Theses and Dissertations in the Digital Library http://www.ndltd.org Croatia - May 1999

Edward A. Fox
Virginia Tech, Blacksburg, VA, USA
fox@vt.edu

NDLTD Project Team

PIs:
Ed Fox - Computer Science (fox@vt.edu)
John Eaton - Graduate School (eaton@vt.edu)
Gail McMillan - Library (gailmac@vt.edu)

GRAs:
Constantinos Phanouriou - Computer Science
Paul Mather - Computer Science

Virginia Tech Background

- Largest university in Virginia, land-grant, town population 35K plus 25K students
- Blacksburg Electronic Village, since 1992, with 80% of community on Internet
- Net Work Virginia, largest ATM network, with over 600 sites, for education, research, govt
- LMDS, Local Multipoint Distribution Service, gigabit wireless networking - 1/3 of Virginia
- Math Emporium, 500 workstations
- Faculty Development Initiative, round 2

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

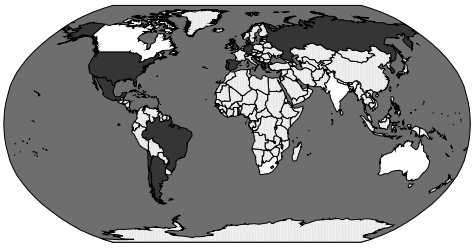
DLs: Why of Global Interest?

- **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**

NUDL

- 1/15/99 NUDL proposal to NSF under DLI2 international program
  - VT: Fox, Kleiner, McMillan, Eaton
  - Partners: UK (2) , Singapore, Russia, Korea, Greece, Germany, plus Iberoamerican group (Spain, Portugal, Argentina, Brazil, Chile, Mexico)
  - Problems: Multilingual search, multimedia submissions, requirements/usability, …
- Start with ETDs, then expand to other student works, portfolios, data sets, (CS) courseware, ...

National Coverage (red/white)



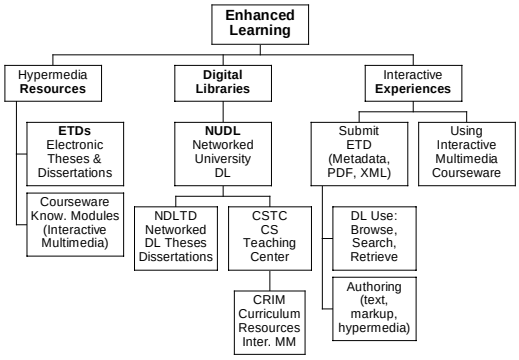
Why of Interest in Computing?

- Next step in fields of DBMS, HT, IR, MM
- **Efficiency** requires advances in, e.g.,
  - algorithms and data structures (ex., MPHF)
  - networking (ex., HTTP-NG)
  - OS (ex., support for streams)
- **Effectiveness** requires advances in, e.g.,
  - AI (ex., multilingual texts, user adaptation)
  - HCI (ex., visualization, DLs embedded in activities)
- **CS Educ.** can benefit; CS can aid **Dist. Educ.**

SMETE Library
(from www.dlib.org)

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

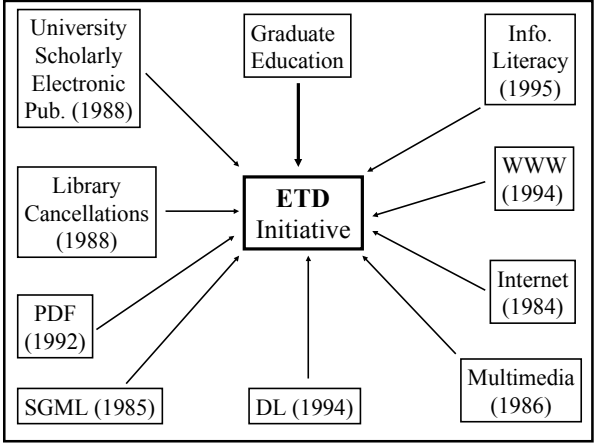
Solution Decomposition



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

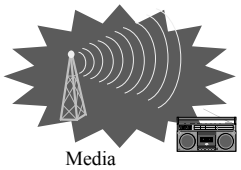


ETDs Got Your Interest?

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:

- Networked infrastructure
- Scalability
  - 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 (MEL): SURA, SOLINET
- 1994 mtg in Blacksburg re ETD project: std of PDF + SGML + multimedia objects
- 1996 funding by SURA, US Dept. of Education (FIPSE) for regional, national projects
- 1997 meetings in UK, Germany, ...

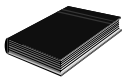
What are we doing?

- Aiding universities to enhance grad educ., 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)

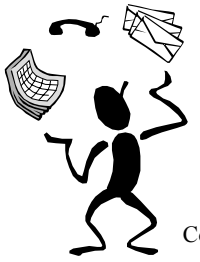
What are the long term goals?

- 400K US students / year getting grad degrees are exposed / involved
- 200K/yr rich hypermedia ETDs that may turn into electronic portfolios
- Dramatic increase in knowledge sharing: lit. reviews, bibliographies, …
- Services providing lifelong access for students: browse, search, prior searches, citation links

