

Beyond Basic Search

Gary Marchionini

University of North Carolina at Chapel Hill

march@ils.unc.edu

CS Colloquium

Computer Science Department

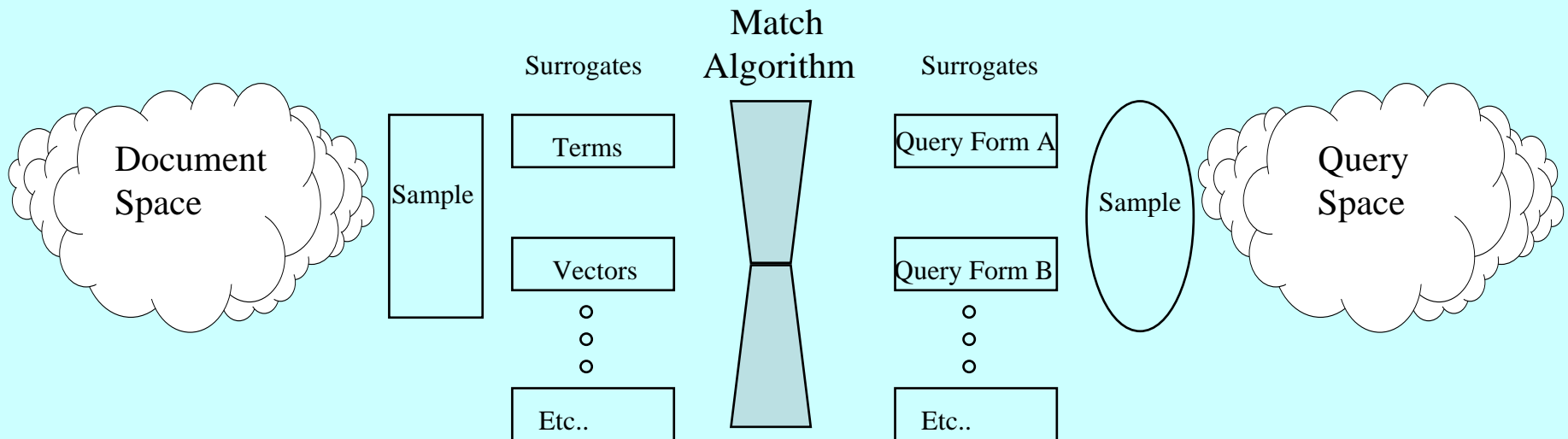
University of Illinois at Urbana-Champaign

