

User Interfaces and Information Seeking in Digital Libraries: A Tutorial

Gary Marchionini

University of North Carolina at Chapel Hill

March@ils.unc.edu

ICADL 2001

Bangalore, India

December 10, 2001



Outline

- First Principles
- Digital Libraries: User Participation & the Sharium
- Interactive Model of Retrieval
- Examples of Agile View Design Techniques
- The Open Video Project Case
- Evaluation
- Implications of DLs

First Principles

Gary Marchionini-UNC-Chapel Hill

First Principles: People

- People and needs are diverse
- People seek least effort and maximum payoff
- Vision is both a primary channel and a metaphor
- People want/need to share
- Innovation adoption is not a linear process

Gary Marchionini-UNC-Chapel Hill

First Principles: Interfaces for Information Seeking

- We do IR research so that people can find information
- The IR problem is distinct from the DB problem
- Treating IR as a query-document matching problem is the technical view
- Information seeking is embedded in real life tasks: The production paradox
- Information seeking is complex: Analytical and browsing strategies

Gary Marchionini-UNC-Chapel Hill

First Principles: User Studies

- Empirical evidence is crucial to design principles involving human interaction
- User studies involve interactions of individual characteristics, system features, content, and context
- Types of studies
 - Expert Critiques
 - Discount usability tests
 - Formal laboratory studies
 - Field testing

Gary Marchionini-UNC-Chapel Hill

First Principles: Libraries

- Libraries assess, collect, preserve, organize, provide access to information and promote its use to improve the human condition.
- Libraries are highly burdened to accommodate physical and electronic materials and increasing needs for information by larger and more diverse populations.
- Internet and WWW offer universal communication and publishing and challenge libraries to reconsider fundamental missions--DLs
- Leverage the DL challenge to develop a broader vision of information services for the common good: Sharium is a metaphor for this

Gary Marchionini-UNC-Chapel Hill

Digital Libraries: User Participation and the Sharium

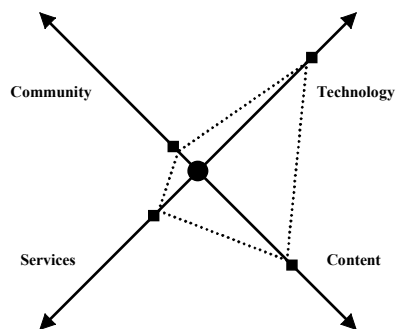
Gary Marchionini-UNC-Chapel Hill

What is a DL?

- Characteristics
 - electronic digital formats
 - networked (sharable information)
 - organization apparent (a library not a pile)
 - Collection development policy
 - Systematic data structuring and tagging
 - use (fair) policy
 - persistent
 - guidance and referral
 - community based
- Motivations: technology, funding, democracy

Gary Marchionini-UNC-Chapel Hill

Digital Library Design Space



Gary Marchionini-UNC-Chapel Hill

DL Missions

- DLs clearly must aim to collect, manage, and preserve electronic expressions of knowledge (this is a well-established mission).
- *Knowledge is in people's heads--DLs should aim to facilitate the use and development of the collective knowledge in human consciousness*
- *Human attention is a fundamental natural resource--DLs should provide tools and resources (material and expertise) to help optimize this resource*

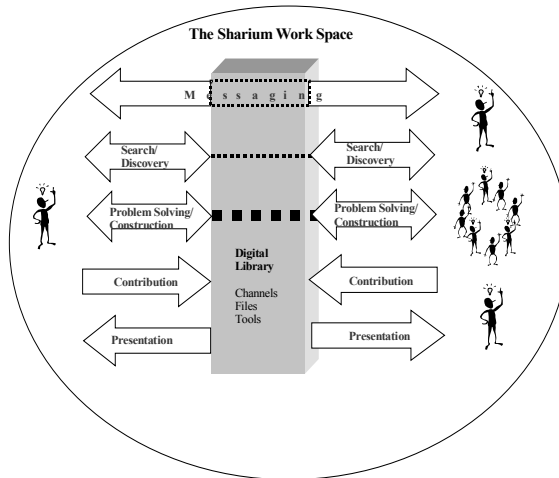
Gary Marchionini-UNC-Chapel Hill

Sharium

- A virtual workspace with rich content and powerful tools where people can work independently or collaborate with each other to learn and solve information problems. A collaborative problem solving environment.
 - Organized around resources and tools
 - Encourages contributions and participation
 - Is sustainable