Student Prepares Thesis or Dissertation



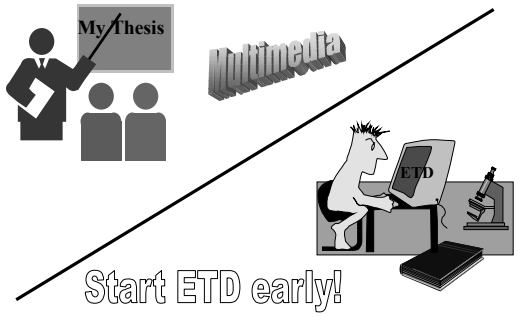
Literature



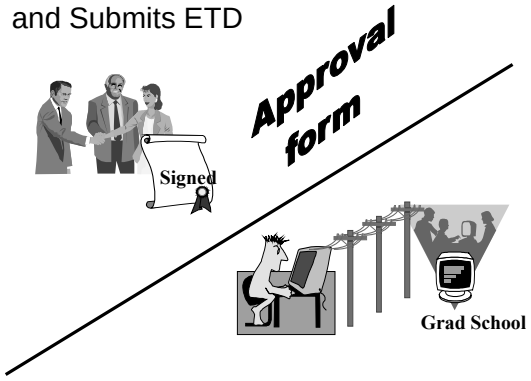
Computer Resources

Research

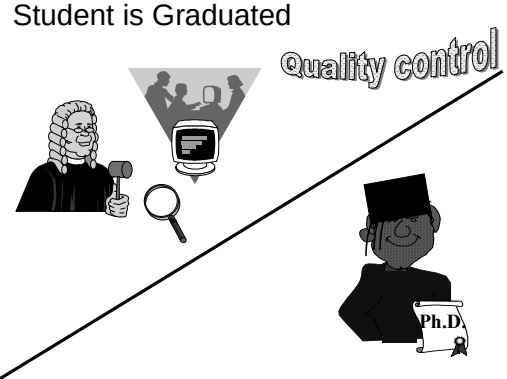
Student Defends and Finalizes ETD



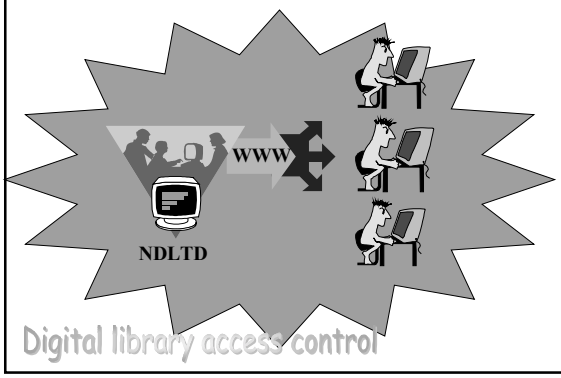
Student Gets Committee Signatures and Submits ETD



Graduate School Approves ETD Student is Graduated



Library Catalogs ETD and New Students Have Access 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 2000 ETDs in collection

Institutional Members

- **Coalition for Networked Information (CNI)**
- **Committee on Inst. Coop. (CIC)**
- **Diplomica.com**
- **Dissertation.com**
- **National Library of Portugal**
- **UNESCO**

US University Members

- Air University (Alabama)
- Cal Tech
- Clemson University
- College of William & Mary
- Concordia University (Illinois)
- East Tenn. State University
- Florida Institute of Tech.
- Florida International University
- Michigan Tech
- Naval Postgraduate School (CA)
- North Carolina State U.
- Penn. State University
- Rochester Institute of Tech.
- U. of Florida
- U. of Georgia
- University of Hawaii, Manoa
- U. of Iowa
- U. of Maine
- U. of Oklahoma
- U. of South Florida
- U. of Tennessee, Knoxville
- U. of Tennessee, Memphis
- U. of Texas at Austin
- U. of Virginia
- U. Wisconsin - Madison
- Vanderbilt U.
- Virginia Tech - required since 1/97
- West Virginia U. - required beginning fall 1998
- Worcester Polytechnic Inst.

Australian Project Members

- U. New South Wales (lead institution)
- U. of Melbourne
- U. of Queensland
- U. of Sydney
- Australian National University
- Curtin U. of Technology
- Griffith U.

German Project Members

- Humboldt University (lead institution)
- 3 other universities
- 5 learned societies
- 1 computing center
- 2 major libraries

Other International Members

- Chinese University of Hong Kong
- Chungnam National U., Dept of CS (S. Korea)
- City University, London (UK)
- Darmstadt U. of Tech. (Germany)
- Gyeongsang National U. (Korea)
- India Institute of Technology, Bombay (India)
- Nanyang Technological U. (Singapore, part)
- National U. of Singapore (Singapore, part)
- *National Library of Portugal
- Polytechnic University of Valencia (Spain)
- Rhodes U. (South Africa)
- St. Petersburg St. Tech.U (Russia)
- Univ. de las Américas Puebla (Mexico)
- U. Laval; U. of Guelph; U. Waterloo; Wilfrid Laurier U. (Canada)

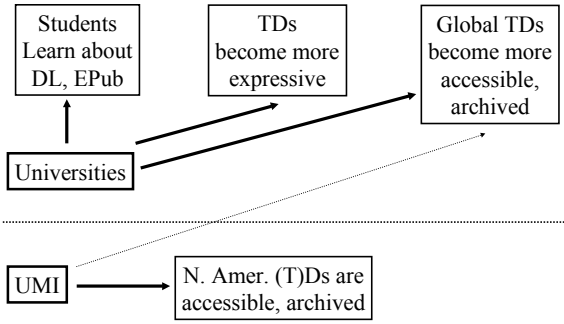
Who are sponsors / cooperators?

- **Funding, Donations of hardware/software**
  - **SURA**
  - **US Dept. of Education (FIPSE)**
  - **Adobe Systems**
  - **IBM**
  - **Microsoft**
  - **OCLC**
- **Others Serving on Steering Committee**
  - **CGS, National Lib. Canada, SOLINET, UMI, UNESCO, UTOG (UK), ...**

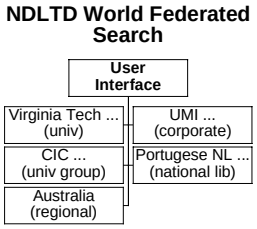
How does this relate to UMI?

- 1987 UMI workshop to explore ETDs
- Support letter for US Dept. of Ed. proposal
- Steering and technical committee membership
- ProQuest Direct pilot of scanning works started 1/1/97
- Collaborating on:
  - **accepting electronic author submissions**
  - **standards (e.g., representation), research**

ETD Initiative (and UMI)



User Search Support (multilingual, XML)



Note: All groups shown are connected with NDLTD.

