User Interfaces and Public Information Spaces

Gary Marchionini, PhD
University of North Carolina at Chapel Hill

March 14, 2006

US EPA Web Workgroup Conference
Research Triangle Park, NC

Outline

- Human-centered design perspective
- BLS longitudinal case
- HCI and IR landscape: human-information interaction perspective
- Relation Browser and Open Video Examples
- Implications for design and practice

- ACM/IEEE Joint Conference on Digital Libraries June 11-15, 2006, Chapel Hill www.jcdl2006.org

Design Perspective

- Interfaces are crucial to successful public websites and intranets
- UI and back end processes (IR) must be closely coupled: Toward HCIR (HCI+IR)
- Iterative analysis and design grounded in the practicalities of large institutions: user needs, prototypes, usability testing, iteration
- Integrating the human and system interaction is the main design challenge: syminforosis—people continuously engaged with meaningful information
- IDL and AgileViews design framework

BLS+ Case: 1995-2005

- Assess User Needs and Influence Design Process: Find what you need; understand what you find
  - Transaction log analyses
  - Query log analysis
  - Email content analysis
  - Focus groups
  - Interviews
  - Prototype developments and testing
    - Laboratory user studies
    - 9 month trial on FedStats with online questionnaire
    - Text mining experiments to generate topical metadata
    - Online help

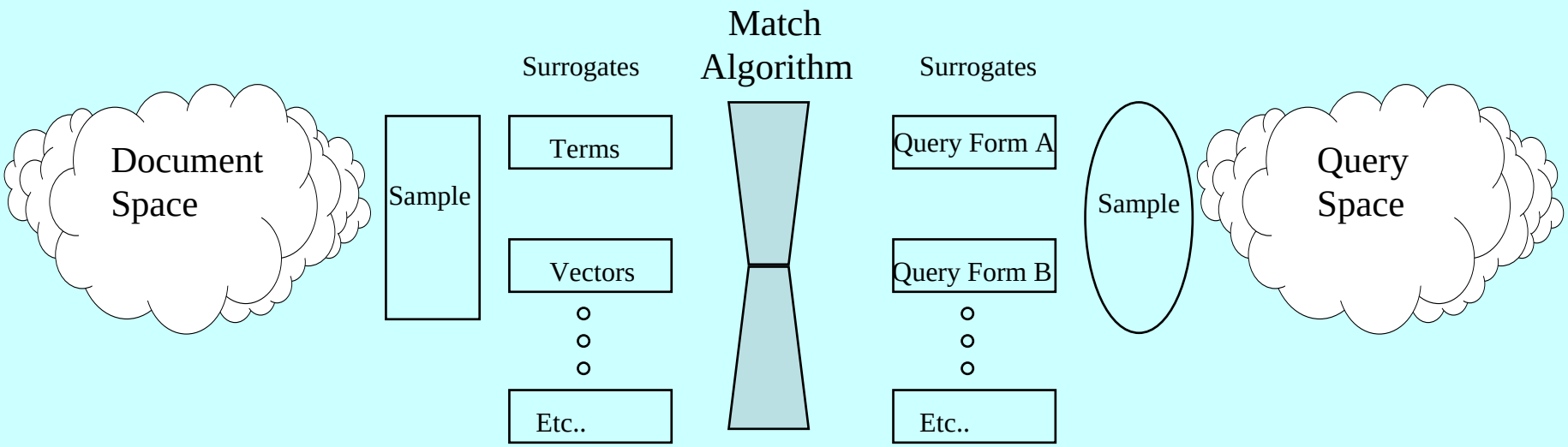
Highlights

- Design and Power
  - Screen real estate wars
  - Branding and symbols
  - Naming and language
  - Technology adoption inertia (HTML—Java—MM XML—AJAX)
- Culture shift
  - User rather than agency driven
  - Changes in the work
    - Broad base of users (universal access, 508, LCD dangers)
    - Email flood (from economists/statisticians to ref librarians)
    - Work flow changes, including security, quality control
  - Changes in information life cycle

Changes in Search and User Interfaces to support Search

Gary Marchionini March 14, 2006

Content-Centered Retrieval as Matching Document Representations to Query Representations



A powerful paradigm that has driven IR R&D for half a century.
Evaluation metric is effectiveness of the match. (e.g., recall and precision).