Gary Marchionini-UNC-Chapel Hill

Sharium Workspace



Gary Marchionini-UNC-Chapel Hill

Query & Selection

- Interfaces
 - Natural language queries
 - Dynamic queries
 - Alternative interfaces
 - Help/support
- Consortia/portals/channels
 - Interoperation
 - Selection and merging

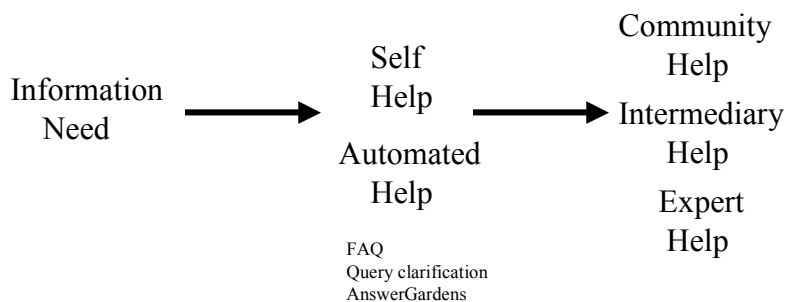
Gary Marchionini-UNC-Chapel Hill

Reference & Question Answering

- Help people help themselves
- Elicitation
- Layered services
- Quality control
- Economic model
- Privacy
- Shared views/clients

Gary Marchionini-UNC-Chapel Hill

Cascading Assistance

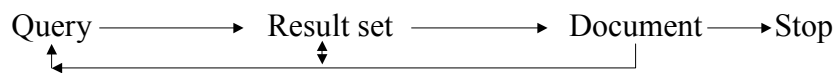


Gary Marchionini-UNC-Chapel Hill

Interactive Model of Retrieval

Gary Marchionini-UNC-Chapel Hill

Iterative Retrieval Model



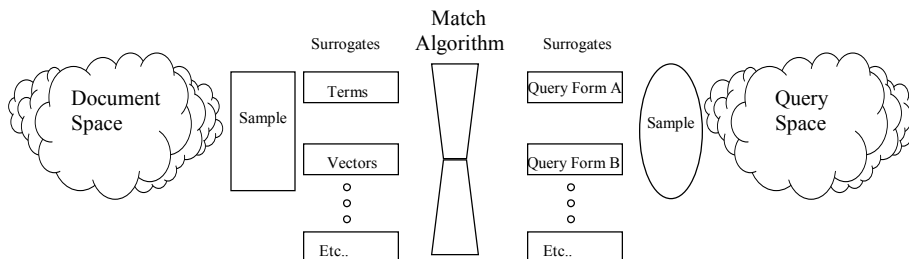
Agileview Interaction Model



- Tight coupling of functions**
- Highly interactive control mechanisms**
- Flexible, non-linear options**
- Result Set manipulations added**
- Document processing tools added**

Gary Marchionini-UNC-Chapel Hill

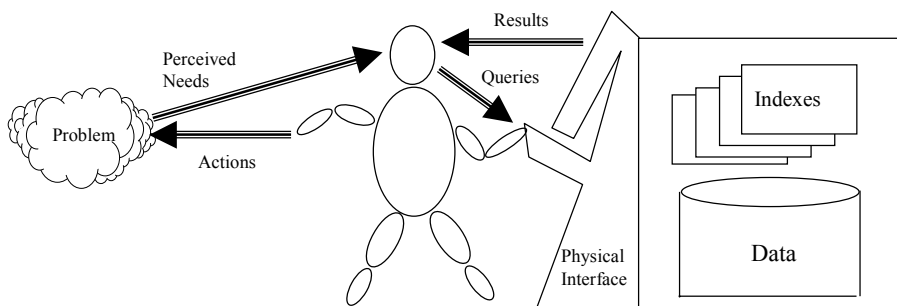
Technical View: Retrieval as Matching Documents to Queries



Retrieval is algorithmic. Evaluation is typically a binary decision for each pairwise match and one or more aggregate values for a set of matches (e.g., recall and precision).

Gary Marchionini-UNC-Chapel Hill

Human View: Information-Seeking Process



Information seeking is an active, iterative process controlled by a human who Changes throughout the process. Evaluation is relative to human needs.