Interoperability Tests Planned

- IBM DL: donated equipment, technical support, powerful IPR (see TOIS, D-Lib)
- Z39.50: OCLC SiteSearch / VT tailored s/w
  - **university libraries w. catalogs of freely shared MARC records pointing to archival copies**
  - **via URNs: handles & PURLs**
- Dienst / NCSTRL - www.ncstrl.org: CS depts., DARPA, NSF, CNRI, Cornell - UVA is working on extensions for ETDs - Portugal is studying use for Europe - VT is working on Dienst to Z39.50 gateway

Access Approaches

- Goal: Maximize access and services, e.g., by encouraging:
- UMI centralized services
- Distributed service: Dienst, Z39.50
- Regional services (e.g., CIC, MEL)
- Local servers with browse, search
  - From local catalogs to local archives
- WWW robot indexing and search services

Why might your university want to be involved?

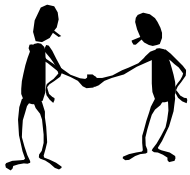
- To improve graduate education / better prepare your students
- To unlock university information
- To save money for students and for the university / improve workflow
- To build an important digital library supported by SURA, FIPSE

How can your university get involved?

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

Contact Our Project Team

E-mail
etd@ndltd.org



Video Tape

Phone Call

Convene Local Planning Group

Build Local ETD Site

Workshop/Training

Policies

Digital Library

Inspection/Approval

Type 1 Members

University Requires ETDs

- Adobe Acrobat and/or 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.)

Support Services Developed

- WWW site with > 300 Mb, CD, videotape
- Automated submission system (UNIX, msql, WWW scripts for grad school & library)
- Student guidelines, style sheets, multimedia training materials, FAQs, press info
- SGML and XML DTDs for ETDs
- SGML to HTML (web generator)
- LaTeX, Word templates, converters

Support Offered

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

PetaPlex

- Parallel computer / storage utility for scale of 1000 to 1,000,000 gigabytes (terabyte/petabyte)
- Knowledge Systems Incorporated will supply VT-PetaPlex-1 with 2.5 terabytes, > 100 processors, high speed backbone connection
- High-performance "superstore" available to help with NDLTD

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

Future Work

- 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

Accessibility

- **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**

Education

- Retain ABDs
- Improve writing (including 2nd language)
- Improve organization
- Improve presentation's technical quality
- Support above in libraries, media centers
- Support above "anytime, anywhere"
- Support faculty as well as students

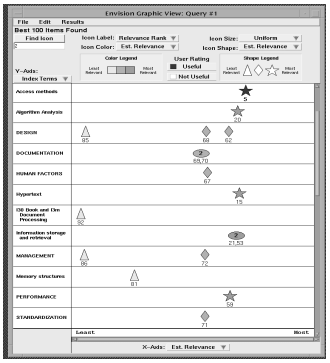
Professional Societies

- Like "writing across the curriculum"
- 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 "On becoming a researcher"

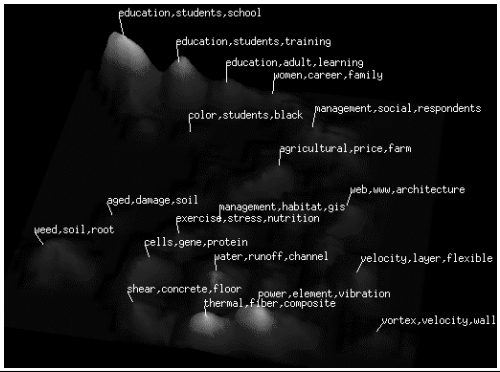
Research Goals: Testbed

- **Heterogeneous:**
- **Language, terminology**
- **Writing style, "structure"**
- **Subjects / topics**
- **Communities of use, "scenarios"/services**
- **Types of related publications and co-uses**
- **Cataloging, classification "spaces" - info. viz.**
- **Personalization**
- **Multimedia "streams"**
- **Variation in size: MARC record, title + abstract, full-text, "container object"**
- **"Societies": dept., college, univ., state,, sports group, nation, language group, ...**

Envision Results Window



SPIRES Visualization



How to Build a Digital Library

Workshop and Training Materials

Neill A. Kipp

May 19, 1999

How to Build a Digital Library

- Understand the problem
- Try to solve it
- Evaluate results
- Iterate

Understand the Problem

Digital Libraries are complex systems that:

1. help satisfy information needs of users	*societies*
2. provide information services	*scenarios*
3. present information in usable ways	*spaces*
4. organize information in usable ways	*structures*
5. communicate information to users	*streams*

5S Framework -- Definitions

Societies
groups that interact

Scenarios
services, functions,
operations, methodologies

Spaces
domains + constraints
(e.g., distance, adjacency)

Structures
nodes and arcs

Streams
sequences of items

5S Framework -- Components

Societies	Scenarios	Spaces	Structures	Streams
Roles	Acquire	Physical	Architectures	Granularities
Rituals	Index	Functional	Taxonomies	Protocols
Reasons	Administer	Presentational	Grammars	Paths
Artifacts	Consult	Temporal	Links	Flows
Relationships	Preserve	Conceptual	Objects	Turbulences