Content Trend

- Content Features (queries too)
  - Not only text
    - Statistics, images, music, code, streams, biochemical
  - Multimedia, multilingual
  - Dynamic
    - Temporal (e,g., blogs, wikis, sensor streams)
    - Conditional (e.g., computed links, recommendations)
- Content Relationships
  - Hyperlinks, new metadata, aggregations
  - Digital Libraries/sharia, personal collections
- Content acquires history=>context retrieval

Responses to Content Trend

Link analysis

- Multiple sources of evidence (fusion)
  - Authors' words (e.g., full text IR)
  - Indexer/abstractor words (e.g., OPACs)
  - Authors' citations/links (e.g., ISI, Google)
  - Readers' search paths (e.g., recommenders, opinion miners)
  - Social tagging
  - Machine generated features and relationships
- Two key challenges:
  - What new relationships can we leverage (human and machine)?
  - How can we integrate multiple sources of evidence?

Installed User Base Trend

- Technical advances and technical literacy allows us to leverage information seeker intelligence
  - Rather than sole dependence on matching algorithms, focus on flow of representations and actions in situ as people think **with** these new tools and information resources
- Web and TV remotes have legitimized browsing as human-controlled information seeking
- To leverage human intelligence and effort, people must assume responsibilities: beyond the two-word, single query
- Aim at understanding rather than retrieval

Responses to People Trend

- Adapt techniques to WWW
  - Relevance feedback
  - Query expansion
  - User modeling/profiles, SDI services
- Recommender systems
  - Explicit and implicit models
- Social tagging
- Capture everything (e.g., Lifebits)
- Human tuning of IR systems
- User Interfaces
  - Dynamic queries
  - Agile views

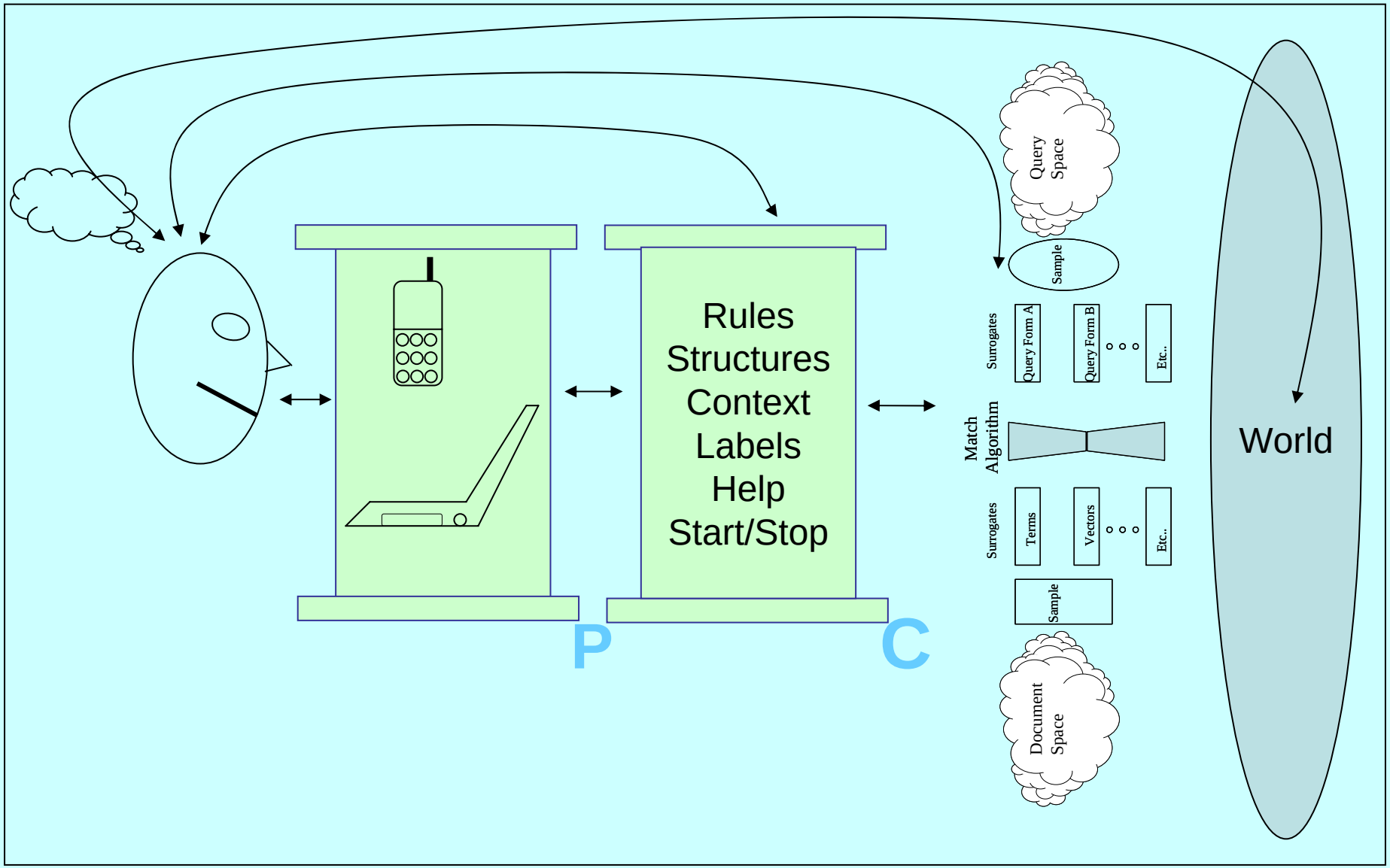
An Expanded II Model:

Think of Information Interaction from the perspective of an active human with information ***needs***, information ***skills***, powerful IR ***resources*** *(that include other humans)*, and situated in global and local connected ***communities***, all of which ***evolve*** over time

HCIR

- Get people closer to the information they need
  - Closer to the backend
  - Closer to the meaning
- Involve IT professionals as integral to the IR system (systems are hybrid)
- Increase responsibility as well as control
- Leverage more demanding and knowledgeable installed base
- Consider ubiquity, digital libraries, e-commerce as extended memories and tools (personal and shared)
- Consider that people often do not start at your home page but enter via referral from search engine

HCIR: Bringing User Closer to World

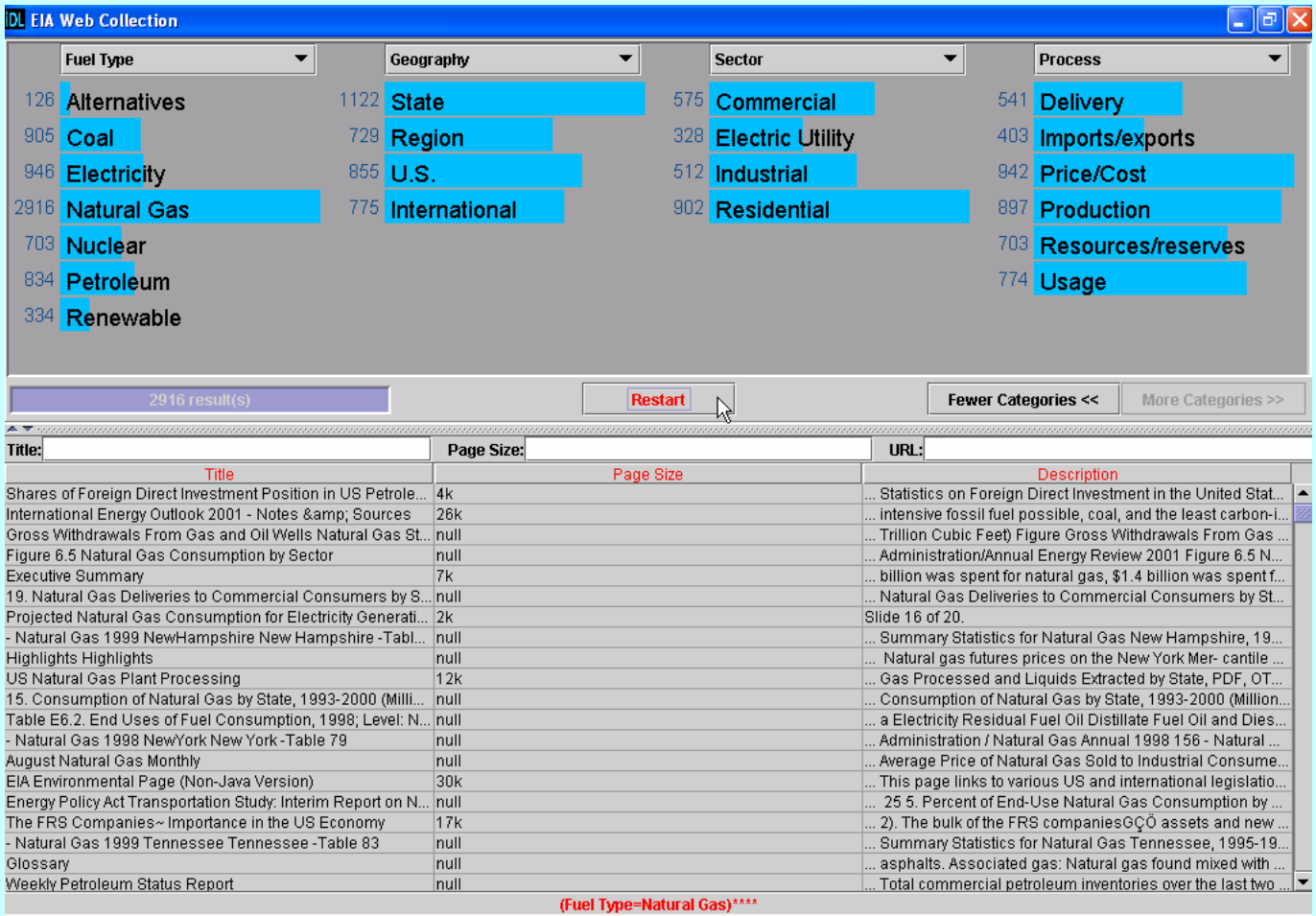


Gary Marchionini March 14, 2006

Key Challenges

- Linking conceptual interface to system backend
  - metadata generation
  - alternative representations and control mechanisms
- Raising user literacy and involvement
  - Engaging without insulting or annoying
- Adding human intelligence to the system
- Moving beyond retrieval to understanding
  - Context

Relation Browser Example with all EIA pages



Gary Marchionini March 14, 2006

RB Goals

- Facilitate exploration of the relationships between (among) different data facets
- Display alternative partitions of the database with mouse actions
- Support string search within partitions
- Serve as an alternative to existing search and navigation tools

Relation Browser Principles

- Architectural Principle: Juxtapose facets
  - Two or more with 5-15 categories per facet
  - Topic is one important facet for most applications
