Practical Digital Libraries Overview

ACM MM'2001 Tutorial
by
Edward A. Fox
fox@vt.edu
http://fox.cs.vt.edu
660 McBryde Hall, M/C 0106
Department of Computer Science
Virginia Tech, Blacksburg, VA 24061 USA
September 30, 2001 (8:30am-12noon)
Ottawa
http://ei.cs.vt.edu/~dlib/tut/MM01.htm

Announcement

Overview: This tutorial will start with an overview of definitions, foundations, scenarios and perspectives. It will cover a variety of issues, including:

- search, retrieval and resource discovery;
- multimedia/hypermedia;
- metadata (e.g., Dublin Core);
- electronic publishing; SGML and XML;
- document models and representations;
- database approaches;
- agents and distributed processing;
- 2D and 3D interfaces and visualizations;
- architectures and interoperability (e.g., OAI); metrics;
- educational (e.g., CSTC, NSDL, NDLTD) and social concerns;
- commerce; and intellectual property rights.

Case studies will illustrate key concepts, including:

- Computer Science Teaching Center (www.cstc.org)
- National Science (, Mathematics, Engineering, Technology Education) Digital Library (NSF NSDL, www.nsdl.nsf.gov)
- Networked Digital Library of Theses and Dissertations (www.ndltd.org)
- Open Archives Initiative (www.openarchives.org)

Level: Introductory or intermediate

Expected audience: researchers, developers, practitioners, librarians, managers, or others who do not have extensive experience in the field of digital libraries and who want a broad overview.

Tutorial presenter: Dr. Edward A. Fox holds a Ph.D. and M.S. in Computer Science from Cornell University, and a B.S. from M.I.T. Since 1983 he has been at Virginia Polytechnic Institute and State University (VPI&SU, also called Virginia Tech), where he serves as Professor of Computer Science. He directs the Digital Library Research Laboratory, the Internet Technology Innovation Center at Virginia

Tech, and varied R&D projects. He is general chair of the First ACM/IEEE Joint Conference on Digital Libraries. He is co-editor-in-chief of ACM Journal of Educational Resources in Computing (JERIC) and serves on the editorial boards of a number of journals. He has authored or co-authored many publications in the areas of digital libraries, information storage and retrieval, hypertext/hypermedia/multimedia, computational linguistics, CD-ROM and optical disc technology, electronic publishing, and expert systems.

Key parts of tutorial

1. **VT Perspective on DLs** - Talk in PowerPoint
2. **Topical Outline** (web format)

Supplemental Information

1. **Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries.** Virginia Tech Department of Computer Science Technical Report TR-01-12, by Marcos Andre Goncalves, Edward A. Fox, Layne T. Watson, and Neill A. Kipp. July, 2001, available as both: PS and PDF
2. **Bibliography for 5S / Star**
3. **DLI Overview of DLI for BASIS - in PDF**
4. **ETD Genre and Examples - in PDF**
5. **DL'99 paper on NDLTD - in PDF**
6. **Selections from Online Courseware:**
   The online courseware accessible through the topical outline above also is accessible for self-study as WWW pages. For convenience, it also is accessible in large part as PDF files:
   - full selections (15M, 1267 pages)
   - 1st page only of selections (9M, 348 pages)
   - outline without many selections (1M, 76 pages)

Tutorial Schedule

First, all tutorial materials will be examined, to orient attendees.

Then the large set of PowerPoint slides will be discussed.

Next, the topical outline will be considered, going through at a high level. This corresponds to the 76 pages of PDF file, which are included in the handout. The presentation will extend this to the 348 page version, as time permits.

Finally, the other supplemental information will be summarized, so attendees can continue their follow-up studies about digital libraries.

Virginia Tech Perspective on Digital Libraries: From Hardware to Software to Projects to Theory

Fall 2001

Edward A. Fox

fox@vt.edu http://fox.cs.vt.edu

CS DLRL Internet TIC

Virginia Tech, Blacksburg, VA, USA

Outline

- **Virginia Tech context**
- **Why DLs? What are DLs? (5S theory)**
- **Case Study: WCA**
- **Case Study: Education: CSTC -> NSDL**
- **Case Study: NDLTD**
- **Accessibility and Visualization**
- **DL Software: MARIAN**
- **DL Hardware: PetaPlex**
- **Interoperability: OAI**

Acknowledgements (Selected)

- **Sponsors:** ACM, Adobe, IBM, Microsoft, NSF, OCLC, SOLINET, SURA, US Dept. of Ed. (FIPSE), …
- **VT Faculty/Staff:** Marc Abrams, Tony Atkins, Thomas Dunbar, Debra Dudley, John Eaton, Gwen Ewing, Peter Haggerty, H. Rex Hartson, Deborah Hix, Gary Hooper, Sunny Kim, JAN Lee, Manan HoLee, Gail McMillan, Len Peters, James Powell, Shalini Urs, …
- **VT Students:** Emilio Arce, Fernando Das Neves, Brian DeVane, Robert France, Marcos Goncalves, Scott Guyer, Robert Hall, Neill Kipp, Paul Mather, Tim McGonigle, Todd Miller, Constantinos Phanouriou, William Schweiker, Ohm Sornil, Hussein Suleman, Patrick Van Metre, Laura Weiss, Wensi Xi, …

Virginia Tech Background

- Largest university in Virginia, land- grant, football, town population 35K plus 26K students
- Blacksburg Electronic Village, since 1992, with > 80% of community on Internet
- Net.Work.Virginia, with sites for education, research, government
- LMDS, Local Multipoint Distribution Service, gigabit wireless networking- 1/3 of Virginia
- Math Emporium, 500 workstations
- Faculty Development Initiative, round 3
- Torgersen Hall, $30M Advanced Communications and Information Technology Center, with DLRL

Internet Technology Innovation Center

Supported by Virginia's Center for Innovative Technology

Statewide University Partners - Governing Board:

- **Christopher Newport University**
  - William Winter, William Muir, Virginia Electronic Commerce Technology Center / Southeastern Virginia Network (VECTEC/SEVAnet)
- **George Mason University**
  - Steven Ruth, International Center for Applied Studies in IT (ICASIT)
- **Old Dominion University –** Kurt Maly (CS Head), …
- **University of Virginia**
  - Alf Weaver, Internet Commerce Group (InterCom)
  - Jim French, Internet Digital Library
- **Virginia Tech**
  - Edward Fox, Digital Library Research Laboratory (DLRL), CC, CS
  - Scott Midkiff, Center for Wireless Telecomm. (CWT), VTISC, ECpE

ITIC @ VT

Research Areas

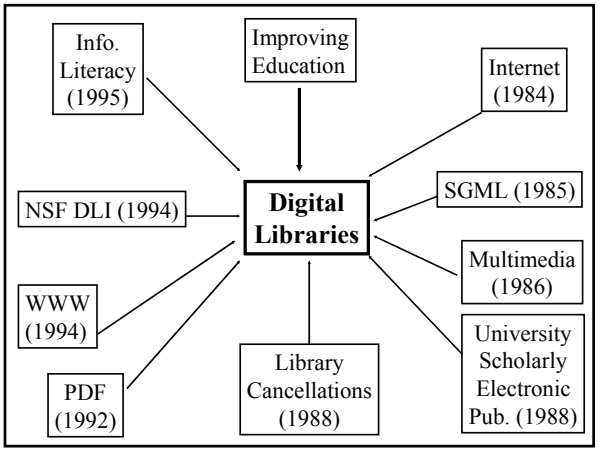
- Collaboration (e.g., group decision support)
- Community networking (e.g., BEV)
- Internet access (e.g., statewide network)
- Information services (e.g., digital libraries)
- Modeling and simulation (e.g., Web traffic)
- Usability (e.g., human factors engineering)
- Virtual environments (e.g., CAVE, visualization)

Digital Libraries --- Virginia Tech

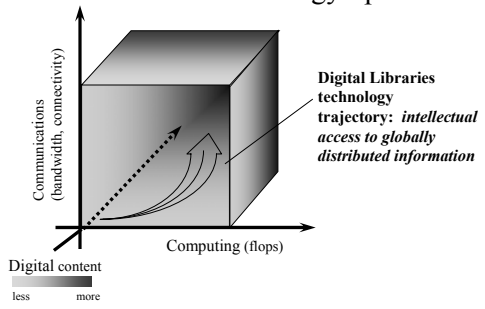
- MARIAN (NLM)
- CS DL Prototype - ENVISION (NSF, ACM)
- TULIP (Elsevier, OCLC)
- BEV History Base (NSF, Blacksburg)
- DL for CS Education - EI (NSF, ACM)
- WATERS, NCSTRL (NSF)
- NDLTD (SURA, US Dept. of Education)
- CSTC (NSF, ACM), CRIM (NSF, SIGMM)
- WCA (Log) Repository (W3C)
- VT-PetaPlex-1 (Knowledge Systems)

Outline

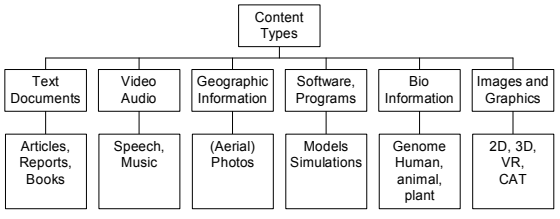
- **Virginia Tech context**
- **Why DLs? What are DLs? (5S theory)**
- **Case Study: WCA**
- **Case Study: Education: CSTC -> NSDL**
- **Case Study: NDLTD**
- **Accessibility and Visualization**
- **DL Software: MARIAN**
- **DL Hardware: PetaPlex**
- **Interoperability: OAI**



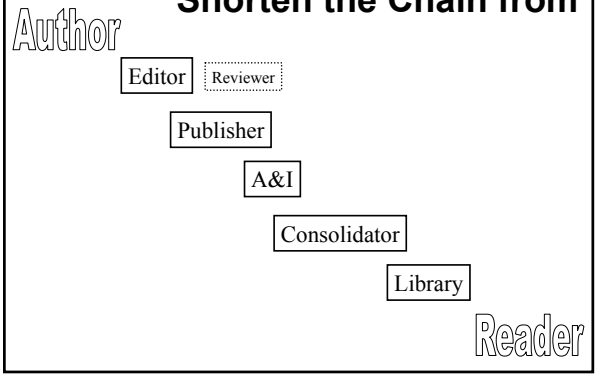
Locating Digital Libraries in Computing and Communications Technology Space



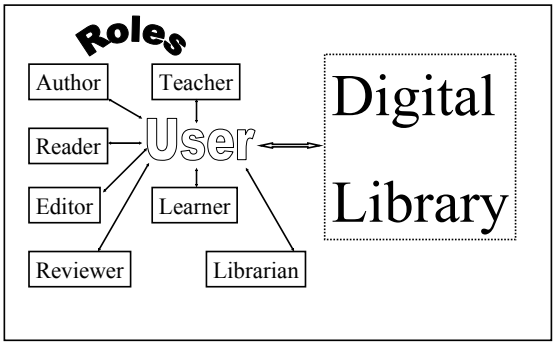
Digital Library Content



Digital Libraries Shorten the Chain from



DLs Shorten the Chain to



Digital Libraries --- Objectives

- World Lit.: 24hr / 7day / from desktop
- Integrated "super" information systems: 5S: streams, structures, spaces, scenarios, societies
- Ubiquitous, Higher Quality, Lower Cost
- Education, Knowledge Sharing, Discovery
- Disintermediation -> Collaboration
- Universities Reclaim Property
- Interactive Courseware, Student Works
- Scalable, Sustainable, Usable, Useful

Benefits

- Ease of use
- Effectiveness
- "The benefits of digital libraries will not be appreciated unless they are easy to use effectively." - IITA Workshop report

DLs: Why of Global Interest?