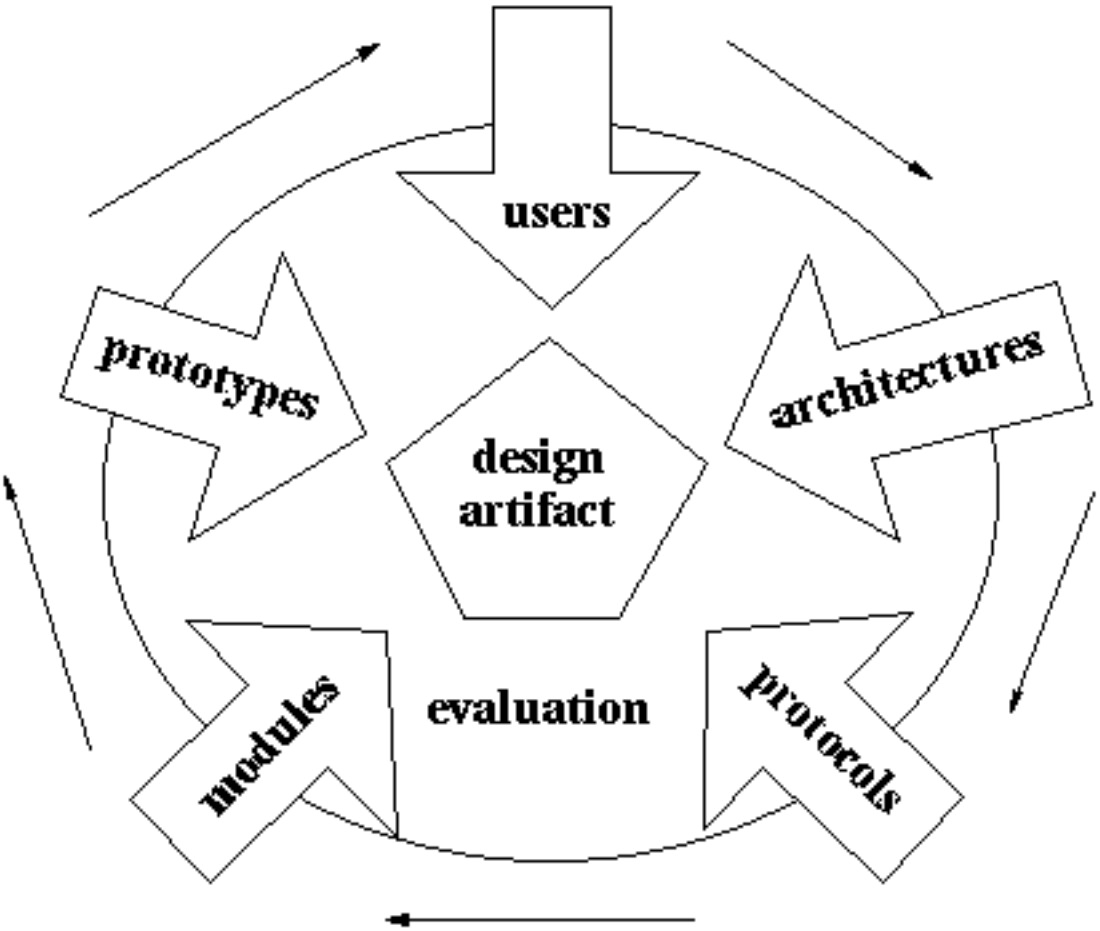
It is Not Enough to Understand the Problem

Hardest problem facing digital library designers:
"What to do next?"

5S Framework and Star Methodology

Framework		Methodology
Classify	-	Evaluate
Analyze	-	Write
Divide	-	Conquer
Understand	-	Build
Think	-	Do

Star Methodology



First Design Meeting

1. Consider societal issues
    - user base
    - funding resources
    - system requirements
2. Determine basic architecture
3. Determine how components communicate
4. Choose shrinkwrap/shareware modules
5. Develop quick prototypes
6. ... evaluate, Evaluate, EVALUATE!
7. Record results

Design Artifact

Contains...
- User requirements
- Evaluation plans
- Figures
- Screen shots
- Reference manuals
- Prototypes

Represented as...
- Hyperdocuments
- Graphics
- Software programs

Obtained by consulting...
- Users
- Architectures
- Protocols
- Modules
- Prototypes

Design Artifact based on 5S Framework (1 of 3)

Societies
- Objectives/goals
- User requirements
- User/reference manuals
- Usability plans/results

Scenarios
- Use cases
- Services
- Functionality

Spaces
- Diagrams
- Screen shots

Design Artifact based on 5S Framework (2 of 3)

Structures
- System requirements
- System architecture
- Field-specific terminology
- Languages/grammars

Streams
- Protocols
- Activity logging
- Timing/synchronization
- Network access
- Chaos control

Also in Combinations (3 of 3)

Societies + Spaces
- User interface look and feel

Societies + Scenarios
- Evaluation plans

Scenarios + Structures
- Object decomposition
- Module choices

Spaces + Structures
- Taxonomies

Structures + Streams
- Documents
- Hypertext

Spaces + Structures + Streams
- Multimedia support

Star Methodology: Users

1. Create glossary of field-specific terminology
2. Collect requirements, tasks, scenarios, use cases
3. Involve users in participatory design
4. Plan usability evaluation of system
5. Collect usability data of interactions
6. Record results in design artifact

Star Methodology: Architectures

1. Separate design into logical, manageable components
2. Determine objects and interconnections
3. Draw structural diagrams
4. Record results

(e.g., Stanford Infobus, IBM Digital Library product, NCSTRL)

Star Methodology: Protocols

1. Collect scenarios of communications between components
2. Determine necessary streams
3. Use standards where applicable
4. Specify syntax and semantics of protocol
5. Record results

(e.g., Michigan Agents, Stanford Infobus, Dienst, Z39.50, HTTP/CGI)

Star Methodology: Modules

1. Find tools:
    - object databases
    - relational databases
    - Web servers/browsers/plugins
    - XML parsers
    - workflow tools
    - authoring tools
2. Align with architectures/protocols
3. Record results

(e.g., IBM Digital Library, IBM QBIC, Carnegie-Mellon digital video tools, OCLC SiteSearch for metadata)

Star Methodology: Prototypes

1. Construct "paper prototypes"
    - use sticky notes, drawing paper, transparencies
2. Build "fake" application
    - use SDKs: VB, Visual Café
3. Link screen shots (GIFs + supertitles)
4. Build real user interfaces
5. Connect GUI to application
6. Record results