- Interaction Principle: Dynamic exploration of relationships between facets and categories
- Database driven to promote flexible applications (requires systematic metadata)

User Study Results

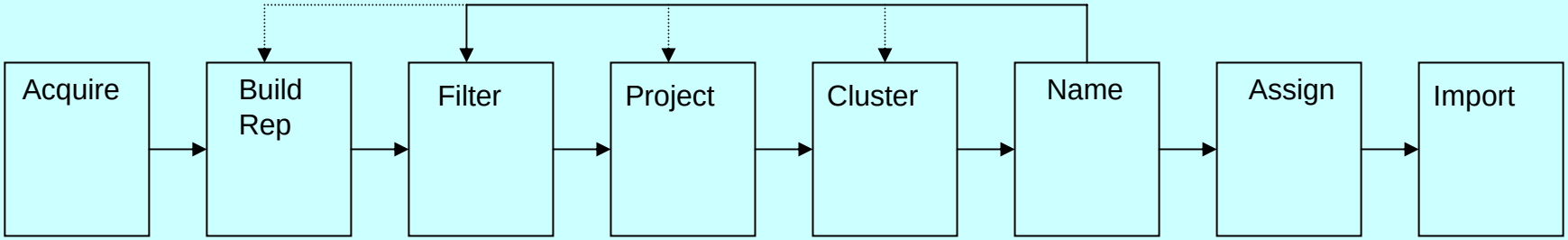
- RB was more effective (less errors) and efficient (shorter time) in completing exploratory tasks than form fill-in interface.
- Users felt more confident and satisfied with RB for exploratory tasks.
- Users felt more satisfied with recommendation task
- Success was mainly due to the interactive category overview and the dynamic keyword searching capability.
  - Interactive category overview helped users to gain insights easily and quickly
  - Dynamic search facilitated narrowing results in a quick and safe way
  - Tightly coupling of search and browse give users more confidence and satisfaction

Key Challenges for RB and similar UIs

- Technical evolutions (Java, metadata to client side)
- User expectations and preparations
- Getting metadata and mapping to RB scheme
  - Given the cost and difficulty with hundreds of thousands of web pages, can we automate this process?

‘Automatic’ classification works best when its application is supported by humans with knowledge of the domain and the techniques at hand. Even then, it tends to be less effective than manual clustering, but may be satisfactory for large-scale indexing.

Behind the RB: Human-Machine Cooperation



Acquire	Build Rep	Filter	Project	Cluster	Name	Assign	Import
Crawl mirror [HTML]	Term/Doc matrix Titles, anchor text, metadata tags	Stop words infrequents	Reduce dimensionality to 50-100 dim PCA LSA ICA	K-means EM Yields prob model	Human effort Frequencies Log-odds	Cataloging (binning) based on model	Pipe to RB Add other facets