September 10, 2007

Outline

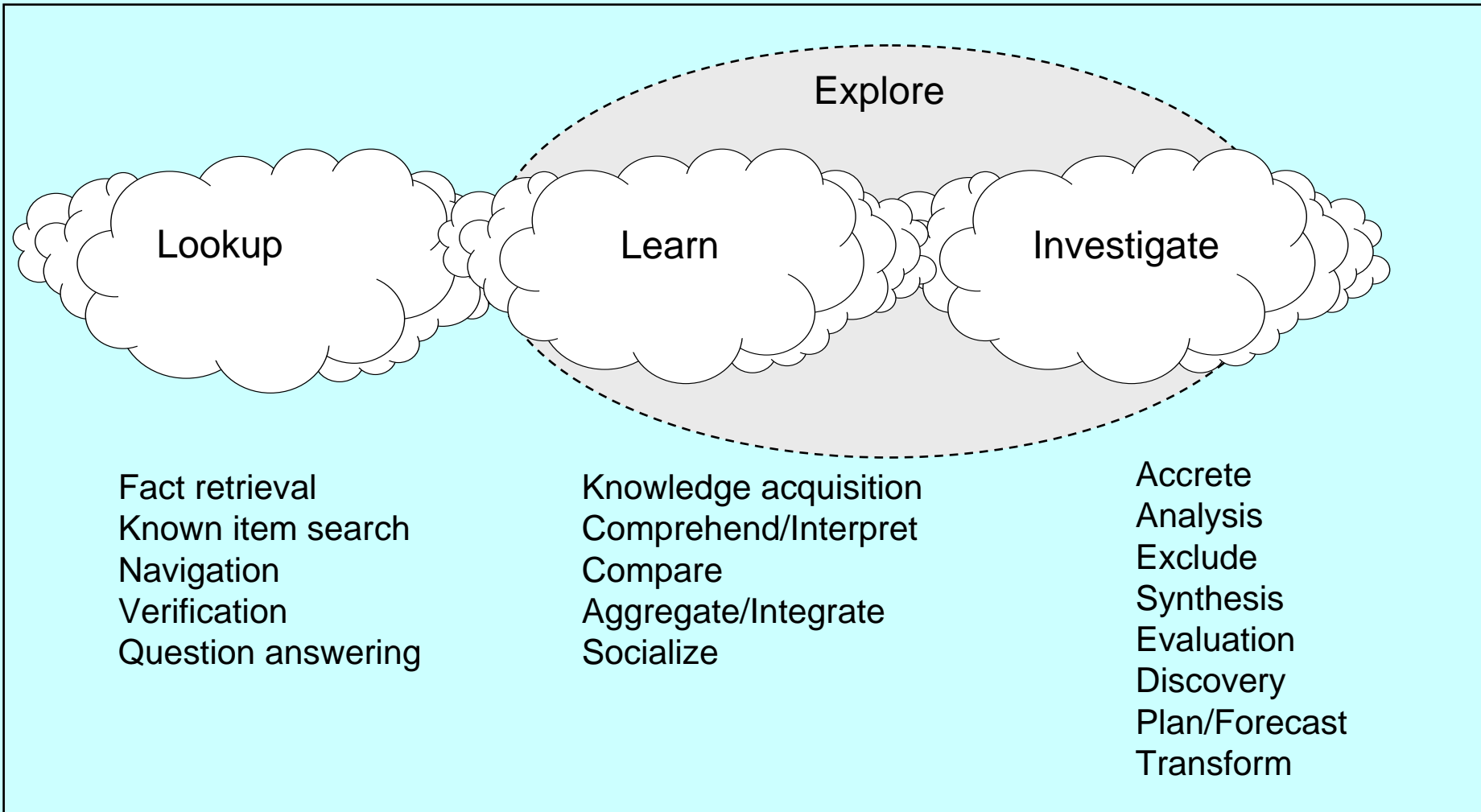
- Information retrieval R&D has stimulated a multi-billion dollar industry
- The challenges of information seeking and exploratory search that get searcher(s) more actively involved
- Some early examples: faceted search, dynamic queries and agile views
- Evaluation challenges

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).
A half duplex process that is strongly dependent on pre-processing

Information Seeking Goals: Focus on Exploratory Search



Characteristics of the Exploratory Search Process

- Multiple sessions
- Multiple queries
- Recall important
- Collaborative
- Substantial time spent in results
- Coordinate with other tools
- Relevance judging more difficult (subjective, domain dependent, nuanced)
- Objects of interest tend toward abstract and complex

Dynamic Queries

- Direct manipulation (e.g., slider move, hover) defines and executes query with immediate feedback (see <http://www.cs.umd.edu/hcil/spotfire/> for history)
- An alternative to text query forms