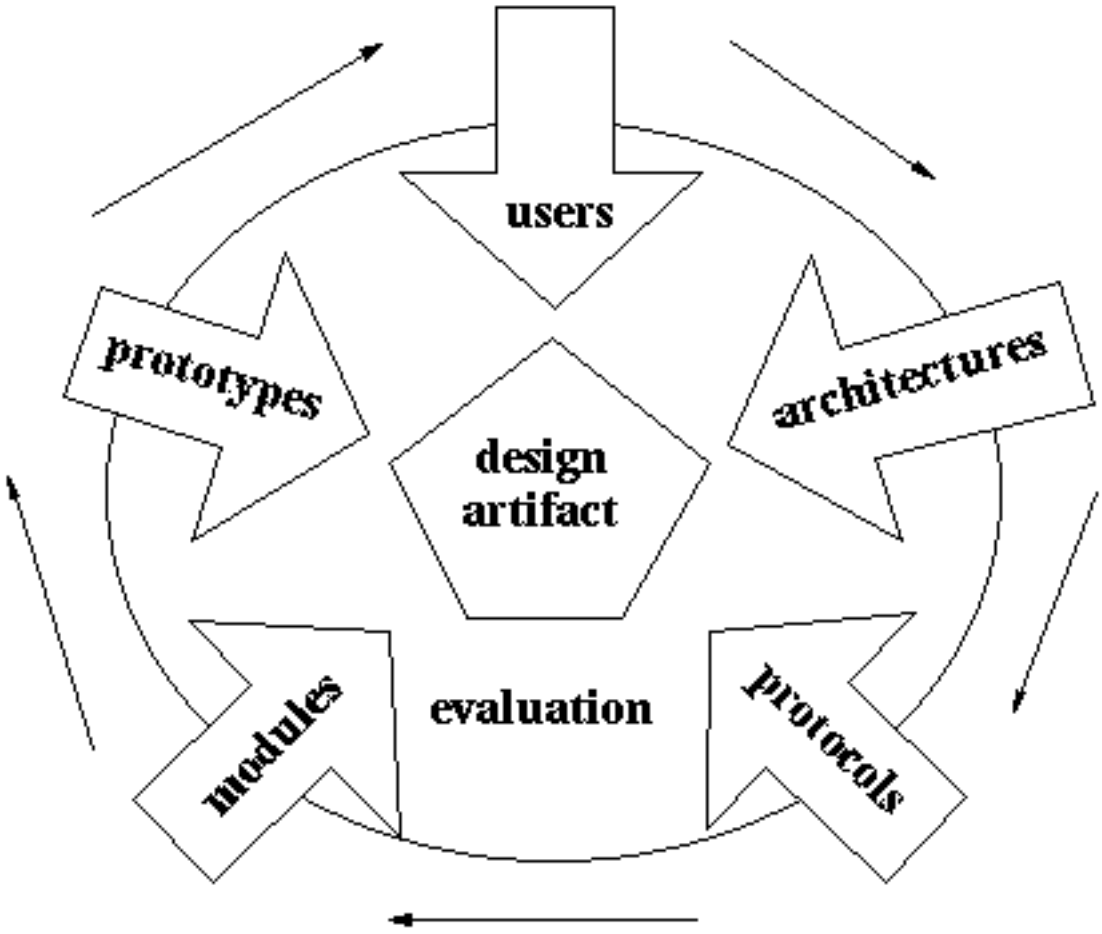
Star Methodology: Evaluation

1. Do "formative analysis"
2. Ensure robustness
3. Provide feedback for designers
4. Ensure robustness---no catastrophic failures allowed!
5. Perform verification and validation
6. Perform usability studies of every "user interface"
7. Record results

- Did we build the right system?
- Did we build the system right?
- Did we log the right data?
- Did we test usability of GUIs, APIs, user manuals, help systems?

Summary

5S Framework	**Star Methodology**
Societies	
Scenarios	
Spaces	
Structures	
Streams	



Questions for Participants

- Did the 5S Framework help you understand digital library components? Why/why not?
- Do you think having the framework is useful for your understanding?
- What are the strengths and weaknesses of the 5S Framework?
- Was the Star Methodology useful for you in your design and development efforts?
- What are the strengths and weaknesses of the Star Methodology?
- Did you have to augment either the framework or the methodology for your work in particular?
- Will you continue to use 5S and Star in this effort? Why/why not?
- Will you recommend 5S and Star for future efforts? Why/why not?

Digital Libraries: Virginia Tech Courseware

To learn about digital libraries, you may wish to visit the Self-Study course materials or the online courses listed below.

- Self-Study Courseware
- Honors 3004: Digital Libraries, Fall 1997, Virginia Tech
- CS6604: Digital Libraries, Fall 1997, Virginia Tech
- PDF file (14M, 639 pages) of all pages of most of readings
- PDF file (8M, 244 pages) of first pages of most of readings
- DL tutorial

Please send comments/suggestions to Ed Fox.

DL Tutorial
DIGITAL LIBRARIES

Virginia Tech
May 20, 1999

Table of Contents

Self-study Courseware on

Digital Libraries

Contents

Introduction: This WWW site has been developed to assist those interested in learning about digital libraries. It is based upon materials tested in 2 Virginia Tech courses taught Fall 1997:

- CS6604
- Honors 3004

Students in those courses especially liked Michael Lesk's "Practical Digital Libraries: Books, Bytes & Bucks" so we refer to it as a supplemental text throughout this site.

There is a set of quizzes to test your knowledge of the chapters in Dr. Lesk's book. We also will support discussion related to these course materials through:

- Hypernews

Revisions: This site will undergo frequent changes, so do check back. The latest revision was completed 6/27/98.

Acknowledgements: This WWW site was developed in part through funding from NSF grants CDA-9312611, DUE-9752408, and DUE-9752190.

Please send comments/suggestions to Ed Fox.

Contents :

Pedagogy:

We recommend that beginners start with the Introduction and then proceed through the Topics, following along with the text by Dr. Lesk. The Resources provide alternate views of the contents, and the References should serve those desiring additional details.

[**Main**]

Please send comments/suggestions to Ed Fox.

Introduction to Digital Libraries:

- Definitions: Some of the attempts made by various people to define 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]**

Definitions :

- "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, Birmingham, 1994)
- "Systems providing a community of users with coherent access to a large, organized repository of information and knowledge."
(Lynch, 1995)
- "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.
(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." (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. 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)
- "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)

- "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, 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, pg. 65)
- "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)
- "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 "
  (Smith, 1995)

Digital Library related terms/glossary

(by Peter Graham, Rutgers University Libraries):

- 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)

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
- What is a "digital library "? (vs. a virtual library)
- UCLA workshop (focusing on user perspectives):
  - Introduction
  - information life cycle
  - Artists
  - Business Records as Artifacts
  - Health-Information Systems
- 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. Also available in PDF.