A Metadata Mining Toolkit is Available

www.ils.unc.edu/govstat/demos.html

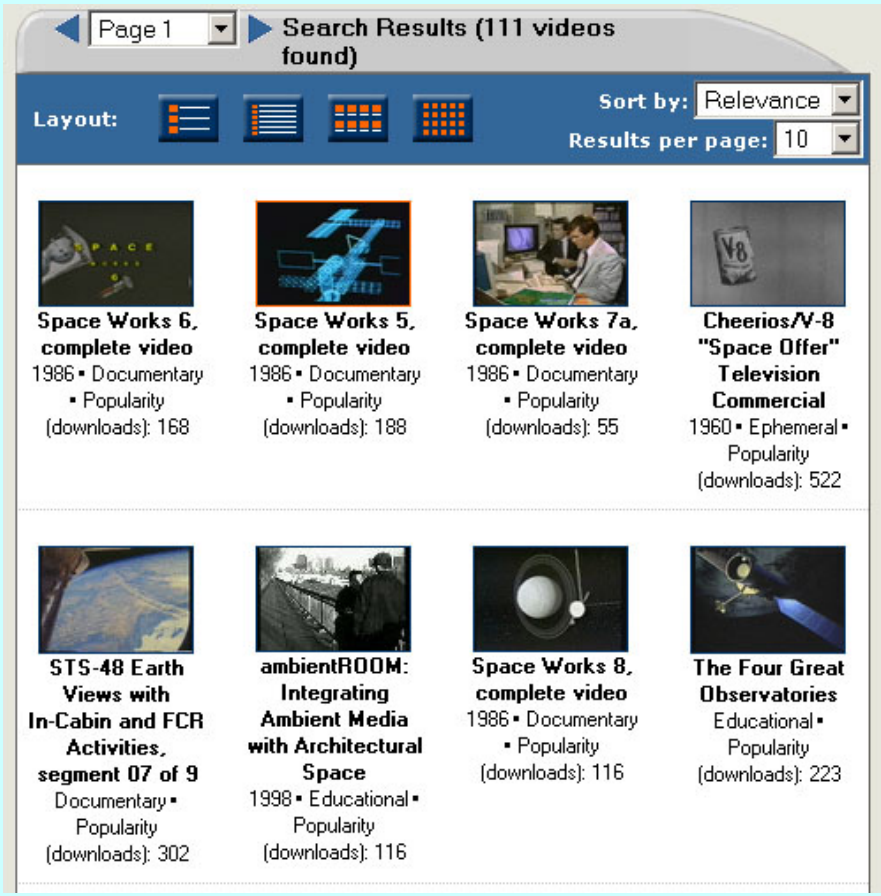
Text Mining/Clustering Experiences

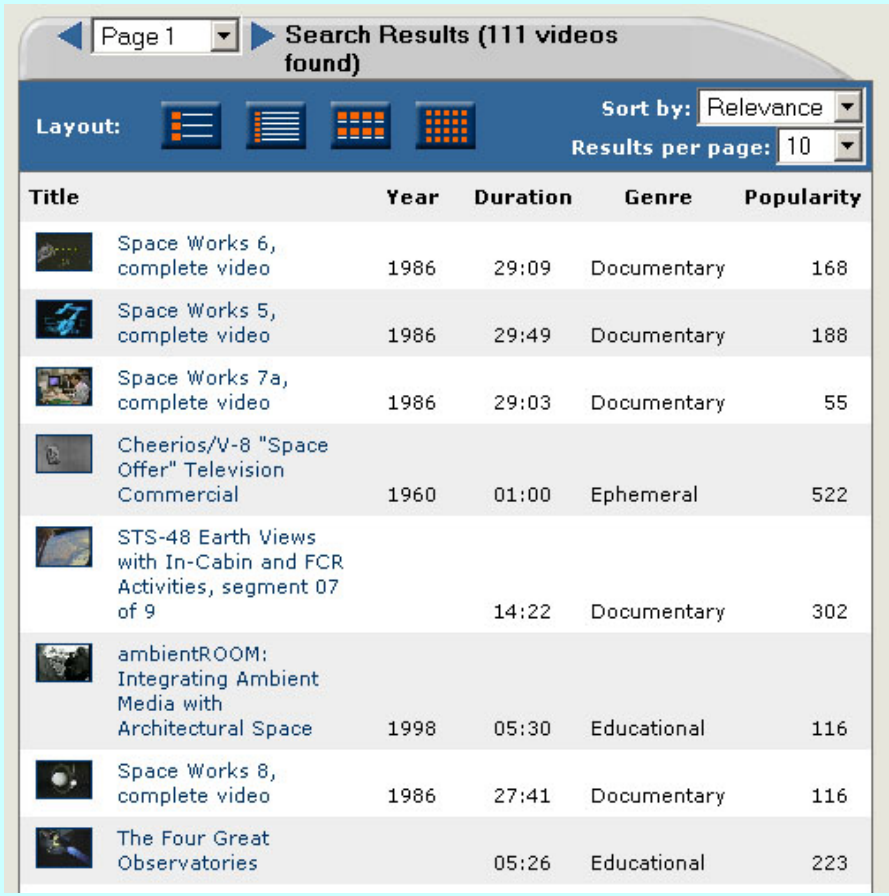
- Anchor text helps a lot
  - Metadata can also help
  - Labeling matters a lot
- Specific partitions of data strong influence
- System tuning and human intervention
- Challenges
  - Acronyms
  - Link pages
  - Template pages (e.g., same report periodically)
  - Page frames
  - Term usage in rows, columns, and titles of tables may each have different functions
- Need for filters and weighting schemes
- Information Alchemy: if you have 14K gold, you can improve it, but you can't turn lead into gold! GIGO