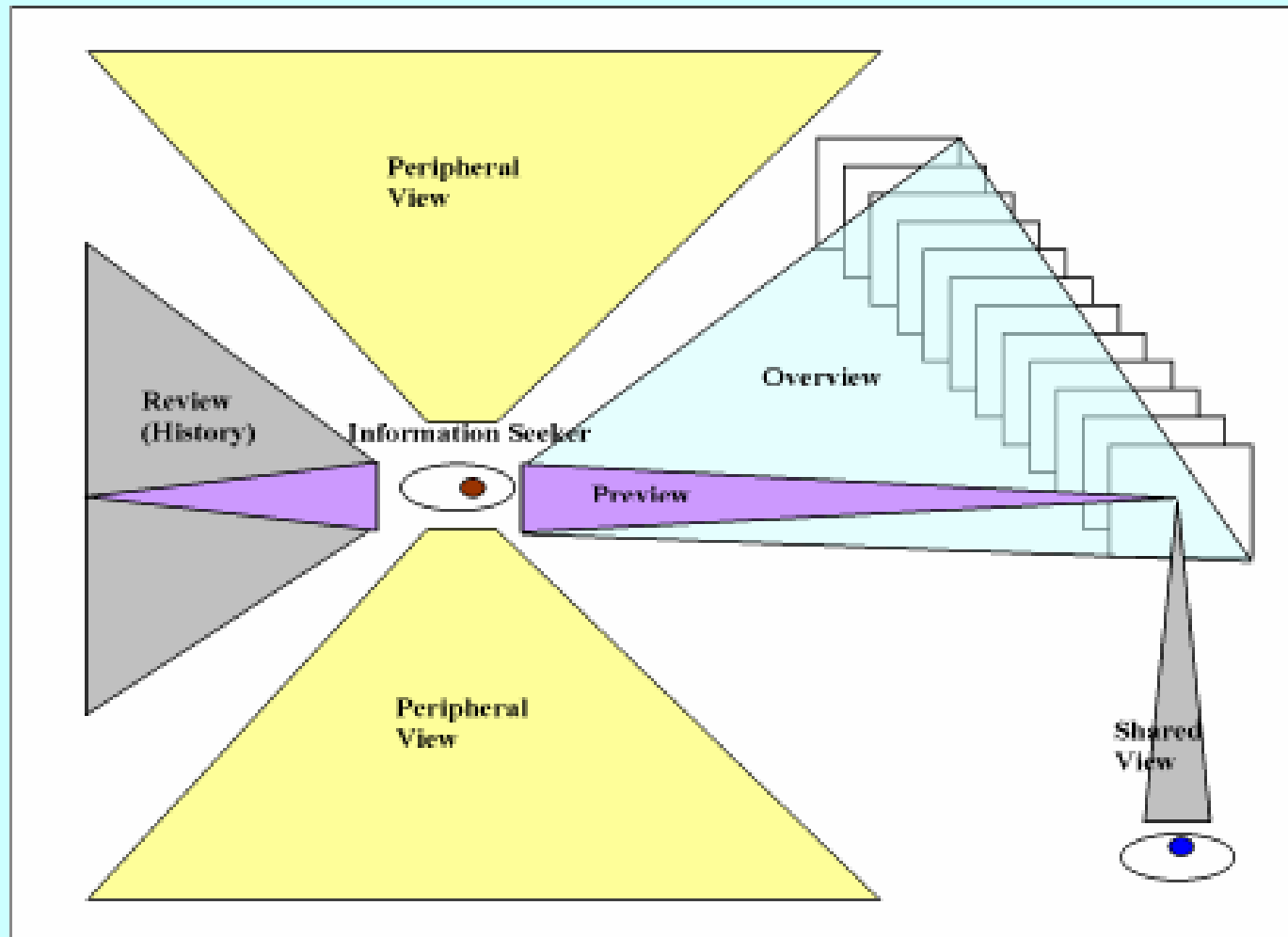
Faceted Search

- Combine text search with category selection
- Many E-commerce sites
- Metadata critical (database backends loved)
- Examples:
 - Flamenco <http://flamenco.berkeley.edu/>
 - mSpace <http://www.mspace.fm/>
 - Endeca <http://endeca.com/> also see <http://www.lib.ncsu.edu/endeca/>
 - Relation Browser <http://idl.ils.unc.edu/rave>

AgileViews

- A view is a partition of an information space
 - There are many possible partitions for any space since many attributes may be used to ‘slice and dice’ the space
- People should be able to **effortlessly** change views
 - Focus can change based on granularity
 - Focus can change based on attribute

AgileViews Framework



Gary Marchionini, UNC-Chapel Hill

What are we trying to support and evaluate?