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:

- 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.

Topics:

- Search, retrieval, resource discovery (See Chapter 2 in Dr. Lesk's book.)
- Multimedia, representations (See Chapter 4 in Dr. Lesk's book.)
- Architectures (See Chapter 6 in Dr. Lesk's book.)
- Interfaces (See Chapter 7 in Dr. Lesk's book.)
- Metadata
- Electronic publishing, SGML
- Database issues
- Agents
- Commerce, economics, publishers (See Chapter 9 in Dr. Lesk's book.)
- Intellectual property rights, copyright laws & security (See Chapter 10 in Dr. Lesk's book.)
- Social issues (See Chapters 11, 12 in Dr. Lesk's book.)

Pedagogy:

We recommend that the topics be covered in the order given above, with the reader examining the material in the book by Dr. Lesk before visiting the online information. Topics that do not correspond to chapters in the book have been included as supplementary material that seemed to be of special interest to students at Virginia Tech, and/or where there is keen interest and progress by the digital library community. However, these can be skipped by novices interested in a general overview.

Search, retrieval, resource discovery:

Searching - LoC

- LoC Home Page
- 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)

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

Columbia

- D-Lib Article on Images/Video
- WebSeek Home Page

BioKleisli

- project
- demos

Filtering

- Defn
- Fast Data Finder: Genetic sequence analysis
- What is *information filtering*? How does it differ from information retrieval?

Cross-Language Information Retrieval Resources

- Eurospider Demos
- Analogical Language Processor Demo
- 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.

Multimedia, Representations:

The Basics:

- text file formats
- graphic file formats
- hypermedia & multimedia

ACM DL'97 Tutorial: Multimedia Information and Systems

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

MHIA Courseware and Curricula

- MHIA Home Page
- SIGIR 96 Workshop
- Drexel 96 Workshop
- IR Courses
- Information Engineering - European Commission: Multimedia Publishing and Distribution, Information Management and Retrieval
- 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 ...)
- **Performance**

[Main] [Contents] [Topics]

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 http://WWW.CNRI.Reston.VA.US/home/cstr/arch/slides.html
  - architecture for information in digital libraries http://www.dlib.org/dlib/february97/cnri/02arm s1.html
  - Digital Object Architecture Project http://www.cnri.reston.va.us/doa.html
- 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.

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 : ENVISION project at Virginia Tech ...

Berkeley: TileBars, Multivalent documents

[Main] [Contents] [Topics]

Please send comments/suggestions to Ed Fox.

Metadata:

- IMS Metadata
- Metadata: the Foundations of Resource Description
- OCLC/NCSA Metadata Workshop Report
- RFC-1807
- TEI
- BASIS article
- D-Lib Working Group on Metadata
- STARTS

[Main] [Contents] [Topics]

Please send comments/suggestions to Ed Fox.

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.

Database Issues:

- UCB database management

[Main] [Contents] [Topics]

Please send comments/suggestions to Ed Fox.

Ontologies and Agents in Digital Libraries

Key topics about *Ontology* adapted from *AI Magazine*, Fall 1997, 18(3), include:

- Defn
- Comparison criteria
- Top level categories, taxonomy. categories, realtions, axioms
- Comparison chart

URLs related include:

- Ontologies
- Indented list diagrams of important ontologies
- CYC Home Page and ontology and table of contents
- WordNet Home Page and online demo
- Generalized Upper Model: model, overall organization, concept hierarchy, relational hierarchy
- UMLS Home Page and fact sheets, MeSH, Grateful Med and demo
- TOVE - Toronto Virtual Enterprise
- KIF - Knowledge Interchange Format and brief intro
- Stanford KSL Network Services and Ontology Editor
- Guided Tour to Developing Ontologies Using Ontolingua
- EUROKNOWLEDGE Glossary etc.
- Stanford DLI and agents, especially for Web browsing
    - InterPay : Shopping Models, Secure Electronic Marketplace for Europe
- ILU and Stanford testbed use
- Agents '97 Conf.
- CHI '97 Software Agents Tutorial by Pattie Maes and her Software Agents Group
- Firefly for music filtering (successor to HOMR from MIT)
- My Yahoo (successor to Webdoggie from MIT)
- IBM Agents, and the Agent Building Environment (ABE): A toolkit for building intelligent agent applications
- NEWS WEEDER - naive Bayes classifier - see AI Magazine Fall 1997 p. 18
- IBM DL: QBIC, watermarking
- Hal Berghel: CACM Nov. 1997 40(11): Watermarking Cyberspace, and IEEE Computer 29:7 article
- DigiCash (Ch. 11)

- Agents: people and places

- iimam@site.gmu.edu adaptatation, intelligence
- yves.Kodratoff@Iri.Iri.fr
- Brian Gaines, U. Calgary: society of agents
- Haynes, Sen : U. Tulsa: cases
- Rus, Dartmouth: gather info
- Decker, Sycara, Williamson: CMU: multiagent society, planning, matchmaker info agent

Questions:

- Try WordNet on "library" and look for coordinate terms on senses 1,2,3
- Try Grateful Med and find MeSH / Meta Terms for "diabetes"

Commerce, Economics, Publishers:

NetBill

- Home Page
- Demo
- Overview article on payment systems from IEEE Spectrum
- Questions: How would this work with ETDs? What are the advantagesand disadvantages relative to other approaches?

Commerce part of CS6604 lecture

- Workshop on Tech. of Terms and Conditions and Final Report to NSF - including Breakout Group Reports
- 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.
- Overview of Copyright Laws in the Digital Domain and References : Check out the references for some very good links and information on copyright laws and related issues.
- Pamela Samuelson and pointers based on her pages and recommendations
- Electronic Commerce
- Workshop on Tech. of Terms and Conditions and Final Report to NSF - including Breakout Group Reports
- 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

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.