- **National projects** can preserve antiquities and heritage: cultural, historical, linguistic, scholarly
- Knowledge and information are essential to economic and technological **growth, education**
- DL - a **domain for international collaboration**
  - wherein all can **contribute** and **benefit**
  - which leverages investment in **networking**
  - which provides useful **content** on Internet & WWW
  - which will **tie nations and peoples together** more strongly and through **deeper understanding**

DL Challenges

- Preservation - so people with trust DLs
- Supporting infrastructure - networks, ...
- Scalability, sustainability, interoperability
- DL industry - critical mass by covering libraries, archives, museums, corporate info, govt info, personal info - "quality WWW" integrating IR, HT, MM, ...
  - Need tools & methods to make them easier to build

Digital Library Courseware

- http://ei.cs.vt.edu/~dlib/
- WWW pages or large PDF copy files
- Online quizzes based on book by Michael Lesk (Morgan Kaufmann Publishers)
- Contents based on book, with several other popular topics added (e.g., agents)
- Separate pages to supplement: Definitions, Resources (People, Projects), and References

Definitions

- Library ++ (library+archive+museum+…)
- Distributed information system + organization + effective interface
- User community + collection + services
- Digital objects, repositories, IPR management, handles, indexes, federated search, hyperbase, annotation

Definition: Digital Libraries are complex systems that

- help satisfy info needs of users (societies)
- provide info services (scenarios)
- organize info in usable ways (structures)
- present info in usable ways (spaces)
- communicate info with users (streams)

5S Layers

Societies

Scenarios

Spaces

Structures

Streams

Document Models, Representations, and Accesses

- Doc = stream + structure + use scenario; hybrid (paper/electronic), digital only
- Multilingual: content, summary, metadata
- Multimedia: structure, quality (oS), search
- Structured: MARC, SGML, by user: MVD
- Distributed collection: Kleisli, CIMI, Z39.50
- Federated search: collecting, picking site(s), parallel search / fall-back, fusing results
- Access: IPR, payment, security, scenarios

Architectural Issues

- Internet middleware
- Independent system / part of federation
- Decompositions vary
  - search engine, browser, DBMS, MM support
  - repository, handle server, client
  - information resources + mediators, bus or agent collection + client with workspace/environment
- Metrics: e.g., for federated search

Standards

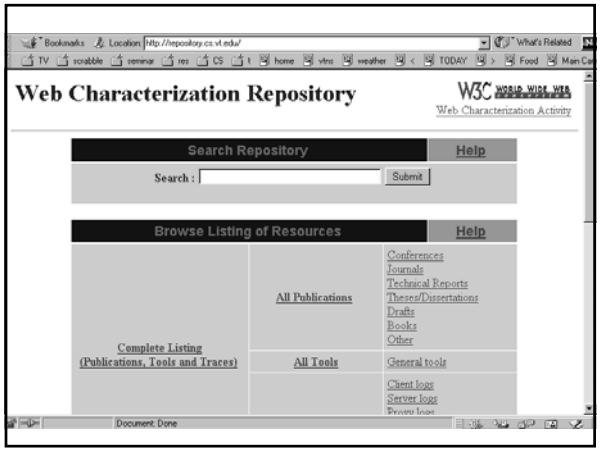
- Protocols/federation
  - Z39.50, CIMI
  - Dienst, NCSTRL
  - OAI protocol
- Metadata
  - TEI: inline, detailed (structure in stream)
  - MARC: two-level, fine-grained
  - Dublin Core: high-level, 15 elements
  - RDF: describing resources/collections, annotation
  - OAMS-->DC and others used in OAI

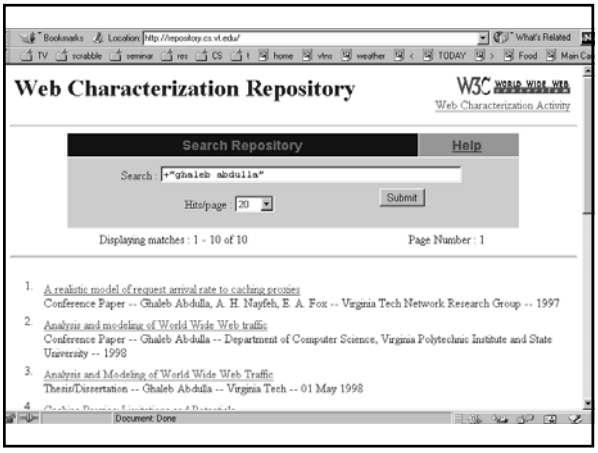
Outline

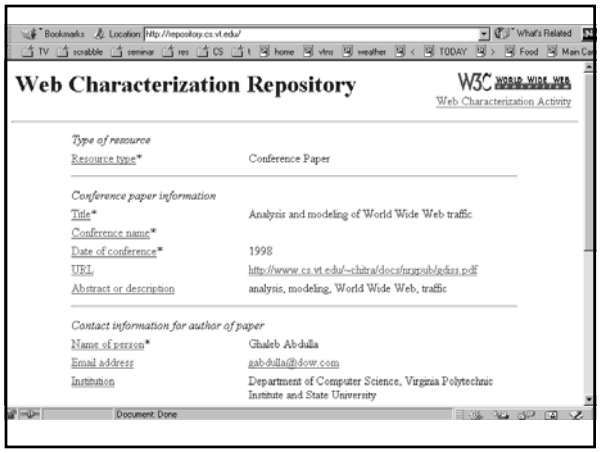
- **Virginia Tech context**
- **Why DLs? What are DLs? (5S theory)**
- **Case Study: WCA**
- **Case Study: Education: CSTC -> NSDL**
- **Case Study: NDLTD**
- **Accessibility and Visualization**
- **DL Software: MARIAN**
- **DL Hardware: PetaPlex**
- **Interoperability: OAI**

W3C Web Characterization Repository

- Online database of metadata related to publications, tools and data sets dealing with Web characterization
- Project of the Web Characterization Activity working group of the World-Wide-Web Consortium (www.w3c.org/WCA)
- http://purl.org/net/repository







Outline

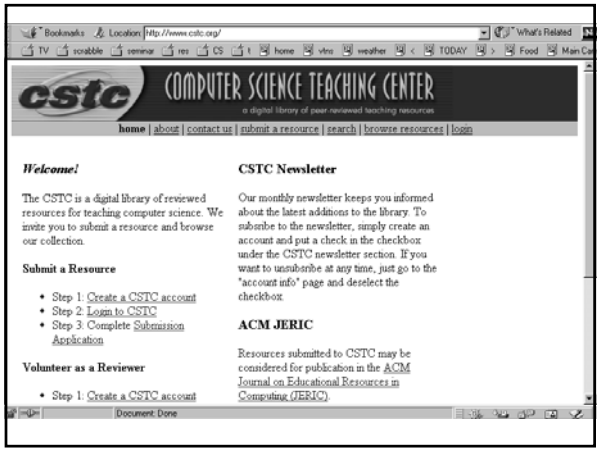
- **Virginia Tech context**
- **Why DLs? What are DLs? (5S theory)**
- **Case Study: WCA**
- **Case Study: Education: CSTC -> NSDL**
- **Case Study: NDLTD**
- **Accessibility and Visualization**
- **DL Software: MARIAN**
- **DL Hardware: PetaPlex**
- **Interoperability: OAI**

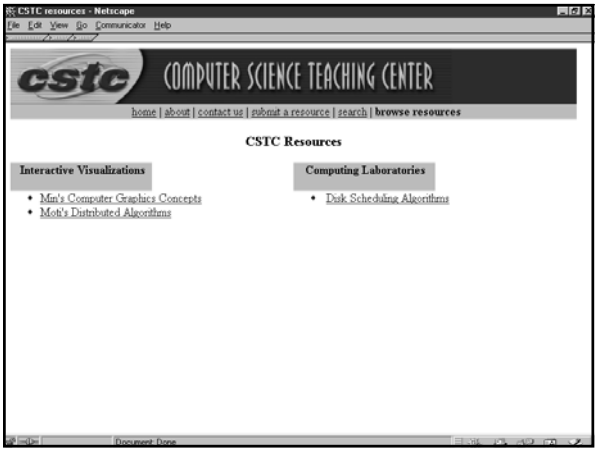
CS -> CSTC -> CRIM

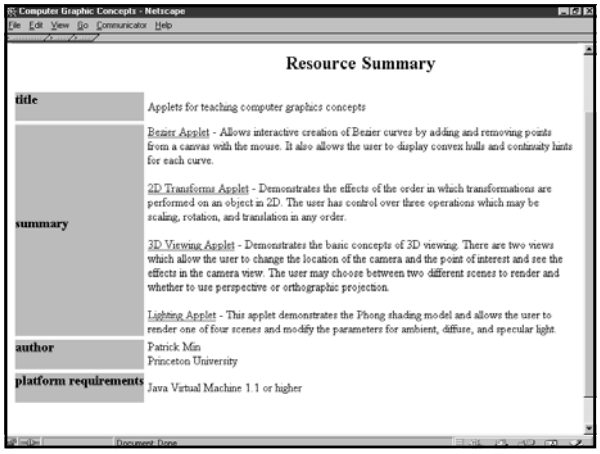
- NSF and ACM Education Committee are funding a 2 year project "A Computer Science Teaching Center" - CSTC - http://www.cstc.org/
- College of NJ, U. Ill. Springfield, Virginia Tech
- Focus initially on labs, visualization, multimedia
- Multimedia part is also supported by a 2nd grant to Virginia Tech and The George Washington University: http://www.cstc.org/~crim/ (with curricular guidelines also under development)

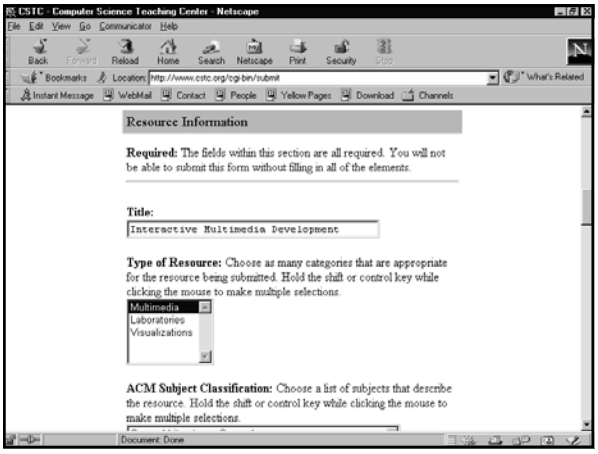
CS Teaching Center (CSTC)

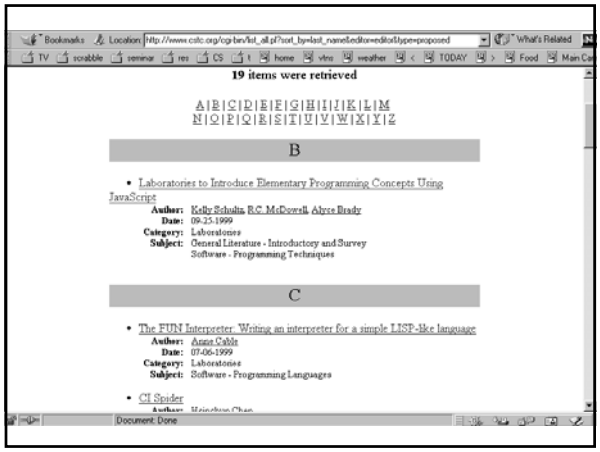
- Instead of building large, expensive multimedia packages, that become obsolete and are difficult to re use, concentrate on **small knowledge units**.
- Learners benefit from having well crafted modules that have been **reviewed and tested**.
- Use digital libraries to build a **powerful base** of support for learners, upon which a variety of courses, self study tutorials & reference resources can be built. [See NSF NSDL- National Science (math, engineering, technology education) Digital Library (formerly SMETE lib) at http://www.dlib.org/smete/public/smete-public.html]
- ACM Education Board and SIG support, new NSF grant with COLLEGIS Research Institute/Eduprise and others …

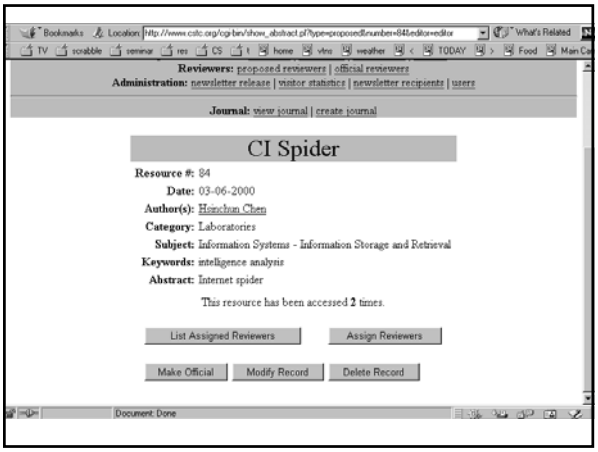












Curriculum Resources in Interactive Multimedia (CRIM)

- MM field needs properly trained personnel
- Support this with resources + curricula
- Benefits will go to teachers (who have more to build upon) and students (who will have a richer environment for learning
- CSTC, CRIM have led to ACM Journal of Educational Resources in Computing, **JERIC**
- Together these help us move forward: DL for Interactive MM -> CS -> NSDL

CRIM Project Activities

- Workshops, other ways to involve community
- WWW site including DL in CSTC re MM
  - Devised cataloging schema, designed interface
  - Referring to all MM syllabi and curriculum
  - Inviting learning resources for the CRIM DL, with reviews, reuse certifications
- Publish report on MM curriculum through ACM and IEEE, after careful review
- Introducing into CC2001: information retrieval, hypertext/hypermedia, multimedia, digital libraries

SMETE Library -> NSDL (from www.dlib.org to NSF DLI-2)

- Context: Global movement toward Digital Libraries (see April 1998 CACM)
- NSF 00-44 effort: Science, Mathematics, Engineering, and Technology Education Digital Library (focussed on undergraduates)
  - 3 workshops, yearly increasing funds / new calls
- NSDL will operate as a distributed federation, with separate parts for each key discipline, and should lead to a global effort.