Gary Marchionini-UNC-Chapel Hill

MultiView Interaction

- Integrate query and browsing
- Closely couple query and results
- Highly interactive control mechanisms (direct manipulation)
- Overviews and Previews
- Alternative interfaces (views)

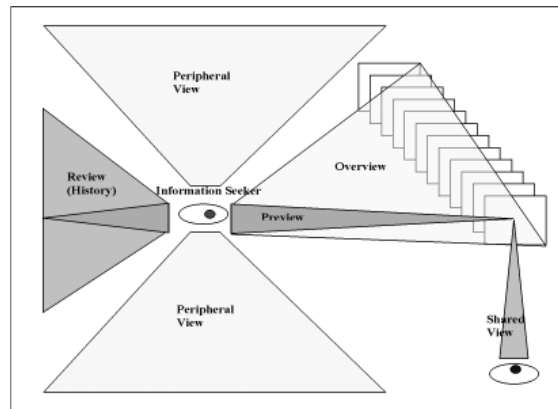
Gary Marchionini-UNC-Chapel Hill

Design Strategies

- Consider the information seeker's context
 - Cognitive accessibility (it does not matter how good the results are if the information cannot be easily understood)
 - Cost-benefit assessment (it does not matter how good results are if there is no time to use it)
- Study special populations (cell biologist vs. practicing physician)
- Usability testing approach (iterative, impressionistic)
- Systematic case studies
- Epidemiology approach (start with outcomes and trace influences)
- Develop an IS interaction model

Gary Marchionini-UNC-Chapel Hill

Agile Views Framework



Gary Marchionini-UNC-Chapel Hill

Theory to Practice

- Design challenge 1: creating views
 - What granularities (collections and items)
 - Which attribute sets?
 - Creating or extracting metadata
- Design challenge 2: manipulating/controlling views
 - Perceptual estimation (e.g., look ahead)
 - Physical and conceptual inertia

Gary Marchionini-UNC-Chapel Hill

Examples of Agile View Design Techniques



- Relation Browser
 - Federal statistics, overviews of relationships (several different partitions). Useful for small number of attribute sets, each with small number of attribute values. Backend database of metadata, Java applet interface
- Enriched Links
 - Complex web sites, previews, overviews, and reviews of pages. Backend computation and Javascript interface
- Integrated overviews and previews
 - Multimedia digital library, backend computation, Java applet interface

Gary Marchionini-UNC-Chapel Hill

Relation Browser

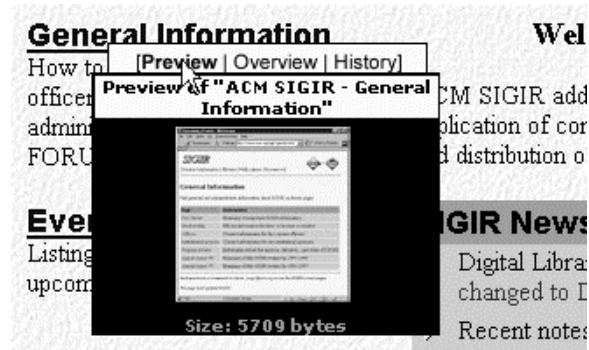


This is an experimental tool that requires Netscape 4.5 (or higher) or Internet Explorer 4.0 (or higher). Please be patient as the tool may take some time to load. After you have used the tool, please take a few minutes to give us some feedback on it. [Feedback](#)

Web Site	International	National	Regional	State	Sub-State
Administration on Aging		X		X	
Administration for Children & Families				X	
Adoption Assistance & Foster Care				X	
Agency for Internal Development	X				
Aging Population		X		X	
Agriculture Research Service		X		X	X
Child Abuse		X		X	
Child Care		X		X	
Child Support Enforcement		X		X	