Open Video Example
www.open-video.org

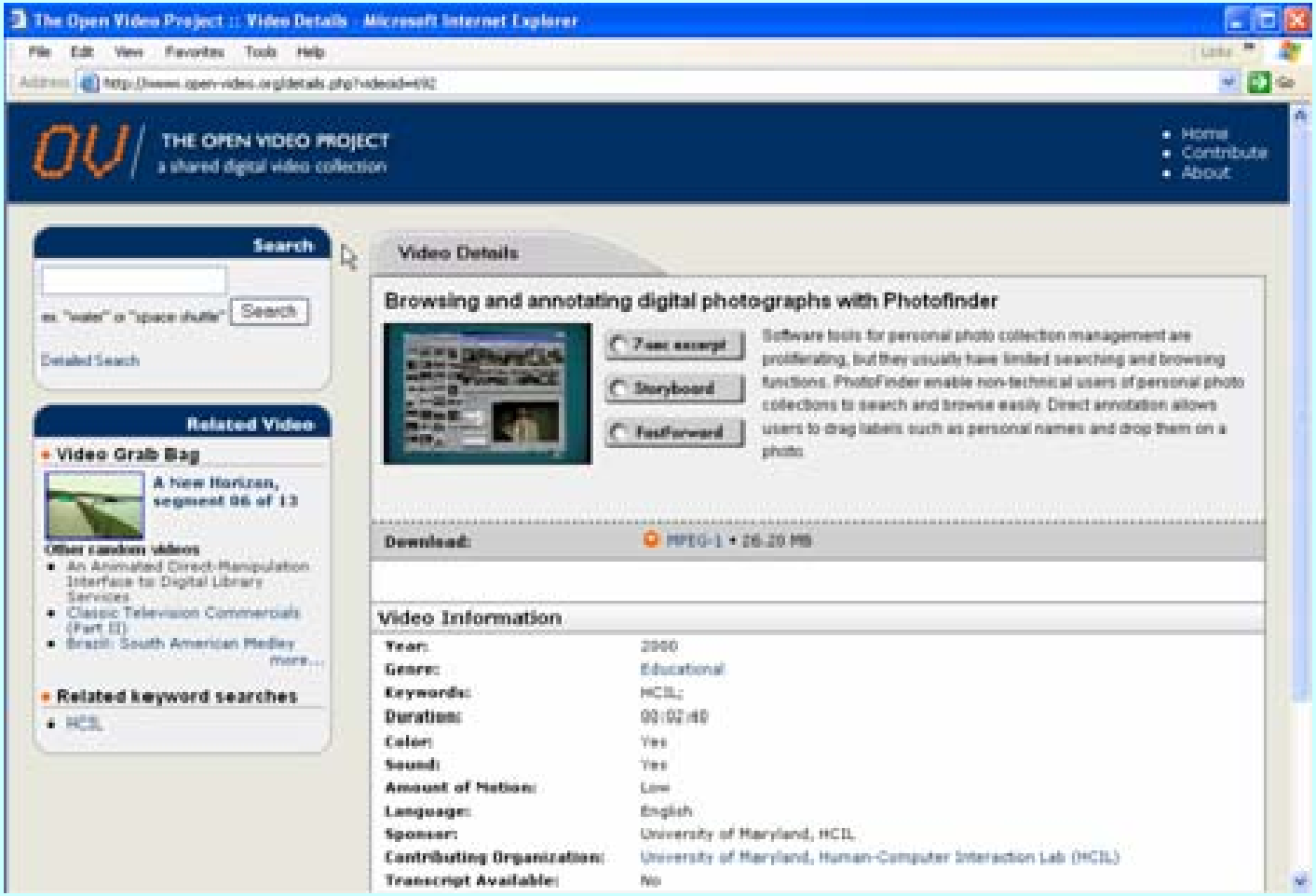
- Open access digital library of digital video for education and research
- 2500+ video segments: MPEG1, MPEG-2, MPEG-4, QuickTime
- Multiple visual surrogates
- Agile Views Design Framework
  - Different types of views
    - Overviews, previews, shared views
  - Multiple examples of views
  - Dynamic control mechanisms