Selected NSDL Early Projects/Topics

COLLEGIS Res. Inst.	IMS, **CS,** Math, Viz., …
Columbia University	Earth sciences
Stanford University	Medicine (images)
U. California Berkeley	Engineering
University of Maryland	K-12 education
U. Texas at Austin	Physical anthropology

Outline

- **Virginia Tech context**
- **Why DLs? What are DLs? (5S theory)**
- **Case Study: WCA**
- **Case Study: Education: CSTC -> NSDL**
- **Case Study: NDLTD**
- **Accessibility and Visualization**
- **DL Software: MARIAN**
- **DL Hardware: PetaPlex**
- **Interoperability: OAI**

A Digital Library Case Study

- Domain: graduate education, research
- Genre:ETDs=electronic theses & dissertations
- Submission: http://etd.vt.edu
- Collection: http://www.theses.org

Project:

Networked Digital Library of Theses & Dissertations

(NDLTD)

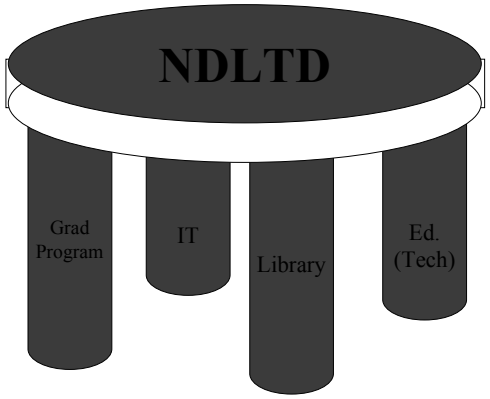
http://www.ndltd.org

The Networked Digital Library of Theses and Dissertations

www.NDLTD.org

Training Authors
Expanding Access
Preserving Knowledge
Improving Graduate Education
Enhancing Scholarly Communication
Empowering Students & Universities

Leader of the Worldwide ETD (Electronic Thesis and Dissertation) Initiative

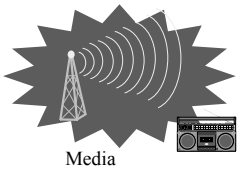


ETDs Got Your Interest?

ETD Web Site
http://www.ndltd.org/

Graduate Students

U. Laval



Media

Singapore AM
Chronicle of Higher Ed.
National Public Radio
NY Times ...

Key Ideas:

Scalability

- Networked infrastructure
- University collaboration
- Workflow, automation
- Education is the rationale
  - 8th graders vs. grads
  - Authors must submit

Maximal Access

- Standards
  - PDF, SGML, MM, MARC, DC, URNs, Federated search

What led to today's meeting?

- 1987 mtg in Ann Arbor: UMI, VT, …
- 1992 mtg in Washington: CNI, CGS, UMI, VT and 10 universities with 3 reps each
- 1993 mtg in Atlanta to start Monticello Electronic Library (regional, US Southeast): SURA, SOLINET
- 1994 mtg at VT: std: PDF + SGML + multimedia objects
- 1996 funding by SURA, US Dept. of Education (FIPSE)
- 1997 meetings in UK, Germany, ...
- 1998 – 1st symposium – Memphis (20)
- 1999 – 2nd symposium – Blacksburg (70)
- 2000 – 3rd symposium – St. Petersburg (225)
- 2001 – 4th symposium – Caltech (200)
- 2002- May 30 – June 1, BYU; 2003 – Spring, in Berlin

What are the long term goals?

- 400K US students / year getting grad degrees are exposed / involved
- 200K/yr rich hypermedia ETDs that may turn into electronic portfolios (images, video, audio, …)
- Dramatic increase in knowledge sharing: literature reviews, bibliographies, …
- Services providing lifelong access for students: browse, search, prior searches, citation links
- Hundreds/thousands of downloads / year / work

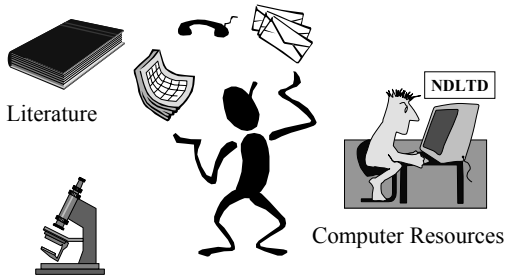
ETDs: Library Goals

- Improve library services
  - Better turn-around time
  - Always available
- Reduce work
  - catalog from e-text
  - eliminate handling: mailing to UMI, bindery prep, check-out, check-in, reshelving, etc.
- Save space

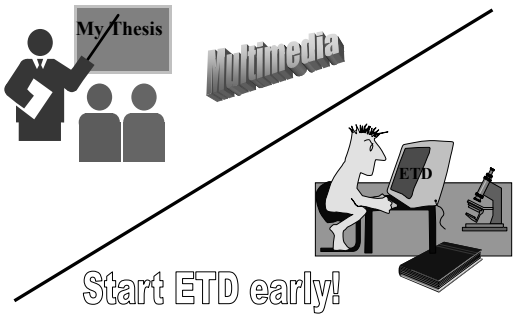
What are we doing?

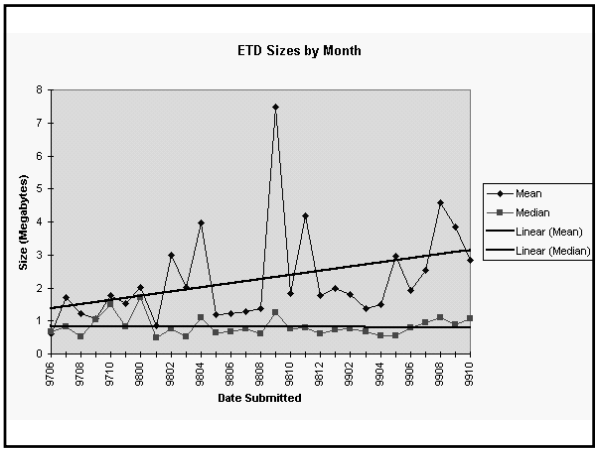
- Aiding universities to enhance graduate education, publishing and IPR efforts
- Helping improve the availability and content of theses and dissertations
- Educating ALL future scholars so they can publish electronically and effectively use digital libraries (i.e., are Information Literate and can be more expressive)

Student Prepares Thesis/Dissertation

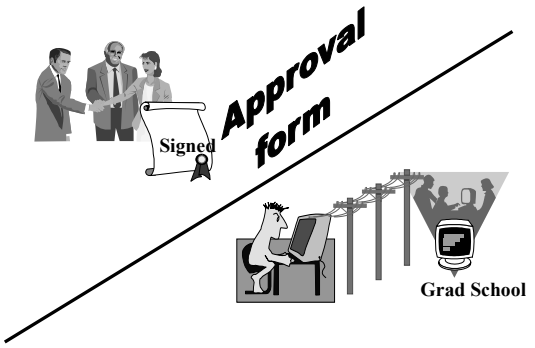


Student Defends & Finalizes ETD

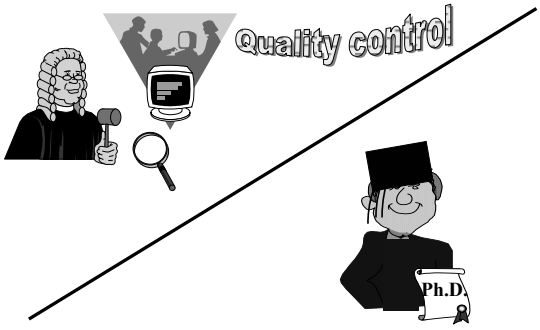




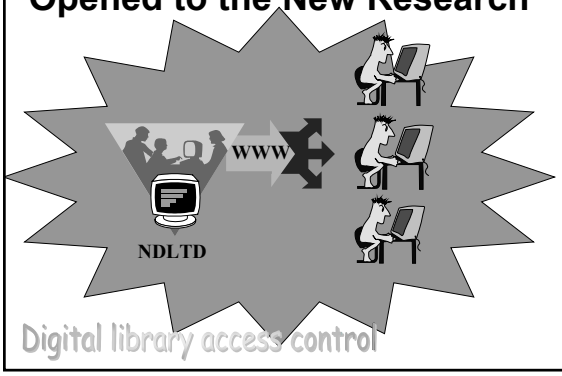
Student Gets Committee Signatures and Submits ETD



Graduate School Approves ETD, Student is Graduated



Library Catalogs ETD, Access is Opened to the New Research



Status of the Local Project

- Approved by university governance Spring 1996; required starting 1/1/97
- Submission & access software in place
- Submission workshops for students (and faculty) occur often: beginner/adv.
- Faculty training as part of Faculty Development Initiative
- Over 3000 ETDs in collection – some have audio, video, large images, software, …

Archiving ETDs

- Every 15 minutes back-ups made of not-yet-approved submissions
- Hourly back-ups of newly approved ETDs
- Weekly back-ups of entire ETD collection
- Copies stored on-site and off-site

VT ETD Cataloging

- same as current cataloging policies, except:
  - author-assigned keywords (not LCSH)
  - generic (not LC) call no.
  - fields/subfields as required for computer files
  - full abstracts
- time savings
  - cataloger familiar with computer files
  - equipment, software for word processing
  - 5 minutes avg. (10-15 minutes for paper TDs)

Library Costs

- $12/vol. for paper thesis processing
  - catalog, bind, security strip, label, shelve
  - @950 vols./yr. = $11,466
- $3.20/vol. ETD processing
  - cataloging @950 vols./yr. = $3040
- $.07/vol. shelving
- $.04/vol. circulation

Costs/Savings at VT

- Graduate School stopped shipping to the library 3000 copies of paper TDs/year
- Library stopped binding, shelving, and circulating 3000 copies of TDs/year
- 166 ft of shelf space saved/year by the library
- VT used existing equipment in Library (vs. start-up costs for staff, hardware and software from from a zero-base estimate: $65,000 – see http://scholar.lib.vt.edu/theses/)

Popular Works 1996

458 Seevers, Gary L. Identification of Criteria for Delivery of Theological Education Through Distance Education: An International Delphi Study (Ph.D., Educational Research and Evaluation, April 1993; 1353Kb)

432 Hohauser, Robyn Lisa. The Social Construction of Technology: The Case of LSD (MS in Science and Technology Studies, Feb. 1995; 244Kb)

390 Childress, Vincent William. The Effects of Technology Education, Science, and Mathematics Integration Upon Eighth Grader's Technological Problem-Solving Ability (Ph.D. in Vocational and Technical Education, July 1994; 285Kb)

310 Kuhn, William B. Design of Integrated, Low Power, Radio Receivers in BiCMOS Technologies (Ph.D. in Electrical Engineering, Dec. 1995; 2Mb)

287 Sprague, Milo D. A High Performance DSP Based System Architecture for Motor Drive Control (MS in Electrical Engineering, May 1993; 878Kb)

165 Wallace, Richard A. Regional Differences in the Treatment of Karl Marx by the Founders of American Academic Sociology (MS in Sociology, Nov. 1993; 479Kb)

150 McKeel, Scott Andrew. Numerical Simulation of the Transition Region in Hypersonic Flow (Ph.D. in Aerospace Engineering, Feb. 1996; 3Mb)

Popular Works 1997

9920 Liu, Xiangdong. *Analysis and Reduction of Moire Patterns in Scanned Halftone Pictures* (Ph.D. in Computer Science, May 1996; 6.6Mb)

7656 Petrus, Paul. *Novel Adaptive Array Algorithms and Their Impact on Cellular System Capacity* (Ph.D. in Electrical Engineering, March 1997; 5Mb)

2781 Agnes, Gregory Stephen. *Performance of Nonlinear Mechanical, Resonant-Shunted Piezoelectric, and Electronic Vibration Absorbers for Multi-Degree-of-Freedom Structures* (Ph.D. in Engineering Mechanics, Sept. 1997; ? + 7926Kb)

2492 Gonzalez, Reinaldo J. *Raman, Infrared, X-ray, and EELS Studies of Nanophase Titania* (Ph.D. in Physics, July 1996; 4607Kb)

1877 Shih, Po-Jen. *On-Line Consolidation of Thermoplastic Composites* (Ph.D. in Engineering Mechanics, Feb. 1997; 3.3Mb)

1791 Saldanha, Kevin J. *Performance Evaluation of DECT in Different Radio Environments* (MS in Electrical Engineering, Aug. 1996; 3.2Mb)

1431 DeVaux, David. *A Tutorial on Authorware* (MS in CS, April 1996; 2.3Mb)

1394 Kuhn, William B. *Design of Integrated, Low Power, Radio Receivers in BiCMOS Technologies* (Ph.D. in Electrical Engineering, Dec. 1995; 2518Kb)

Institutional Members

- **Cinemedia**
- **Coalition for Networked Information (CNI)**
- **Committee on Institutional Cooperation (CIC)**
- **Consorci de Biblioteques Universitàries de Catalunya**
- **Diplomica.com**
- **Dissertation.com**
- **Dissertationen Online (Germany)**
- **ETDweb, a Division of Answer4.com**
- **Ibero-American Science & Technology Education Consortium (ISTEC)**
- **National Documentation Centre (NDC), Greece**
- **National Library of Portugal (for all universities)**
- **OCLC Online Computer Library Center**
- **OhioLINK**
- **Organization of American States (SEDI/OAS)**
- **Southeastern Library Network (SOLINET)**
- **UNESCO (www.unesco.org/webworld/etd)**