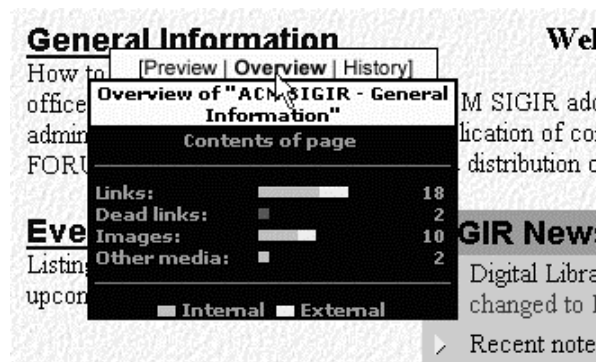
Gary Marchionini-UNC-Chapel Hill

Enriched Links: Preview



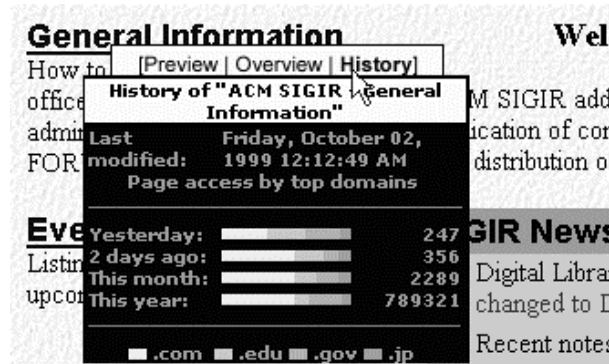
Gary Marchionini-UNC-Chapel Hill

Enriched Links: Overview



Gary Marchionini-UNC-Chapel Hill

Enriched Links: Shared View



General Information

How to [Preview | Overview | History]

History of "ACM SIGIR General Information"

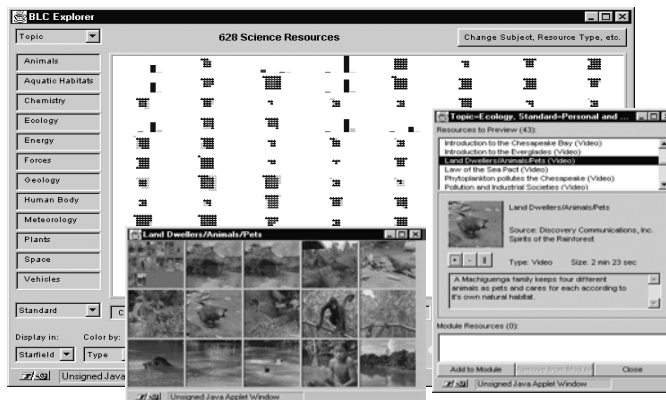
Last modified: Friday, October 02, 1999 12:12:49 AM
Page access by top domains

Yesterday:	247
2 days ago:	356
This month:	2289
This year:	789321

.com
 .edu
 .gov
 .jp

Gary Marchionini-UNC-Chapel Hill

Overviews and Previews: One Screen



628 Science Resources

Change Subject, Resource Type, etc.

Topic: Ecology, Standard-Personal and

Resources to Preview (4):
 Introduction to the Chesapeake Bay (Video)
 Introduction to the Everglades (Video)
 Land Dwellers/Animals/Pets (Video)
 Law of the Sea Pact (Video)
 Physiography of the Chesapeake (Video)
 Pollution and Industrial Societies (Video)

Land Dwellers/Animals/Pets

Source: Discovery Communications, Inc.
 Sparks of the Rainforest

Type: Video Size: 2 min 23 sec

A Machiguenga family keeps four different animals as pets and cares for each according to its own natural habitat.

Module Resources (0)

Add to Module Close

Gary Marchionini-UNC-Chapel Hill

The Open Video Project Case

Gary Marchionini-UNC-Chapel Hill

A New Testbed for Agile Views

- Open Video Project (www.open-video.org)
 - Research community
 - Contributory facility
- Granularities: collection of videos and collection of segments
- Attributes: three levels of metadata

Gary Marchionini-UNC-Chapel Hill

Overviews and Previews

- Need to gain understanding of neighborhood of objects (the aboutness problem)
- Need to quickly understand whether an object is interesting (the relevance problem)
- Digital Libraries exacerbate the problems
 - one view fits all (screen, levels of granularity, media)

Gary Marchionini-UNC-Chapel Hill

Why Video Browsing?

- Digital Libraries and Video-on-demand applications->lots of digital video
- As part of retrieval
 - embedded in larger task
 - quick decisions about rejection
- To speed basic understanding (accretion)
- To save time, bandwidth, money

Gary Marchionini-UNC-Chapel Hill

Key Problems

- What to represent (the representation problem)
 - Case Dependence: “The usefulness of a representation depends upon how well-suited it is to the purpose for which it is used.”
David Marr, 1980, *Vision*
 - Implies a need for MULTIPLE LEVELS of representation
- How to control the representations (the user control mechanism problem)

Gary Marchionini-UNC-Chapel Hill

Video Hierarchy

- Video
 - Clip (segment) [conceptual/editorial; physical]
 - Sequence (scene) [conceptual/editorial]
 - Shot [camera specific]
 - » Frame [physical]