Social Issues:

- Social Aspects D-Lib Working Group
- UCLA Workshop, Social Aspects of Digital Libraries, Feb. 16-17, 1996 http://www-lis.gseis.ucla.edu/DL/
- Life Cycle http://www-lis.gseis.ucla.edu/DL/UCLA_DL_model.gif

---

[Main] [Contents] [Topics]

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 submissions

DLI-1

- DLI-1 home page at U. Illinois
- DLI-1 information & resources
- 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

American Memory Project (Library of Congress)

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
  - Project Report for NSF EI PI Meeting, Nov. 1996
- **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**
  - overview http://ei.cs.vt.edu/~cs5604/Adv/Adv-MARIAN.html
  - system http://opac3.cc.vt.edu/htbin/marian
- 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

Some Extra Virginia Tech resources on various projects:

- Build upon existing electronic materials
  - Netlib (numerical analysis) http://www.netlib.org/
    - Attribute/value 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:5047/oclc/research/projects/core/
  - Elsevier
    - Science Direct 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
  - 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 server features
  - 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 orientation
- Harmony screens http://ei.cs.vt.edu/~cs5604/Adv/Adv-Harmony.html
- Harmony information structuring
- Harmony document viewers
- 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 1992
    - Screen for Article
- Discipline - Greek Literature http://www.perseus.tufts.edu/
  - Evaluation - 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/approval.htm
    - Letter to students http://etd.vt.edu/submit/letter.htm
    - Standards http://etd.vt.edu/submit/mm.htm
  - Collection http://www.theses.org
  - 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 proposal
            - abstract http://www.ndltd.org/support/fipseabs.htm
            - full-text http://www.ndltd.org/support/fipse10.pdf

- PDF PART 1 - PROJECTS
    - Fig. 1: Timeline of Recent Information & DL Systems
    - Fig. 2: NCSTRL Architecture
    - NETLIB (numerical analysis)
    - CORE (chemistry)
    - TULIP (material science & engineering, with Elsevier, OCLC)
    - IBM digital libraries products and projects
    - Hyper-G/HyperWave (clients and servers)
    - BEV HistoryBase
    - CS technical reports (CS-TR, WATERS, NCSTRL) and related efforts
    - CS education (ACM literature, courseware on IR, multimedia, hypertext, history)
    - Digital Library Initiative (CMU, Michigan, Stanford, UC Berkeley, UC Santa Barbara, University of Illinois Urbana-Champaign)
    - ETD (electronic theses and dissertations - NDLTD)
- PDF PART 2 -SOURCES, RESEARCH
    - Digital Library conferences
    - IITA meetings (e.g., May 1995 workshop)
    - Allerton Institutes (from U. Illinois, NSF)
    - D-Lib (research, magazine, working groups)
    - D-Lib research articles (architecture, metadata, URNs, use)
    - Virginia Tech information (DL page, Sourcebook)
    - Virginia Tech projects (Envision, ILDLCS, WWW traffic analysis)
    - Z39.50 (overview, OCLC, CNIDR)
    - Library of Congress
    - CNRI (architecture, handles)
    - UMBC agents

- LIS: preservation, TEI, ...

SMETE-lib (undergraduate education in US)

SMETE - NSF Science Mathematics, Engineering and Technology Education Digital Library

NEEDS - National Engineering Delivery System

Efforts Outside US

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

Electronic Theses and Dissertations Initiative: NDLTD project, Collection, Submission Instructions

International Digital Libraries Association: home page

NSF-EU Working Groups and Meetings: home page

Singapore Network: SINGAREN

[Main] [Contents] [Resources]

Please send comments/suggestions to Ed Fox.

People:

Rob Akscyn of Knowledge Systems Incorporated with its PetaPlex Project

Dan Atkins University of Michigan Digital Library Project Director.

Howard Besser of School of Information Management and Systems at Berkeley

Bill Birmingham University of Michigan Digital Library Researcher.

Chris Borgman of Information Studies at UCLA

Hsinchun Chen Head of the AI Lab of U. Arizona

Edward A. Fox Director of the Digital Libraries Research Group at Virginia Tech.

Rick Furuta of CS at Texas A&M Univ.

Hector Garcia-Molina In the Stanford DB Group

Henry Gladney at IBM Almaden Research Laboratory

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

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 in that year's volume online in ACM DL

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

Bruce Schatz Principal Investigator of University of Illinois at Urbana-Champaign, DLI 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

Robert Wilensky Principal Investigator of Berkeley DLI Project

---

[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

---

[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 Digital Library

IBM

- IBM DL Home page
- IBM Renaissance Consortium Panel and workshop
- images - QBIC

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/oclc/research/index.htm
  SiteSearch http://www.oclc.org/oclc/menu/site.htm

Xerox

- 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.

Library of Congress:

- American Memory http://lcweb2.loc.gov/
- Call about American Memoryhttp://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.

DLI - Stanford:

- Home Page
- IEEE Computer article
- testbed development
- info finding
- user interfaces
- DLITE (task env)
- DLITE comps
- DLITE screens

[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
- 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
- Tutorial
- World Spatial Data
- Annual Report
- 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, concept extraction, concept spaces, crack semantic barrier
- Interspace Reports
- Social Science Home Page: Fn Reqts
- 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.

References:

- Courses: Digital Library and related courses being offered at various Universities.
- Conferences/Workshops: Links to various conferences/workshops that have been held in the recent past or will be held in the near future.
- Journals: Digital Library related journal information with links.
- Repositories & Bibliographies: contains information and links to some of the repositories maintained by various organizations such as the D-Lib Magazine.
- Books: Some books that contain valuable information on Digital Libraries (along with links to some publishers)

[Main] [Contents]

Please send comments/suggestions to Ed Fox.