National / Regional Projects

- **Australia**
  - U. New South Wales (lead)
  - U. of Melbourne
  - U. of Queensland
  - U. of Sydney
  - Australian National U.
  - Curtin U. of Technology
  - Griffith U.
- **Germany**
  - Humboldt University (lead)
  - 3 other universities
  - 5 learned societies: Math, Physics, Chemistry, Sociology, Education
  - 1 computing center
  - 2 major libraries
- OhioLINK: 79 colleges/univs
- Consorci de Biblioteques Universitàries de **Catalunya**, as group, www.cbuc.es:
  - Universitat de Barcelona
  - Universitat Autonòma de Barcelona
  - Universitat Politècnica de Catalunya
  - Universitat Pompeu Fabra
  - Universitat de Girona
  - Universitat de Lleida
  - Universitat Rovira i Virgili
  - Universitat Oberta de Catalunya
  - Biblioteca de Catalunya

OhioLINK

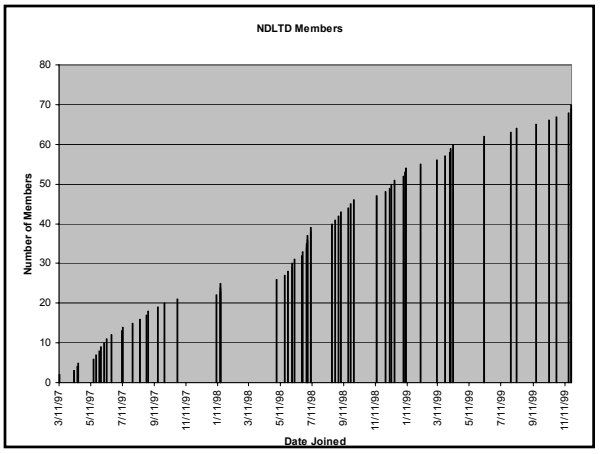
- Statewide Consortium
- Represents 79 colleges, universities, libraries
- Public Universities
- Private Universities and Colleges
- 2-Year Colleges
- Only a few (e.g., Miami U. of Ohio) are also NDLTD members on their own

US University Members (52)

- **Air University (Alabama)**
- **Baylor University**
- **Brigham Young University (part, whole)**
- **Caltech**
- **Clemson University**
- **College of William & Mary**
- **Concordia University (Illinois)**
- **East Carolina University**
- **East Tenn. State U. – require fall 2000**
- **Florida Institute of Technology**
- **Florida International University**
- **George Washington University**
- **Louisiana State University**
- **Marshall University (W. Va.)**
- **Miami University of Ohio**
- **Michigan Tech**
- **Mississippi State University**
- **MIT**
- **Montana State University**
- **Naval Postgraduate School (CA)**
- **New Jersey Inst. of Technology**
- **New Mexico Tech**
- **North Carolina State University**
- **Northwestern University**
- **Penn. State University**
- **Regis University**
- **Rochester Institute of Tech.**
- **Texas A&M**
- **U. of Colorado Health Science Center**
- **U. of Florida**
- **U. of Georgia**
- **University of Hawaii, Manoa**
- **U. of Iowa**
- **U. of Kentucky**
- **U. of Maine**
- **U. of North Texas – required since 8/99**
- **U. of Oklahoma**
- **U. of Pittsburgh**
- **U. of Rochester**
- **U. of South Florida**
- **U. of Tennessee, Knoxville**
- **U. of Tennessee, Memphis**
- **U. of Texas at Austin – required in 2001**
- **U. of Virginia**
- **U. of West Florida**
- **U. of Wisconsin - Madison**
- **Vanderbilt U.**
- **Virginia Commonwealth U.**
- **Virginia Tech - required since 1/97**
- **West Virginia U. - required fall 1998**
- **Western Michigan U.**
- **Worcester Polytechnic Inst.**

Other Countries - 52 Members

- Australia
- Belgium
- Brazil
- Canada
- China, Hong Kong
- Columbia
- Germany
- India
- Italy
- Korea
- Mexico
- Netherland
- Norway
- Russia
- Singapore
- S. Africa
- S. Korea
- Spain
- Sudan
- Sweden
- Taiwan
- UK



Type 1 Members

University Requires ETDs

- Adobe Acrobat and/or XML/SGML tools
- Automated submission & processing
- Archive/access through UMI, (OCLC,) Virginia Tech, ...
- (Local) WWW site, publicity
- (Local) Assistance provided as requested: email, phone, listserv(s)

Type 2 Members

University Agrees to Require ETDs

- Like Type 1 but set date not reached
- Usually has an option or pilot
- May: wait for new AY; start with all who enter after; …
- Build grass roots support
  - **Advisory committee: representative? expert?**
  - **Champions to spread by word of mouth**
  - **Approval: Senates, Commissions, Deans, Students**
  - **Publicity to reach community**

NDLTD Members, Types 3-7

- 3. Part of university requires ETDs
- 4. University allows ETDs
- 5. University investigating, has pilot
- 6. University consortium joins:
  - CIC (Big 10 coordinating body)
- 7. Non-university organization joins
  - CNI (Coalition for Networked Info.)

Counts of ETDs at Selected U's

University/Institution	ETD Collection Size
ADT: Australian Digital Thesis Program (Australia)	238
University of Bergen (Norway)	45
California Institute of Technology	2
Consorci de Biblioteques Universitaries de Catalunya (Spain)	151
East Tennessee State University	106
Humboldt-University (Germany)	430
Louisiana State University	3
Mississippi State University	33
MIT	62
North Carolina State University	301
Pennsylvania State University	83
Pontifical Catholic University (PUC) (Brazil)	90
Gerhard Mercator Universitat Duisburg (Germany)	126
Universitat Politecnica de Valencia (Spain)	189
University of Florida	174
(continued)	

Counts of ETDs at Selected U's (cont'd)

University/Institution	ETD Collection Size
University of Georgia	121
University of Iowa	6
University of Kentucky	19
University of Maine	27
University of North Texas	337
University of South Florida	25
University of Tennessee	12
University of Tennessee, Knoxville	28
Uppsala University (Sweden)	178
Virginia Tech	3393
West Virginia University	1006
Worcester Polytechnic Institute	83
TOTAL	**7268**

Counts of University Scanned ETD Collections

University/Institution	ETD Collection Size
MIT	5,581
National Documentation Center, Greece	12,000
New Jersey Institute of Technology	26
University of South Florida	150
TOTAL	**17,763**

VT ETD Access Logs

	1997/98	1998/99	Increase 1997/98 - 1998/99	1999/00	Increase 1998/99 - 1999/00
Requests for PDF files (mostly full ETDs)	221,679	481,038	117.0%	578,152	20.2%
Requests for HTML files (mostly tables of contents and abstracts)	165,710	215,539	30.1%	260,699	21.0%
Requests for multimedia	1,714	4,468	160.7%	12,633	182.7%
Distinct files requested	6,419	21,451	234.2%	16,409	-23.5%
Distinct hosts served	29,816	57,901	94.2%	87,804	51.6%
Average data transferred daily	156 MB	219 MB	40.4%	382 MB	74.4%
Data transferred	55 GB	78 GB	40.4%	137 GB	75.6%

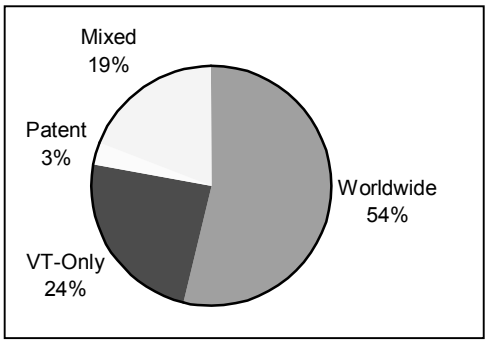
VT ETD Access by Int'l Sites

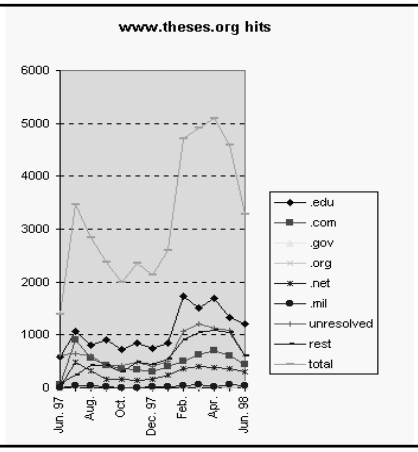
International Domain	1997/98	1997/98 rank	1998/99	1998/99 rank	Increase 1997/98 -1998/99	1999/00	1999/00 rank	Increase 1998/99 -1999/00
United Kingdom	6,735	1	11,347	1	68.5%	25,583	1	125.5%
Malaysia	876	16	4,190	6	378.3%	16,147	2	285.4%
France	2,138	7	4,797	5	124.4%	14,960	3	211.9%
Germany	6,727	2	3,374	9	-49.8%	14,384	4	326.3%
Canada	3,413	4	9,632	3	182.2%	13,543	5	40.6%
Spain	590	18	3,647	8	518.1%	9,918	6	171.9%
Italy	1,430	12	3,095	10	116.4%	9,300	7	200.5%

Multimedia Use in ETD Collection

File type	Examples	Count
Still image	**BMP, DXF, GIF, JPG, TIFF**	**328**
Video	**AVI, MOV, MPG, QT**	**58**
Audio	**AIFF, WAV**	**18**
Text	**PDF, HTML, TXT, DOC, XLS**	**7601**
Other	**Macromedia, SGML, XML**	**51**

Access Choices at VT (7/2000)





Who are sponsors / cooperators?

- **Funding, Donations of hardware/software**
  - **SURA**
  - **US Dept. of Education (FIPSE)**
  - **Adobe Systems**
  - **IBM**
  - **Microsoft**
  - **OCLC**
- **Others Serving on Steering Committee**
  - **National/Regional Projects: Australia, French speaking group, Germany, IberoAmerica (ISTEC), UK (UTOG)**
  - **CGS, National Lib. Canada, NSF, OAS, SOLINET, UMI, UNESCO, ...**

For professional societies

- Like "writing across the curriculum", e.g., Chemical Markup Language, MathML, …
- Besides writing: computing/communications, information literacy, personal digital library management, tool use, research methods, collaboration, archiving/preservation
- Data sets, communities of users of them
- Classification systems / browsing / searching
- NRC's "Issues for Science and Engineering Researchers in the Digital Age", 57 pages

Relationship with publishers

- **Concern** of faculty and students that still wish to publish books or journal articles, voiced: campus, Chronicle, NPR, Times
- **Solution**: Approval Form gives students, faculty choices on access, when to change access condition; use IPR controls in DL
- **Solution**: by case, work with publishers and publisher associations to increase access
  - **AAP, AAUP**
  - **AAAS, ACM, ACS, Elsevier, ...**

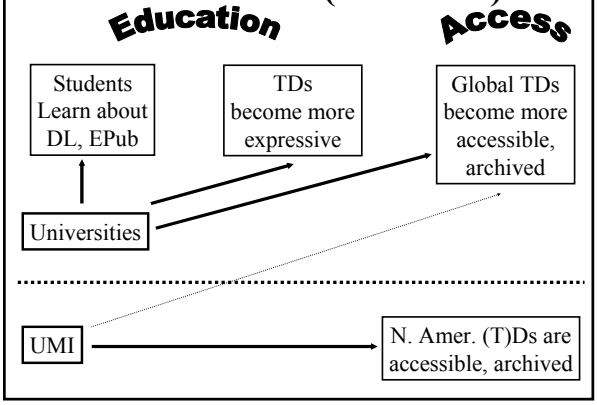
Some responses from publishers

- **ACM**: need to acknowledge copyright
- **Elsevier**: need to acknowledge copyright
- **IEEE-CS**: endorse initiative
- **ACS**: After first publication, can release
- **Textbook publishers**: different market, manuscript significantly reworked
- **General**: restricting access to local campus will not cause any problems

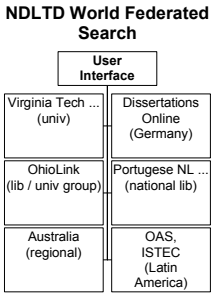
How does this relate to UMI?

- **Generally, they are independent decisions.**
- 1987 UMI workshop was first to explore ETDs.
- UMI wrote support letter for US Dept. of Ed. proposal.
- UMI is on Steering Committee.
- ProQuest Direct pilot of scanning works started 1/1/97, with free 2 yr access to front part.
- We are collaborating on:
  - **accepting electronic author submissions**
  - **standards (e.g., representation)**

ETD Initiative (and UMI)



User Search Support
(multilingual, XML)



Note: All groups shown are connected with NDLTD.

www.theses.org

- James Powell student project, D-Lib Magazine description in Sept. 1998
- XML description of each site
  - type of search engine / service
  - language
  - coverage (for resource discovery)
- Adding Z39.50 gateway capability and integrating with MARIAN, along with Harvest and Open Archives protocols

Access Approaches

- Goal: Maximize access and services, e.g., by encouraging:
- UMI centralized services
- VTLS: planned free union collection of metadata
- Distributed service: Dienst, Z39.50
- Regional services (e.g., OhioLinkh)
- Local servers with browse, search
  - From local catalogs to local archives
- WWW robot indexing and search services

Access Possibilities



Web search engines	www.theses.org	www.openarchives.org	library catalog clients	3rd Party Services (e.g., UMI)

Virginia Tech	MIT	National Library of Portugal	CBUC (Spain)	Ohio Link	National Projects: AU, GE, …

Why might a university want to be involved?