Alternative Overviews of Result Sets





Alternative Previews for a Specific Video Segment



Gary Marchionini March 14, 2006

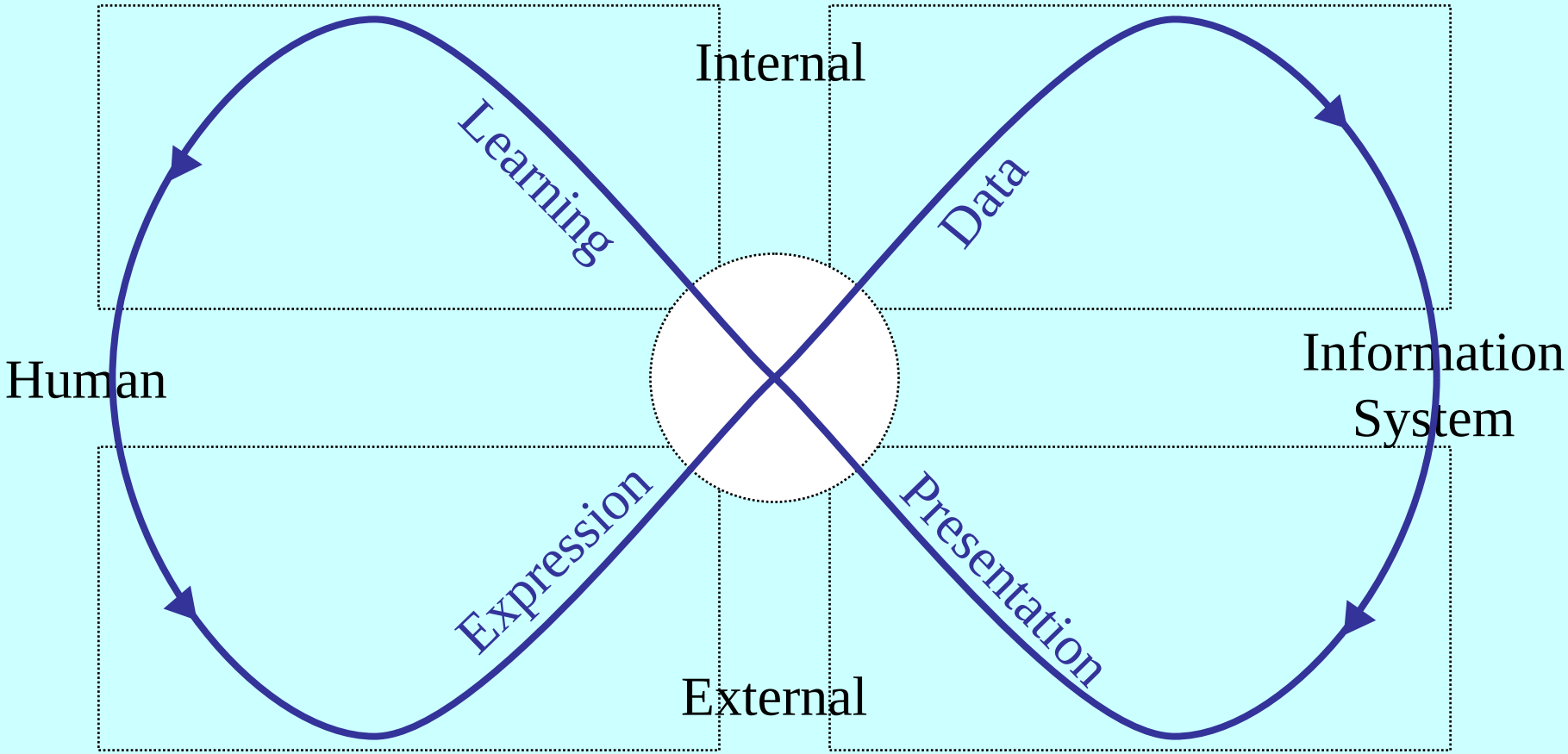
Some Interaction Principles and Caveats in These Examples

- Principles
  - Look ahead without penalty
  - Minimize scrolling and clicking
  - Alternative ways to slice and dice
  - Closely couple search, browse, and examine
  - Continuous engagement—useful attractors
  - Treasures to surface
- Caveats
  - Scalability (getting metadata to client side)
  - Metadata crucial
    - We are working on automatically creating partitions
  - Increasing expectations about useful results (answers!)

