- [Copyright Basics Circular from the U.S. Copyright Office](#)
- [Copyright Clearance Center \(CCC\) Online](#)
- [Digital Future Coalition \(DFC\)](#)
- [IIP Policy Gateway, Harvard Information Infrastructure Project](#)
 - [Bibliography](#)
 - [Policy resources in the area of Internet governance](#), supplement to MIT Press [book](#)
 - The Impact of the Internet on Communications Policy [conference](#)
- [ALAWON](#) - ALA (American Library Association) Washington Office Newslines providing urgent and late breaking news
- [ARL Federal Relations and Information Policy Program](#), Prue Adler

Social Issues:

- Social Aspects [D-Lib Working Group](#)
- UCLA Workshop, Social Aspects of Digital Libraries, Feb. 16-17, 1996
<http://is.gseis.ucla.edu/research/dl/index.html>
 - Life Cycle http://www-lis.gseis.ucla.edu/DL/UCLA_DL_model.gif
- [The social functions of digital libraries: designing information resources for virtual communities](#) - by Peter Lyman

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998-2001, Edward A. Fox, Rajat Gupta