- To improve graduate education / better prepare your students / increase their knowledge and visibility
- To unlock university information
- To save money for students and for the university / improve workflow
- To build an important digital library

Multiple objectives

- **Sharing research results**
  - **Decrease costs, increase services**
  - **Increase knowledge of users**
- **Adding to author knowledge/skills**
  - **Epub, DL, IPR**
- **Enhancing organization's infrastructure**
  - **CS department, library**
  - **University, Laboratory**

DL Submission Software

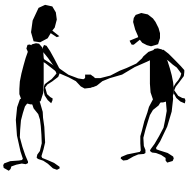
- **Similar software developed for W3C's WCA, CSTC, and NDLTD**
- **CSTC version field-tested to manage papers for ACM Digital Libraries '99**
- **May generalize for**
  - **conferences**
  - **electronic journal**
  - **resource description (e.g., courses, Web content)**

How can a university get involved?

- Select planning/implementation team
  - **Graduate School**
  - **Library**
  - **Computing / Information Technology**
  - **Institutional Research / Educ. Tech.**
- Send us letter, give us contact names
  - **www.ndltd.org/join**
- Adapt Virginia Tech solution
  - **Build interest and consensus**
  - **Start trial / allow optional submission**

Contact Our Project Team

E-mail
etd@ndltd.org

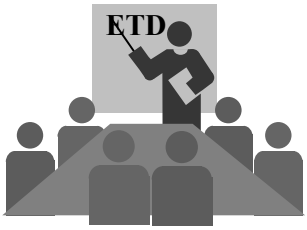


Video Tape

Visit

Phone Call

Convene Local Planning Group



Build Local ETD Site

Workshop/Training

Policies

Digital Library

Inspection/Approval

Support Offered

- Software, documentation, tech support
- Email, listservs (etd-l@listserv.vt.edu, -eval, -grad, -library, -technical)
- Donations: Adobe, Microsoft
- Evaluation: instruments, analysis http://scholar.lib.vt.edu - solutions/statistics
- (Temporary storage / archiving; aid - in setting up an int'l service & archive)

NUDL

- 1/15/99 NUDL proposal to NSF under DLI2 international program, later redone as separate bilateral projects
  - Partners: Germany, Mexico (Puebla and Monterrey), Brazil
  - Problems: Multilingual search, multimedia submissions, requirements/usability, …
- Start with ETDs, then expand to other student works, portfolios, data sets, (CS) courseware, ...

Future Work - 1 of 2

- Working with publishers to increase level of access as much as possible
- Interoperability tests among universities and with UMI to provide integrated services
- Study with testbed that emerges, to improve information retrieval, browsing, interface, and other types of user support
- Evaluation, improving learning experience, spread to worldwide initiative, sustainable support and coordination

Future Work - 2 of 2

- Adding services currently prototyped
  - annotation and SDI (routing) capabilities
  - Dublin Core metadata, crosswalk to MARC
- Adding other services planned
  - building and using citation database (w. SFX)
  - implementing plagiarism check (like "SCAM")
- Developing NUDL as a sustainable self governing global institution (w. committees)

Outline

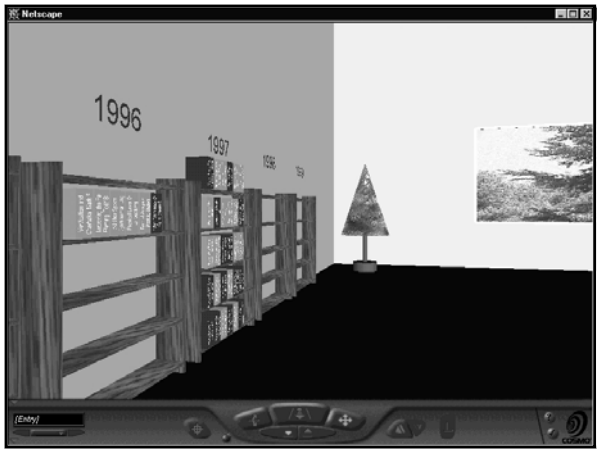
- **Virginia Tech context**
- **Why DLs? What are DLs? (5S theory)**
- **Case Study: WCA**
- **Case Study: Education: CSTC -> NSDL**
- **Case Study: NDLTD**
- **Accessibility and Visualization**
- **DL Software: MARIAN**
- **DL Hardware: PetaPlex**
- **Interoperability: OAI**

Accessibility Activities / Plans

- **Interface design (simple, 3D, VR)**
- **Usability studies**
- **Generic multi-lingual support**
- **Support for those with disabilities**
- **Hybrid collection (paper, MARC, abstracts, full-text, multimedia)**
- **Disciplinary classifications, tools**
- **Visualization of results, collection**

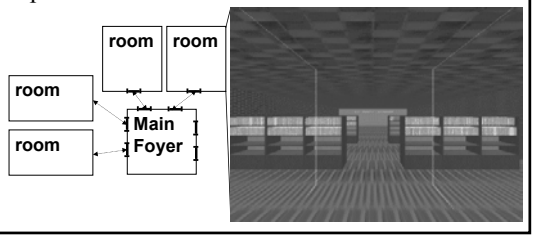
CAVE Experiments

- **Use a familiar metaphor**
  - **building / floor / room / shelf / book**
- **Rearrange orderings / shelving**
  - **use categories, clustering, ranking**
  - **use visualization: colors and gaps**
  - **study space mappings: physical, logical**
- **Simplify movement for key tasks**

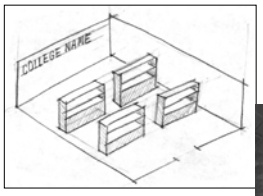


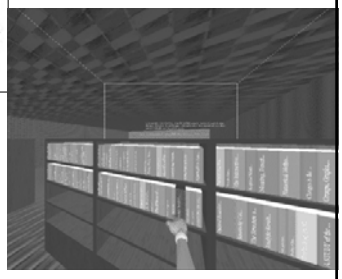
CAVE-ETD

- CAVE-ETD is a simulation of a library that runs in a CAVE (VR environment).
- Populated with a subset of ETD records.

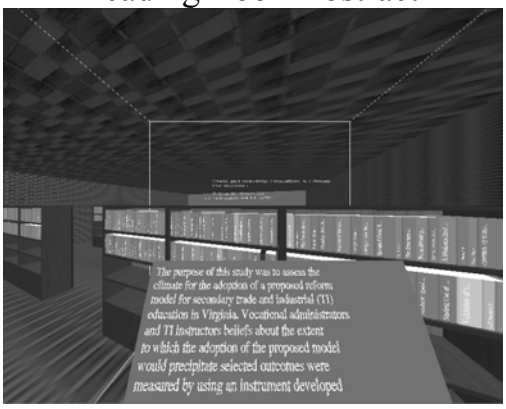


Book Browsing





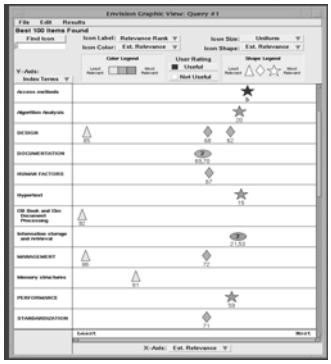
Reading Book Abstract

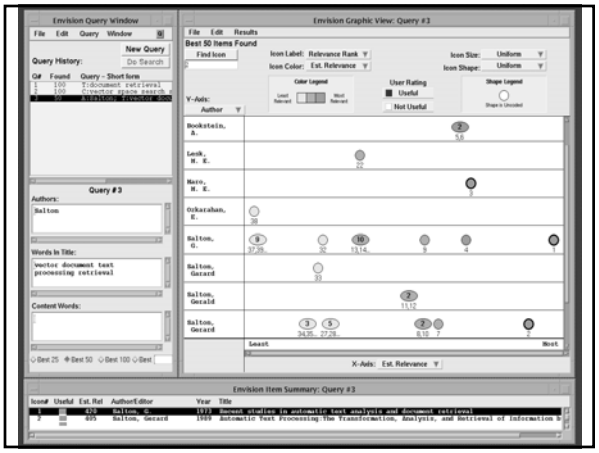


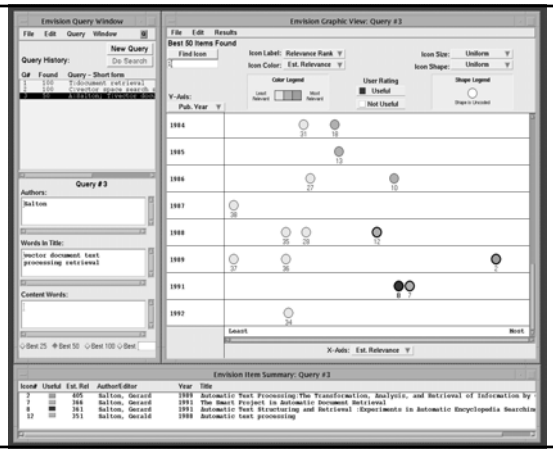
ENVISION

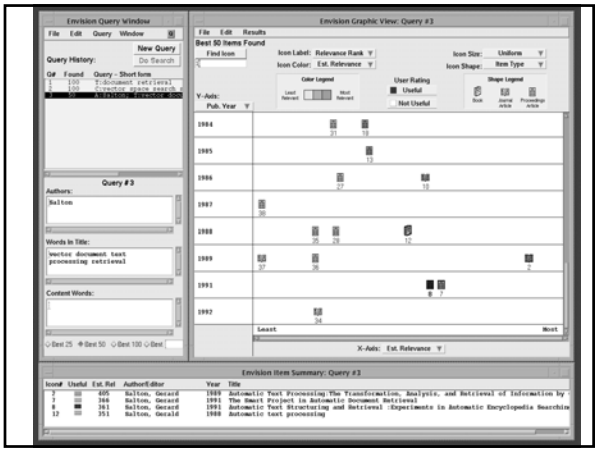
- NSF "A User- Centered Database from the Computer Science Literature" (1994 - 98)
- Collected bib/typesetter data, converted to SGML
- Scanned thousands of page images
- MARIAN search engine- can be made available (also applied to the Virginia Tech library catalog) used as part of a prototype object- based DL, with tailored visualization interface (L. Nowell dissertation)

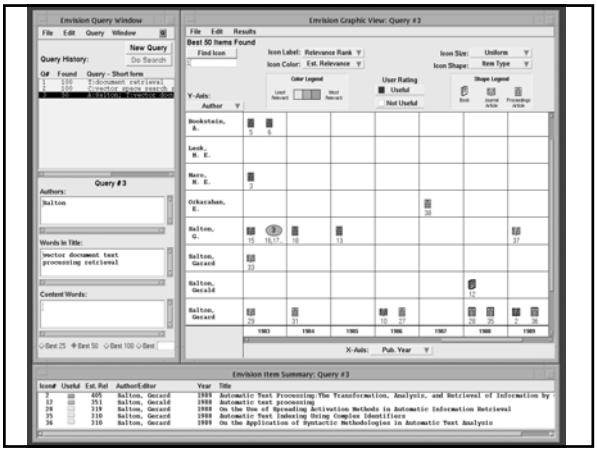
Envision Results Window



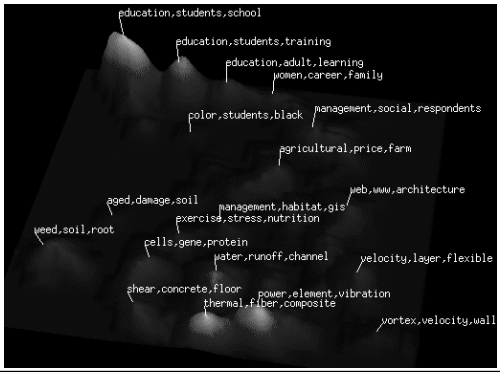








SPIRE Visualization



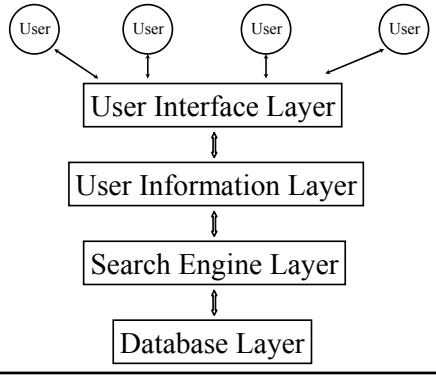
Outline

- **Virginia Tech context**
- **Why DLs? What are DLs? (5S theory)**
- **Case Study: WCA**
- **Case Study: Education: CSTC -> NSDL**
- **Case Study: NDLTD**
- **Accessibility and Visualization**
- **DL Software: MARIAN**
- **DL Hardware: PetaPlex**
- **Interoperability: OAI**