Recommendations

- Implement iterative schedules with reasonable milestones
- Monitor the installed user base
  - Users, potential users, larger global culture
- Monitor technology trends but do not let technology drive
  - HTML→XML/XSLT→AJAX→???
  - Form factors (IPods, PDAs, cell phones, wall screens)
  - Interaction styles
  - Data structures and algorithms
- Do not underestimate economic and political factors

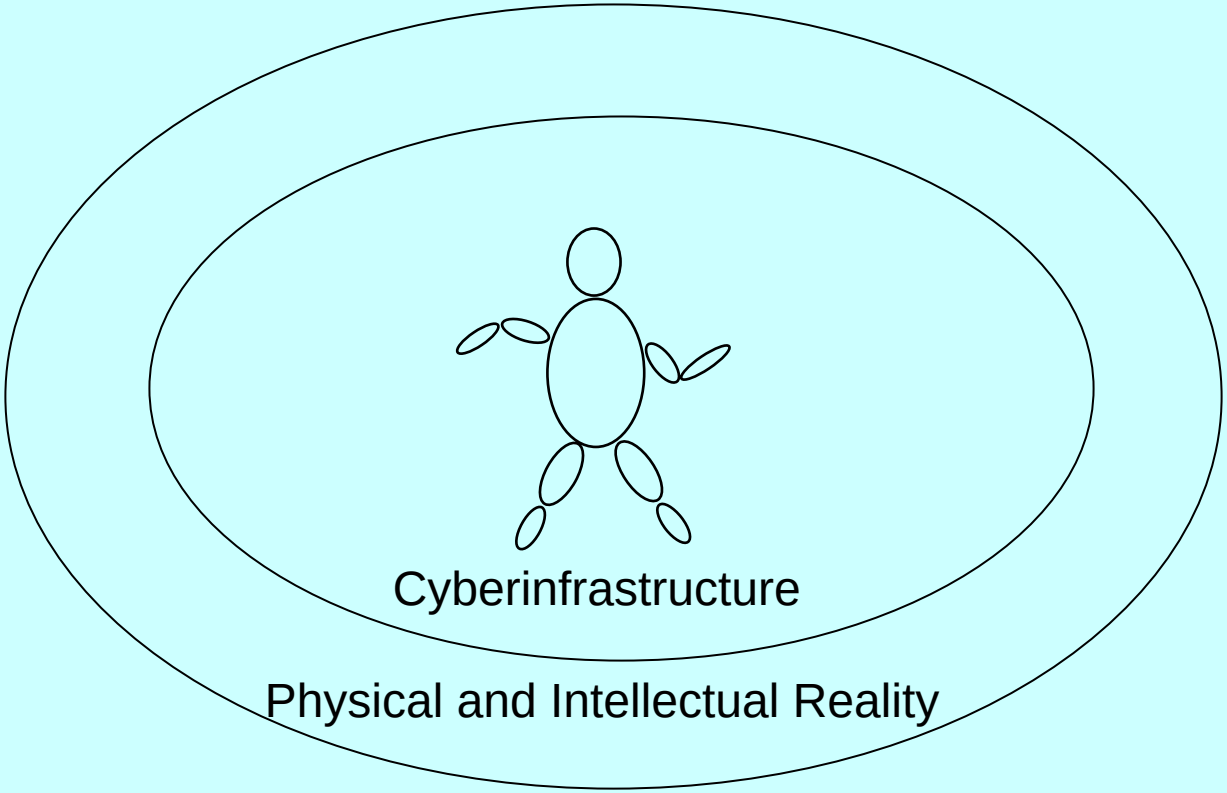
Human-Information Ecology Model



Gary Marchionini March 14, 2006

Long Term Paradigm: Information Interaction as Core Life Process

RB represents one early way to get the information seeker more involved in the information seeking process—there is plenty more to do. Like eating we have varying expectations, invest different levels of effort, and use diverse and ubiquitous infrastructures. Key challenge is to span boundaries between cyberinfrastructure and the 'real' world.



Gary Marchionini March 14, 2006

Thank You!

Questions and Discussion

march@ils.unc.edu

www.ils.unc.edu/govstat

www.open-video.org

NSF Grants EIA 0131824 and IIS 0099638