Gary Marchionini-UNC-Chapel Hill

Video Surrogates

- Linguistic information
 - bibliographic records
 - descriptors, extracts (e.g., transcripts, close cap)
 - reviews
- Audio information (speech, music, effects)
- Clips
 - rushes/out takes/trailers/teasers
 - original vs extractions
- Fast Forwards (compress time)
- Key Frames (aka poster frame, thumbnail, storyboard)

Gary Marchionini-UNC-Chapel Hill

Key Frames

- Video segmentation (chunking)
 - scene changes
 - other large changes (camera, sound)
 - signal processing: color histogram; motion; luminosity; texture; voice; music
- A frame from the change is significant (thesis sentence?)
- Salient stills

Gary Marchionini-UNC-Chapel Hill

Control Mechanisms

- Dynamic (surrogates move, user may have some control over movement)
 - Slide shows
 - display rate
 - image size
 - Fast forward
 - not effective for key frames
 - Multiple concurrent surrogates
 - 2 feasible, 3, 4 problematic
 - multi-video view (Bellcore) within single video

Gary Marchionini-UNC-Chapel Hill

Control Mechanisms

- Static
 - Story Boards (aka filmstrip, v-wall, v-array)
 - layout
 - array size
 - with or without labels (e.g., words, time codes)
 - Hierarchical
 - key frame structure (Singapore)
 - Network
 - Scene Transition Graph (IBM)
 - Extract
 - Streamer (MIT)
 - Salient Still (MIT)

Gary Marchionini-UNC-Chapel Hill

Control Mechanisms

- Hybrid
 - image tree (Bellcore) TOC+full
 - video skim (Informedia) story board+mini-clips+sound (e.g., show frames 1-5; 50-55, 106-120, etc. OR base on audio, e.g., best key words from transcript)













Gary Marchionini-UNC-Chapel Hill

Compression and Compaction

- Compression (e.g., MPEG)
 - save space
 - save bandwidth (transmission time)
- Compaction: the HUMAN perspective
 - 2 minute clip, show 10 key frames for 500 ms each
 - $30 \text{ fps} * 120 \text{ sec} = 3600 \text{ frames} \rightarrow 360:1 \text{ compression?}$
 - $120 \text{ seconds} : 5 \text{ seconds} \rightarrow 24:1 \text{ compaction}$
 - We prefer the human-meaningful ratio

Gary Marchionini-UNC-Chapel Hill

Overview: Open Video

Search Results						
Preview	Video Title	Video Segment	Download	Details	File Size (MB)	Duration (secs)
	Women Scientists	Segment 1			63.104	323.19
	Women Scientists	Segment 2			22.272	113.88
	Women Scientists	Segment 3			25.6	130.96
	Women Scientists	Segment 4			116.736	598.5

Gary Marchionini-UNC-Chapel Hill

One of many previews: Open Video

Search 1

Preview

FACES BRIGHTNESS SUPER

— Segment Storyboard - Index By Faces —

SORT BY:







cs)

9

8

6

5

Britton, South Dakota, Segment 4

1 to 12 of 43 shown

Gary Marchionini-UNC-Chapel Hill

Give People Flexibility!

- Multiple views require rich and accessible metadata
- Control mechanisms are kludges in today's WWW environment
- A click is a terrible thing to waste!

Gary Marchionini-UNC-Chapel Hill

Evaluation

Gary Marchionini-UNC-Chapel Hill

Evaluation Perspective

- Need to choose
 - product testing
 - controlled comparisons
- Need to assess
 - system performance
 - outcome research (e.g., social programs)
- Need to understand
 - basic research

Gary Marchionini-UNC-Chapel Hill

Existing Models

- Library Effectiveness
 - circulation
 - collection size
 - reference encounters
 - satisfaction
- Information Retrieval
 - recall/precision tradeoff
 - satisfaction

Gary Marchionini-UNC-Chapel Hill

Claims

- Today's IR systems are not comparable to paper-based systems.
 - bibliographic, full-text, and multimedia IR systems are not comparable
- Complex systems are greater than the sum of their respective components.
 - systems that include human components are inherently complex
- Information seeking is an interactive process.
 - different users, domains, and settings require distinct IR system capabilities

Gary Marchionini-UNC-Chapel Hill

Evaluating New Systems

- “We may never know quantitatively the impact of these combined effects, partly because we don't know what would have happened without the collaboratory.”