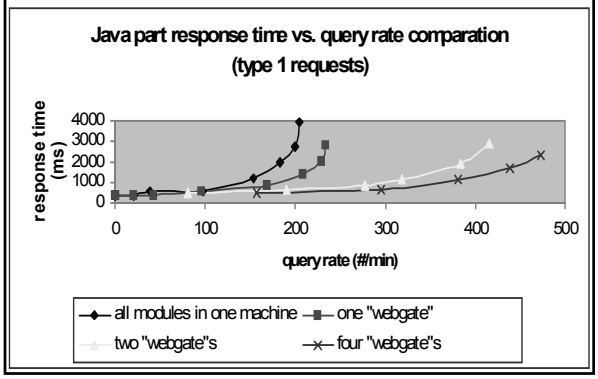
MARIAN

- **Multiple Access Retrieval of Information with Annotations**
- **(Marian the Librarian …)**
- **Evolved from CODER system to a distributed Online Public Access Catalog (OPAC), then DL backend, now becoming a full DL system**
- **From C/C++ to Java**
- **Future: NDLTD, NUDL, PetaPlex**
- **Use for campus collection management**
- **Use for www.theses.org as centralized system with gateway services: OAi, Harvest, Z39.50, …**

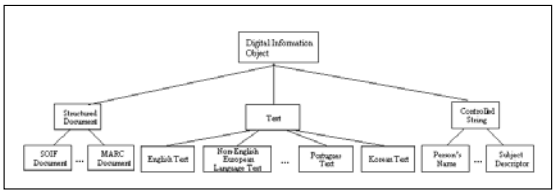
MARIAN Layers



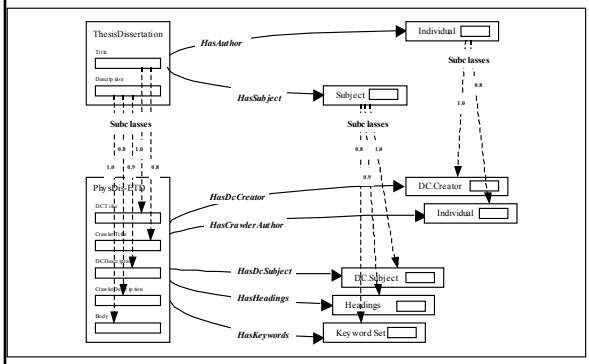
MARIAN Parallelism

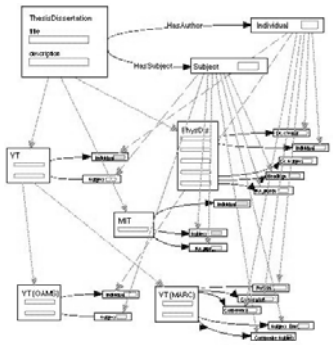


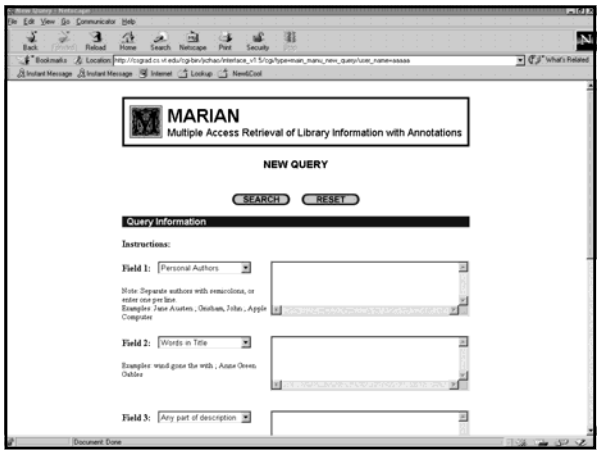
MARIAN – Part of Class Hierarchy

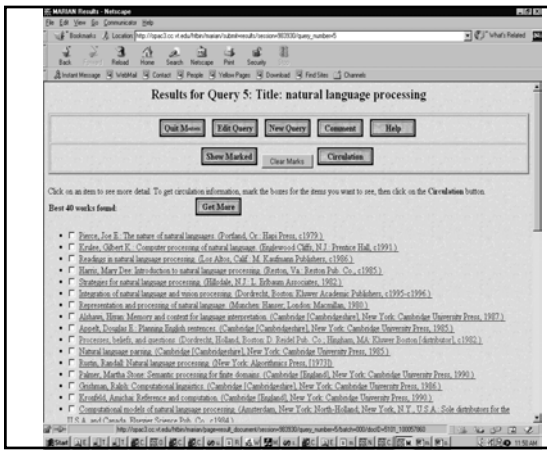


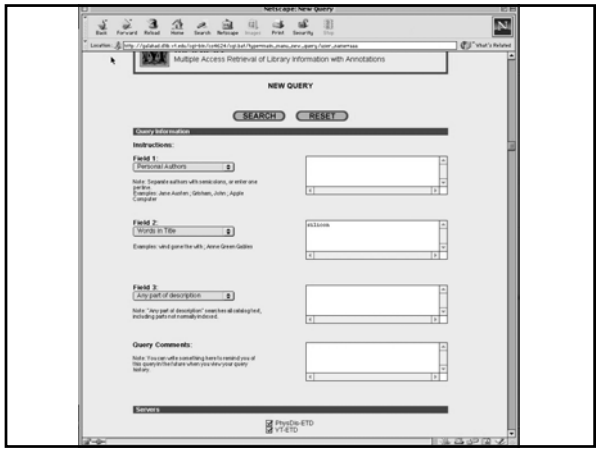
PhysDis Collection View

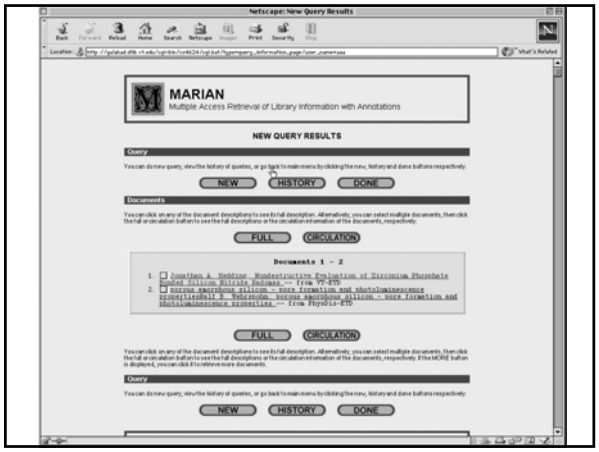


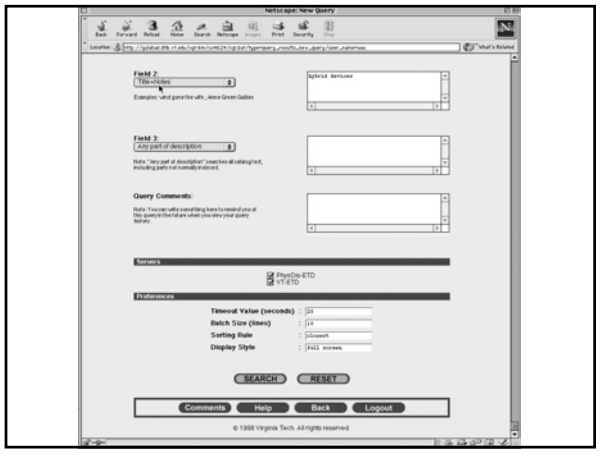










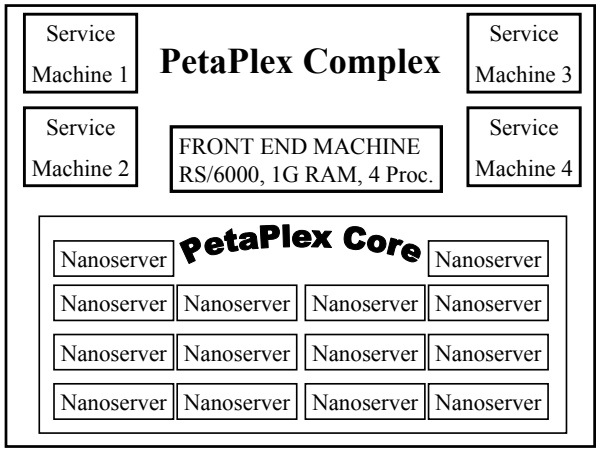


Outline

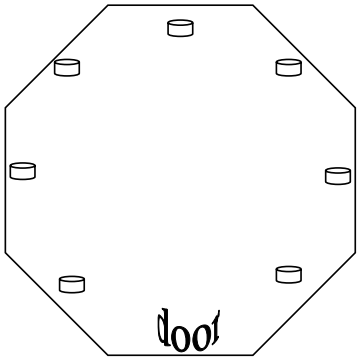
- **Virginia Tech context**
- **Why DLs? What are DLs? (5S theory)**
- **Case Study: WCA**
- **Case Study: Education: CSTC -> NSDL**
- **Case Study: NDLTD**
- **Accessibility and Visualization**
- **DL Software: MARIAN**
- **DL Hardware: PetaPlex**
- **Interoperability: OAI**

PetaPlex

- Digital Library Machine ("super" object store): Parallel computer / storage utility
- Research: inverted files, video server, …
- Knowledge Systems Incorporated is supplying VT-PetaPlex-1 with 2.5 terabytes through 100 nodes:
  - Net connection + 25GB disk + 233 MHz Pentium + Linux

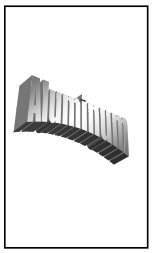


PetaPlex Top View



PetaPlex Side View

Roles:
* Support
* Cooling
* Power



15 shelves

8 ft. high

4 ft. wide

PetaPlex Service Machine Possibilities

- Front-end provides handle/repository abstraction through hashing
- Small object server
- Large object server
  - video on demand
  - streaming audio
- Information retrieval server
- Proxy / cache server (e.g., 1 terabyte server of 1000 worldwide for Comsat/Intelsat)

Sornil & Mather Dissertations

- Mather: efficiently handling very large numbers of objects of varying sizes
- Sornil: efficiently handling IR for very large dynamic collections, large numbers of users, high transaction rates, large inverted files
  - modeling and simulation
  - data organization
  - parallelization of algorithms, alone and in combination for retrieval (related) tasks

Comparison

	Network of Workstations (NOW)	Beowulf	PetaPlex
Archi-tecture	Cluster of general purpose workstation class machines using off-the-shelf network interconnect	General purpose PCs, interconnected with a custo-mized network	Special purpose architecture tuned for superstorage. Uses a mix of off-the-shelf PC compo-nents and specialized network interconnects.
Cost per node	Workstation prices. Between $2000-$2500/node	Mid to low-end PC prices. Between $1200-$1800 per node	Mass produced components will reduce price to around $100/node
Target area	Computation	Computation	Storage, computation is a secondary function
Filesystem support	UNIX flavors	UNIX flavors	Replaces location dependant files with location independent fine-grained URN named objects

Outline

- **Virginia Tech context**
- **Why DLs? What are DLs? (5S theory)**
- **Case Study: WCA**
- **Case Study: Education: CSTC -> NSDL**
- **Case Study: NDLTD**
- **Accessibility and Visualization**
- **DL Software: MARIAN**
- **DL Hardware: PetaPlex**
- **Interoperability: OAI**

Open Archives Initiative

OAI

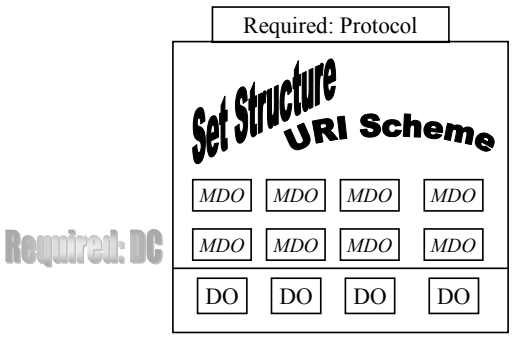
www.openarchives.org

openarchives@openarchives.org

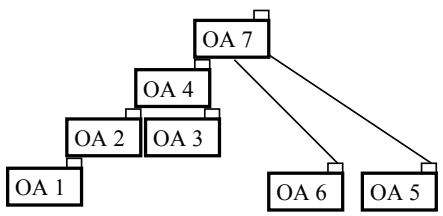
Open Archives Initiative (OAI)

- xxx@LANL, high-energy physics (Ginsparg, 1991)
- CSTR + WATERS = NCSTRL (Lagoze,1994)
- xxx + NCSTRL = CoRR collaboration (1998)
- Universal Preprint Service protoproto, Oct. 21-22, 1999, Santa Fe – led by LANL, CNI, DLF, Mellon -->OAi
- Santa Fe Convention (see Feb. D-Lib Magazine article)
- Follow-on mtgs: 6/3@San Antonio, 9/21@Lisbon (ECDL)
- Archives->Open Archives
  - Support unique archive identifiers
  - Implement Open Archives metadata set (DC, using XML)
  - Implement OA harvesting protocol (derived from Dienst protocol)
  - Register the archive
- Build tools, layer other services: linking, searching, …