Digital Library and related courses:

- Digital Library course offered at Pittsburgh
- Michael Lesk's Digital Library course at Columbia University
- University of Missouri course on Library Information Systems
- Virginia Tech
  - CS6604 (1997) Digital Libraries
  - UH3004 Fall 1997 Honors 3004 - Digital Libraries
  - CS5604 Information Storage and Retrieval
  - CS4624 Multimedia, Hypertext and Information Access
  - CS6604 (1995) Interactive Accessibility
- CSEI: NSF CS Education Innovation - projects around the nation
- Furman University:
  - Exploring the Digital Domain
  - Web site on creating WWW pages
- 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.
- Digital Library (Alexandria) Online Tutorial at UCSB

[Main] [Contents] [References]

Please send comments/suggestions to Ed Fox.

Conferences/Workshops:

- ACM DL'99: Berkeley, Aug. 11-14 http://fox.cs.vt.edu/DL99/
- ACM DL'98: Pittsburgh, June 23-26 http://www.ks.com/DL98/
- ACM DL'97: Philadelphia, July 23-26 http://www.lis.pitt.edu/~diglib97/
- DL'96: Bethesda, March (1st ACM ...) http://fox.cs.vt.edu/DL96/
- DL'95: Austin, June http://csdl.tamu.edu/DL95/
- DL'94: Texas A&M University
- CoLIS3: Third Int'l Conf. on Conceptions in Library and Information Science: Digital Libraries: Interdisciplinary Concepts, Challenges and Opportunities, Dubrovnik, May 1999
- Santa Fe Workshop, Digital Knowledge Work Environments, March 9-11, 1997 http://www.si.umich.edu/SantaFe/
- UCLA Workshop, Social Aspects of Digital Libraries, Feb. 16-17, 1996 http://www-lis.gseis.ucla.edu/DL/
  - life cycle
- IITA Digital Libraries Workshop, 1995
- Allerton, 1996 http://edfu.lis.uiuc.edu/allerton/96/ and map
- Allerton, 1995 http://edfu.lis.uiuc.edu/allerton/95/
- ADL 99, IEEE Advances in Digital Libraries May 19-21, 1999, Baltimore, MD
- ADL 96, Forum on Research and Technology Advances in Digital Libraries May 13-15, 1996, Washington, D.C.
- KOLISS DL 96, Proc. Int'l Conf. on Digital Libraries and Information Services for the 21st Century, Sept. 10-13, 1996, Seoul, Korea

- DLI Funded Workshops http://www.dli2.nsf.gov/workshops.html

- D-Lib supported meetings, conferences and workshops http:/ /www.dlib.org/groups.html

---

[Main] [Contents] [References]

Please send comments/suggestions to Ed Fox.

Journals:

Selected special issues include:

- Commun. ACM
  - April 1995: 38(4)
  - April 1998: 41(4)
- IEEE Computer, May 1996
- J. American Society for Information Science, Sept. 1993: 44(8)
- J. of Visual Communication and Image Representation, 7(1), March 1996

There also are closely related journals like:

- J. of Digital Information (British Computer Society)

[Main] [Contents] [References]

Please send comments/suggestions to Ed Fox.

Repositories & Bibliographies:

- **D-Lib** http://www.dlib.org/
  - Articles (by author) http://www.dlib.org/author-index.html
  - Articles (by title) http://www.dlib.org/title-index.html
  - Research Projects (incl. DLI) http://www.dlib.org/projects.html
  - D-Lib Working Groups http://www.dlib.org/groups.html
    - Metadata http://www.dlib.org/metadata/overview.html
    - Naming http://www.dlib.org/naming/overview.html
    - Repository Interfaces http://www.dlib.org/repository/overview.html
    - Social Aspects http://www.dlib.org/social/overview.html
  - D-Lib Magazine Articles on Key Topics
    - Agents http://www.dlib.org/dlib/July95/07birmingham.html
    - Architecture (incl. handles) http://www.cnri.reston.va.us/home/dlib/July95/07arms.html
    - Metadata http://www.dlib.org/dlib/July95/07weibel.html
    - Uniform Resource Names (URNs) http://www.dlib.org/dlib/february96/02arms.html
    - Use http://www.dlib.org/dlib/october95/10bishop.html
    - Informedia http://www.dlib.org/dlib/july96/07wactlar.html
    - Variations http://www.dlib.org/dlib/june96/variations/06fenske.html
    - Access Control: Articles by Gladney et al.

- UIUC Pointers to Publications http://dli.grainger.uiuc.edu/pubsnatsynch.htm

- **Annotated Bibliography for Digital Libraries** - Christine Woerner, 1996 (Case Studies, DL as Place, Archive/Organization/Preservation, Librarianship, Mediation/Interaction, Authoring/Authenticity/Originality)

- Virginia Tech
  - Digital Library Research Laboratory Publications
  - BibTeX file for article: E. Fox and O. Sornil. Digital Libraries. Chapter 11 in Modern Information Retrieval, AWL England, 1999: Ricardo Baeza-Yates and Berthier Ribeiro-Neto, eds., to appear.
  - misc ptrs http://scholar.lib.vt.edu/digilib/

Books:

There is only one really good book on digital libraries:

- Michael Lesk, Practical Digital Libraries, Morgan Kaufmann, 1997, San Francisco

For a history of many digital library activities through Fall 1993, including reports on key workshops, see:

- Digital Library Source Book, Edward Fox, ed., 1993 http://fox.cs.vt.edu/DLSB.html

In the related field of Information Retrieval the best set of readings is:

- Karen Sparck Jones and Peter Willett, Readings in Information Retrieval, Morgan Kaufmann, 1997, San Francisco

Some miscellaneous related works include:

- Elsevier, TULIP Final Report, 1996, New York. This booklet was distributed after completion of the TULIP digital library prototype project by Elsevier, and led to their current digital library effort, EES.

- Hermann Maurer, ed., Hyper-G/Hyperwave: The Next Generation Web Solution, Addison Wesley Longman, 1996, Harlow, England

- Setrag Khoshafian, A. Brad Baker, MultiMedia and Imaging Databases, Morgan Kaufmann, 1996, San Francisco

- V.S. Subrahmanian, Sushil Jajodia, eds., Multimedia Database Systems: Issues and Directions, Springer, 1996, Berlin

---

[Main] [Contents] [References]

Please send comments/suggestions to Ed Fox.