William Wulf, *The National Collaboratory--
A White Paper*, 1989

Gary Marchionini-UNC-Chapel Hill

Evaluate Systems

- TREC ad hoc and routing evaluations
- TREC interactive track
 - introduces the user as a component but not the problems, perceived needs, and actions
- Hybrid solutions
 - human + automatic
 - statistical + natural language processing

Gary Marchionini-UNC-Chapel Hill

Evaluate Actions: Medical Case

- Does the patient recover?
- Were good decisions made?
 - patient, physician, hospital, HMO views?
- Difficult (impossible?) to disambiguate component effects
- Task-oriented studies (e.g., Hersh's medical student decisions)

Gary Marchionini-UNC-Chapel Hill

Evaluate Interactions



- Think aloud protocols
- Observations, Transaction log analysis
- Interviews, Stimulated recall
- Error analysis
- Time on task
- Cost-benefit analysis
- Questionnaires
- Simulations

Gary Marchionini-UNC-Chapel Hill

New, User-oriented Questions



- Given many relevant documents, which can be most easily processed/understood?
- What are the cost-benefits to different stakeholders?
- What are the organizational/institutional changes due to a system?
- What are the most useful surrogates (representations) for multimedia objects?
- How to best integrate results
 - multiple retrieved sets
 - multiple evaluation efforts

Gary Marchionini-UNC-Chapel Hill

Alternative Strategies

- Consider the information seeker's context
 - Cognitive accessibility (it does not matter how good the results are if the information cannot be easily understood)
 - Cost-benefit assessment (it does not matter how good results are if there is no time to use it)
- Study special populations (cell biologist vs. practicing physician)
- Usability testing approach (iterative, impressionistic)
- Systematic case studies
- Epidemiology approach (start with outcomes and trace influences)
- Develop an IR interaction model

Gary Marchionini-UNC-Chapel Hill

The Perseus Case

www.perseus.tufts.edu

- Multiple stakeholders, methods, and components
- A set of evaluation questions (learning, teaching, system, publishing)
- Longitudinal effects
 - mechanical advantages
 - side effects
 - new types of learning and teaching
 - systemic change

Gary Marchionini-UNC-Chapel Hill

Implications of DLs

Gary Marchionini-UNC-Chapel Hill

Long-Term Implications for Social Sciences Research

- What does it mean when you can have **everything you can possibly access anywhere, available everywhere?** Removing the bounds of access implies ubiquity and augmented memory. What does it then mean to be informed? Intelligent? There are a cascading set of issues: trust, ownership/IP, communication, human relationships, socio-technical symbiosis, etc.

Gary Marchionini-UNC-Chapel Hill

Short-Term Implications for SS Research

- Internet Impact Group distinguishes using Internet to do SS research and SS research on impact of Internet on SS problem—we can make similar distinction with DLs. Examples
 - Statistical Data collection and access
 - Scholarly communication
 - Human behaviors
 - Information flow

Gary Marchionini-UNC-Chapel Hill

Statistical Data Archives

- Odum Institute archives
 - Quick search (collection overview)
 - Catalog
 - Archives (some online)
- ICPSR, other consortia
- From FedStats to the Bureau of Labor Statistics
 - Data, reports, summaries
 - Much more citizen-oriented (selective access to public data query)
- Online data collection (from CATI/CAPI to web-surveys)

Gary Marchionini-UNC-Chapel Hill

Scholarly Communication and Learning

- Access to primary materials: e.g., UNC Library online journals and databases
 - How does this change your writing & citing?
 - How does it change your teaching?
- Improved collaboration potential
- New scholarly structures
 - Crane's shift from monastery to university with library as cathedral, and now from university to network community
- New publication venues
 - Libraries as publishers versus as IP police
 - Print becoming invisible (e.g., Nature paper)

Gary Marchionini-UNC-Chapel Hill

Human Behavior

- How do people seek information? Learn?
 - Who uses DLs? For what? How/Why? How do these results differ from physical libraries?
- How do people work? Collaborate?
 - How do we conceptualize productivity? What are the mechanisms to assure quality? Reliability? Build trust?
- Techniques
 - Transaction log analysis
 - Online surveys