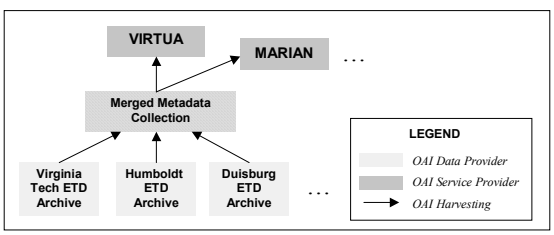
OAI – Repository Perspective



OAI – Black Box Perspective



ETD Union Collection (OAI)



Tiered Model of Interoperability

Mediator services

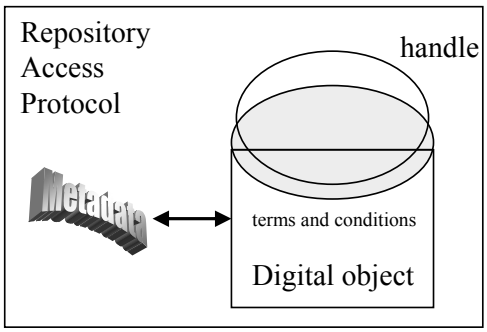
Metadata harvesting

Document models

OAi Philosophy

- Self-archiving = submission mechanism
- Long-term storage system = archive
- Open interface = harvesting mechanism
- Data provider + service provider
- Start with "gray literature"
  - e-prints/pre-prints, reports, dissertations, …

Repository of Digital Objects



Open Archives (protoproto)

- **ArXiv** & Los Alamos National Lab
- **CogPrints** & U. Southampton
- **NACA** & NASA (reports)
- **NCSTRL** & Cornell U.
- **NDLTD** & Virginia Tech
- **RePEc** & U. Surrey
- Total of around 200K records

Original Open Archives Members

- **American Physical Society**
- **California Digital Library**
- **Caltech**
- **Coalition for Networked Info.**
- **Cornell University**
- **Harvard University**
- **Library of Congress**
- **Los Alamos Nat'l Lab**
- **Mellon Foundation**
- **NASA Langley Research Cntr**
- **Old Dominion University**
- **Stanford University**
- **U. of Ghent**
- **U. of Surrey**
- **U. of Southampton**
- **Vanderbilt University**
- **Virginia Tech**
- **Washington University**

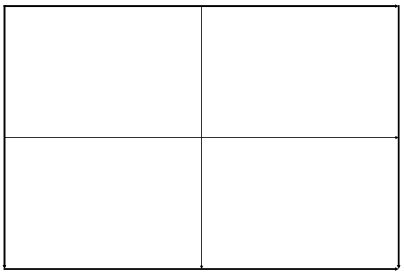
Open Archives Future

- EconWPA (U. Washington)
- e-biomed -> PubMed Central (NIH)
- PubScience (DOE)
- Clinical Medicine Netprints (+ other HighWire Press holdings)
- University ePub (California Digital Library)
- All public e-prints (MIT)
- Scholar's Forum (Caltech)
- Int'l: CERN, Germany, India, Mexico, …
- **Goal: millions of books/articles/reports / yr**

Approaches to Open Archives

Build By Institution

Build By Discipline



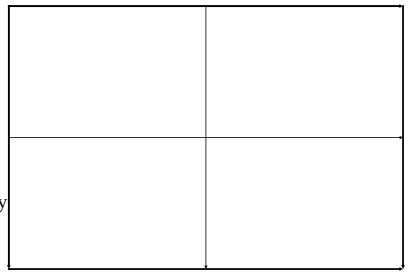
Approaches to Open Archives

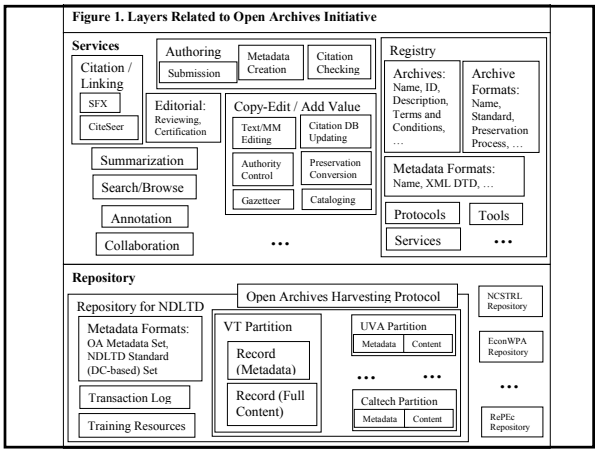
Build By Institution

Build By Discipline

Access by

Author
Category
Interdisciplinary
Year
Language
Query …





Mechanisms

- **Sharing**
  - **Join federation, run software**
  - **Make metadata and archive available**
- **Aggregating**
  - **By discipline**
  - **By institution**
  - **By genre**
- **Automating**
  - **Workflow**
  - **Harvesting and providing services**
  - **Federated searching**
  - **Dynamic linking (e.g., with SFX (OpenURLs))**

VT View of the Open Archives Initiative (OAI)

- Enable sharing of publication metadata and full-text by digital libraries
- Standardize low-level mechanisms to share contents of libraries
- Build higher-level user-centric and administrative services in meta-libraries
- Install organizational mechanisms to support the technical processes

Virginia Tech Projects

- MARC XML-DTD
- Computer Science Teaching Centre (CSTC)
- W3C Web Characterization Repository
- OAI Repository Explorer
- Networked Digital Library of Theses and Dissertations (NDLTD)

MARC XML-DTD

- XML Transport format for US-MARC records
- Standardized metadata exchange format for traditional library services joining OAI

OAI Repository Explorer

- Serves as a compliancy test
- Allows browsing of open archives using only OAI protocol
- Sends requests on behalf of user, parses and checks responses and displays browsable interface
- Will detect most discrepancies in protocol
- http://purl.org/net/explorer

Request, Response – OAI, VT ETDs

Request

http://scholar.lib.vt.edu/theses/OAI/cgi-bin/index.pl?
verb=GetRecord&metadataPrefix=oai_etdms&identifier=oai:VTETD:etd-52011285965179 1

Response

<?xml version="1.0" encoding="UTF-8" ?>
- <GetRecord xmlns="http://www.openarchives.org/OAI/1.1/OAI_GetRecord"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/1.1/OAI_GetRecord
http://www.openarchives.org/OAI/1.1/OAI_GetRecord.xsd">
<responseDate>2001-08-18T18:03:28-05:00</responseDate>
<requestURL>http://scholar.lib.vt.edu:80/theses/OAI/cgi-bin/index.pl?
verb=GetRecord&metadataPrefix=oai_etdms&identifier=oai:VTETD:etd-
52011285965179 1</requestURL>
- <record>
- <header>
<identifier>oai:VTETD:etd-52011285965179 1</identifier>
<datestamp>1996-06-05</datestamp>
</header>
- <metadata>
- <thesis xmlns="http://www.ndltd.org/standards/metadata/etdms/1.0/"
xsi:schemaLocation="http://www.ndltd.org/standards/metadata/etdms/1.0/
http://www.ndltd.org/standards/metadata/etdms/1.0/etdms.xsd">
<title>Analysis of Tow-Placed, Variable-Stiffness Laminates</title>
<creator>Waldhart, Chris</creator>
<subject>variable-stiffness laminates</subject>
<subject>curvilinear fibers</subject>
<subject>tow placement machine</subject>
<subject>buckling</subject>
<description>It is possible to create laminae that have spatially varying fiber
orientation with a tow placement machine. A laminate which is composed of
such plies will have stiffness properties which vary as a function of position.
Previous work had modelled such variable-stiffness laminae by taking a

Summary

- **Virginia Tech context**
- **Why DLs? What are DLs? (5S theory)**
- **Case Study: WCA**
- **Case Study: Education: CSTC -> NSDL**
- **Case Study: NDLTD**
- **Accessibility and Visualization**
- **DL Software: MARIAN**
- **DL Hardware: PetaPlex**
- **Interoperability: OAI**

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

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.

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 a by a defined community or set of communities."
  (Digital Library Federation)
- "Digital libraries are complex data/information/knowlege (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 hat 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.

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.

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.

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.

Resources:

- Projects
- People
- Countries and regions
- Centers, sites and organizations

[Main] [Contents]

Please send comments/suggestions to Ed Fox.

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 Inititives

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://wwwqbic.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.

DIGITAL LIBRARIES INITIATIVE

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

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 - Informedia
- Informedia-II (for DLI-2)
- IEEE Computer article
- NetBill

[Main] [Contents] [Resources] [Projects]

Please send comments/suggestions to Ed Fox.

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.

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.

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 mutilingual IR demos")

[Main] [Contents] [Resources] [Projects]

Please send comments/suggestions to Ed Fox.

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.

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

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.

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.

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.

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 Mellom 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 deifferent 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 understading. 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 fro digital library projects. The program originally called FIGIT, now known as E-LIB funded 35 projects. Work includes ctaloguinf 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 CAntersbury Tales project that involves scanning all pre-1500 manuscripts and some ohe 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.

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.

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.

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.

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.

ETD Digital Library

Networked Digital Library of Theses and Dissertations: Federated Search

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

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

About Emerge
Distributed Search
Emerge Architecture
Standards
Emerge Components
Gazebo
Gazelle
CTK
Grunk
Download
Documentation
Gazebo
Gazelle
CTK
Grunk
Links

NCSA
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.

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.

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 if the internet and other networked information, research in automatic filtering of networked information has exploded in recent years. Becuase 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 refered 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

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 - 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*TF/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: Cand * MIN + (1-Cand) * MAX
    - OR: Cor * MAX + (1-Cor) * 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) * di
    - 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=infinity 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=infinity 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 - 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): MMM < Paice < P-norm
- Effectiveness (i.e., precision at given recall level) seems to be: MMM < Paice < 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:

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.

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.

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.

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.

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.

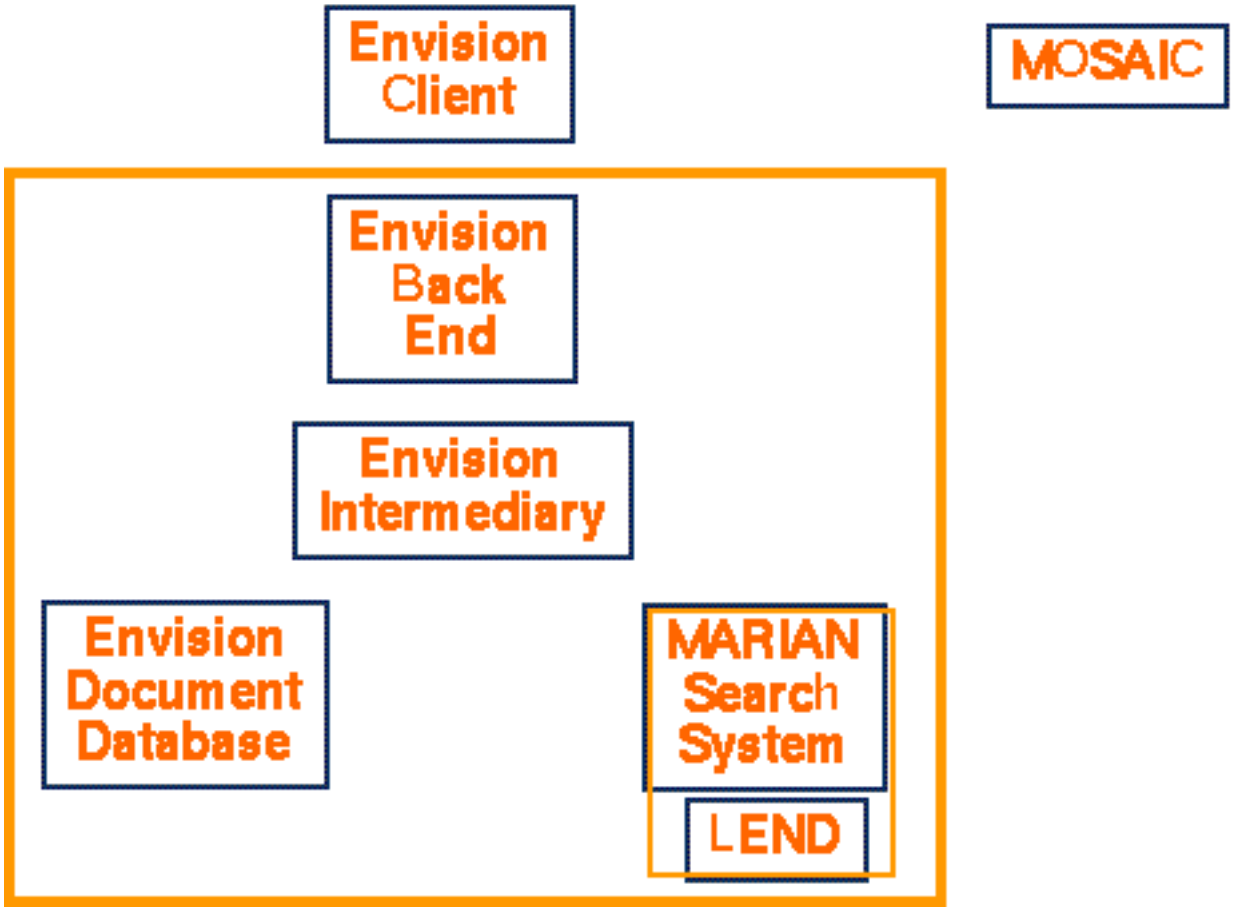
Efforts have concentrated on building an archive based upon SGML, developing an object-oriented database, applying the MARIAN retrieval system and WWW, and constructing a special search interface based upon user wishes.

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:

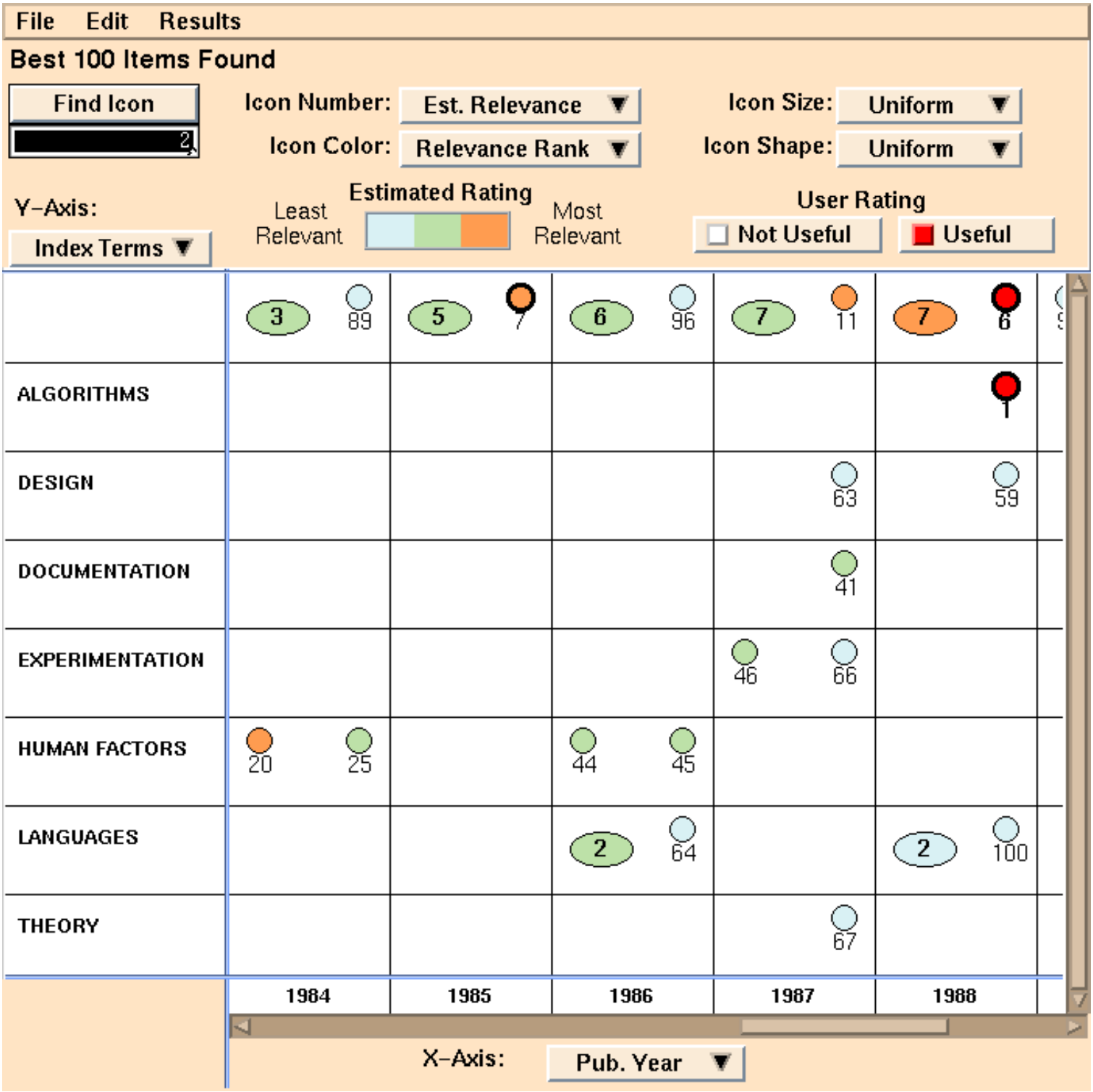
Envision


Envision Client
MOSAIC
Envision Back End
Envision Intermediary
Envision Document Database
MARIAN Search System
LEND

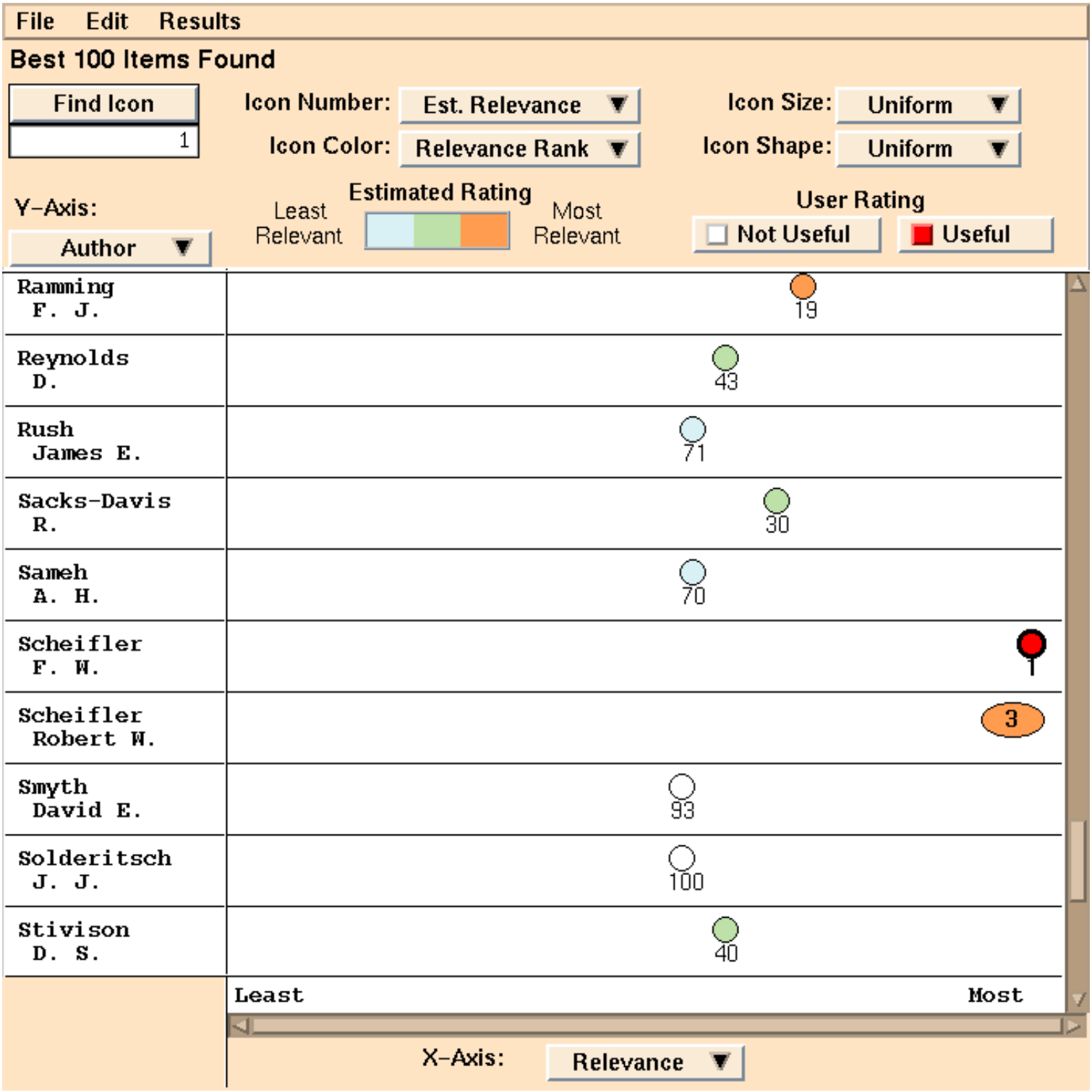
Envision - Results Screens

The interface includes:

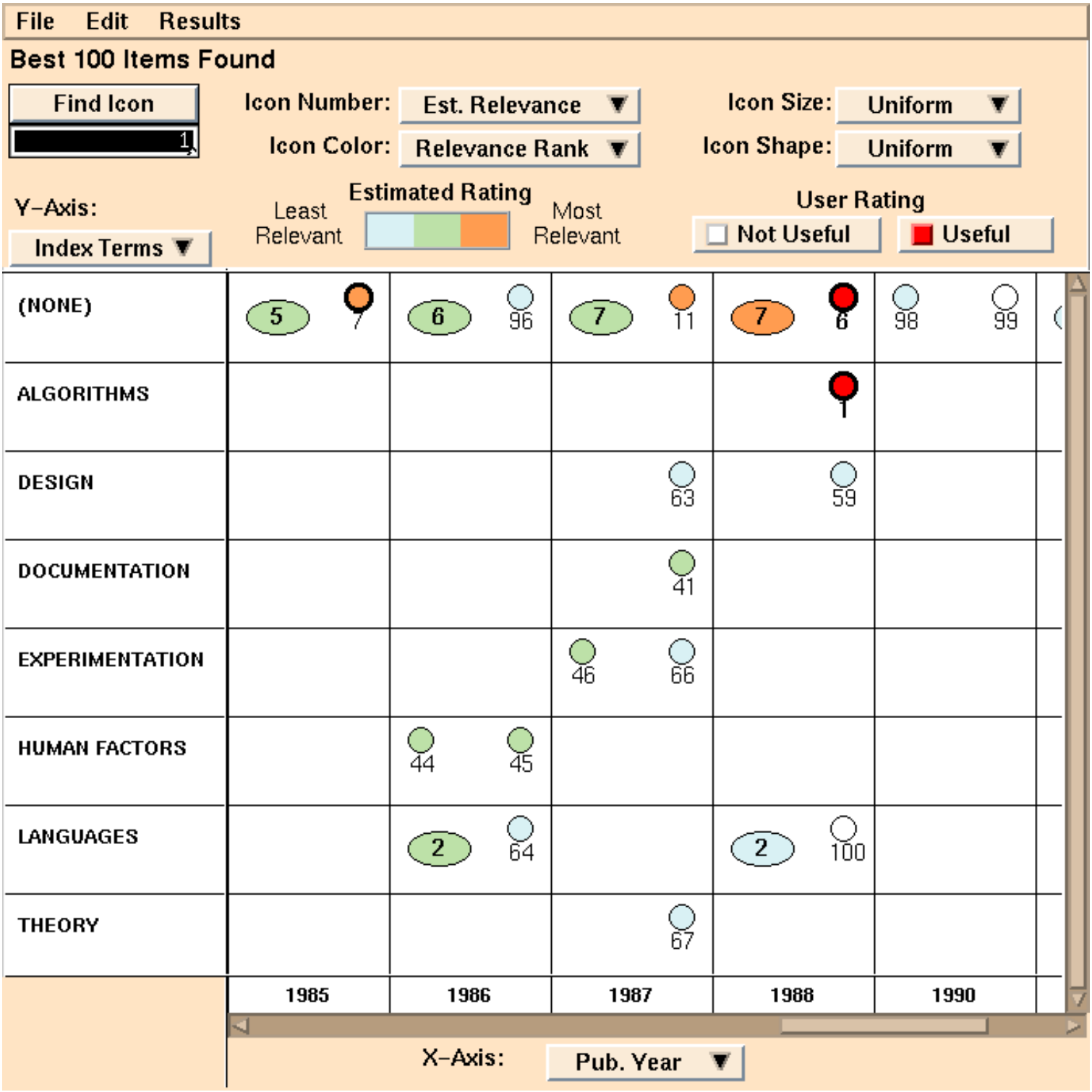
- Graphic View 1:



- Graphic View 2:



- Graphic View 3:


File Edit Results
Best 100 Items Found
Find Icon
1
Icon Number: Est. Relevance
Icon Color: Relevance Rank
Icon Size: Uniform
Icon Shape: Uniform
Y-Axis:
Index Terms
Estimated Rating
Least Relevant
Most Relevant
User Rating
Not Useful
Useful
(NONE)
ALGORITHMS
DESIGN
DOCUMENTATION
EXPERIMENTATION
HUMAN FACTORS
LANGUAGES
THEORY
1985
1986
1987
1988
1990
X-Axis:
Pub. Year

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.

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 their 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.

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 occured 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.

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 advantagesand 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.

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.

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 Newsline 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.