Gary Marchionini-UNC-Chapel Hill

Information Flow

- Information Life Cycle changes?
 - Creation → pub/review/dissemination → use → re generation/dispensation
 - Accelerate cycle rates?
 - Add new feedback channels? (e.g., at BLS, I hypothesize that good user interfaces attract more diverse users, which in turn not only affects the publication phase but also propagates back to the creation phase, i.e., affect the survey(s))

Gary Marchionini-UNC-Chapel Hill

Integration Hypothesis

- As information resources and technologies are integrated as digital libraries (sharia or collaboratories), institutional boundaries will blur. Examples:
 - Types of learning (formal, informal, professional)
 - Types of libraries
 - Levels of government (local, state, federal)

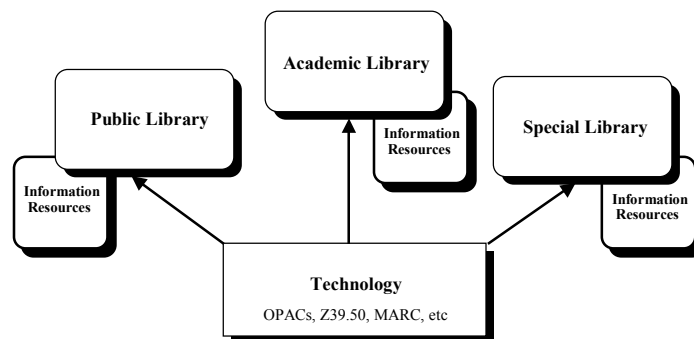
Gary Marchionini-UNC-Chapel Hill

Key Social Challenges

- Why do people share?
- Models?
 - open source?
- Creation-ownership relationship
- Creating and maintaining trust
- Quality assurance
- Sustaining participation

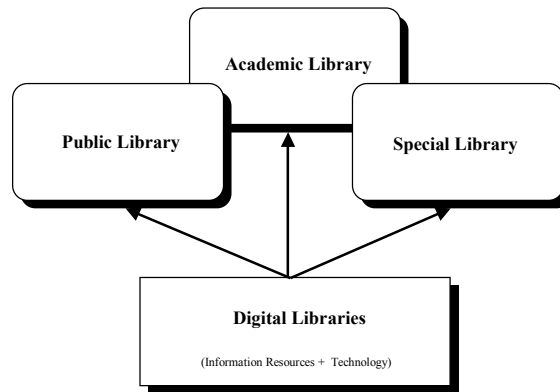
Gary Marchionini-UNC-Chapel Hill

Current model of technological support for types of libraries



Gary Marchionini-UNC-Chapel Hill

Shared Digital Libraries Lead to Integrated Resources and Services (Federation)



Gary Marchionini-UNC-Chapel Hill

Selected Pointers

see www.ils.unc.edu/~march

- Marchionini, G. (2001). Evaluating Digital Libraries: A Longitudinal and Multifaceted View. *Library Trends*, 49(2), 304-333.
- Fox, E. & Marchionini, G. 2001. Digital libraries (Guest editors introduction to special issue). *Communications of the ACM*, 44(5), 31-32.
- Marchionini, G. & Fox, E. (1999). Progress toward digital libraries: Augmentation through integration (Guest Editors' Introduction). *Information Processing & Management*. 219-225.
- Marchionini, G., Plaisant, C., & Komlodi, A. (1998). Interfaces and tools for the Library of Congress National Digital Library Program. *Information Processing & Management*, 34(5), 535-555.
- Fox, E. & Marchionini, G. (1998). Toward a worldwide digital library. *Communications of the ACM*, 41(4).
- Marchionini, G., & Mauer, H. (1995) The roles of digital libraries in teaching and learning, *CACM*, 38(4), 67-75.
- Marchionini, G. (1998). Digital Library Research and Development. In A. Kent (Ed.) *Encyclopedia of Library and Information Science*. Vol 63, Supplement 26. NY:Marcel Dekker. 259-279.
- Marchionini, G. (1999). Augmenting library services: Toward the sharium. *Proceedings of International Symposium on Digital Libraries 1999* (Tsukuba, Japan, September 28-29, 1999). 40-47.
- Marchionini, G., Plaisant, C., & Komlodi, A. (forthcoming). The people in digital libraries: Multifaceted approaches to assessing needs and impact. In A. Bishop, B. Buttenfield, & N. VanHouse (Eds.) *Digital library use: Social practice in design and evaluation*. Cambridge, MA: MIT Press.

Gary Marchionini-UNC-Chapel Hill