Active humans 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

Relation Browser Example with all EIA pages [RB demo here]

EIA Web Collection

Fuel Type: **Natural Gas** (2916 result(s))

Geography: **State** (1122), **Region** (729), **U.S.** (855), **International** (775)

Sector: **Commercial** (575), **Electric Utility** (328), **Industrial** (512), **Residential** (902)

Process: **Delivery** (541), **Imports/exports** (403), **Price/Cost** (942), **Production** (897), **Resources/reserves** (703), **Usage** (774)

Title	Page Size	URL
Shares of Foreign Direct Investment Position in US Petrole...	4k	... Statistics on Foreign Direct Investment in the United Stat...
International Energy Outlook 2001 - Notes & Sources	26k	... intensive fossil fuel possible, coal, and the least carbon-l...
Gross Withdrawals From Gas and Oil Wells Natural Gas St...	null	... Trillion Cubic Feet) Figure Gross Withdrawals From Gas ...
Figure 6.5 Natural Gas Consumption by Sector	null	... Administration/Annual Energy Review 2001 Figure 6.5 N...
Executive Summary	7k	... billion was spent for natural gas, \$1.4 billion was spent f...
19. Natural Gas Deliveries to Commercial Consumers by S...	null	... Natural Gas Deliveries to Commercial Consumers by St...
Projected Natural Gas Consumption for Electricity Generati...	2k	Slide 16 of 20.
- Natural Gas 1999 NewHampshire New Hampshire - Tabl...	null	... Summary Statistics for Natural Gas New Hampshire, 19...
Highlights Highlights	null	... Natural gas futures prices on the New York Mer- cantile ...
US Natural Gas Plant Processing	12k	... Gas Processed and Liquids Extracted by State, PDF, OT...
15. Consumption of Natural Gas by State, 1993-2000 (Milli...	null	... Consumption of Natural Gas by State, 1993-2000 (Million...
Table E6.2. End Uses of Fuel Consumption, 1998; Level: N...	null	... a Electricity Residual Fuel Oil Distillate Fuel Oil and Dies...
- Natural Gas 1998 NewYork New York -Table 79	null	... Administration / Natural Gas Annual 1998 156 - Natural ...
August Natural Gas Monthly	null	... Average Price of Natural Gas Sold to Industrial Consume...
EIA Environmental Page (Non-Java Version)	30k	... This page links to various US and international legislatio...
Energy Policy Act Transportation Study: Interim Report on N...	null	... 25 5. Percent of End-Use Natural Gas Consumption by ...
The FRS Companies~ Importance in the US Economy	17k	... 2). The bulk of the FRS companiesGÇÖ assets and new ...
- Natural Gas 1999 Tennessee Tennessee -Table 83	null	... Summary Statistics for Natural Gas Tennessee, 1995-19...
Glossary	null	... asphalts. Associated gas: Natural gas found mixed with ...
Weekly Petroleum Status Report	null	... Total commercial petroleum inventories over the last two ...

(Fuel Type=Natural Gas)****

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)

Key Challenges

- 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?

Recent User Studies

- Facets (ACM/IEEE JCDL 2007; DL 05):
 - Relation Browser, Vanilla Facet, BLS website
 - Known item and exploratory tasks in BLS data
 - Between and within subject designs
 - Results
 - NSR differences
 - familiarity influences expectations—installed base syndrome
 - Automatically generated categorization comparable to carefully crafted website layouts

Open Video Example

www.open-video.org

- Open access digital library of digital video for education and research
- 4000+ video segments: MPEG1, MPEG-2, MPEG-4, QuickTime
- Multiple visual surrogates
- Agile Views Design Framework
 - Facet partitions (collections, genre, length, etc.)
 - Different types of views
 - Overviews, previews, shared views
 - Multiple examples of views: Surrogates as previews (textual metadata, storyboard, except, fast forward, spoken descriptions/keywords)
 - Dynamic control mechanisms
- Basic search (MySQL indexes)

Alternative Previews for a Specific Video Segment [OV demo here]

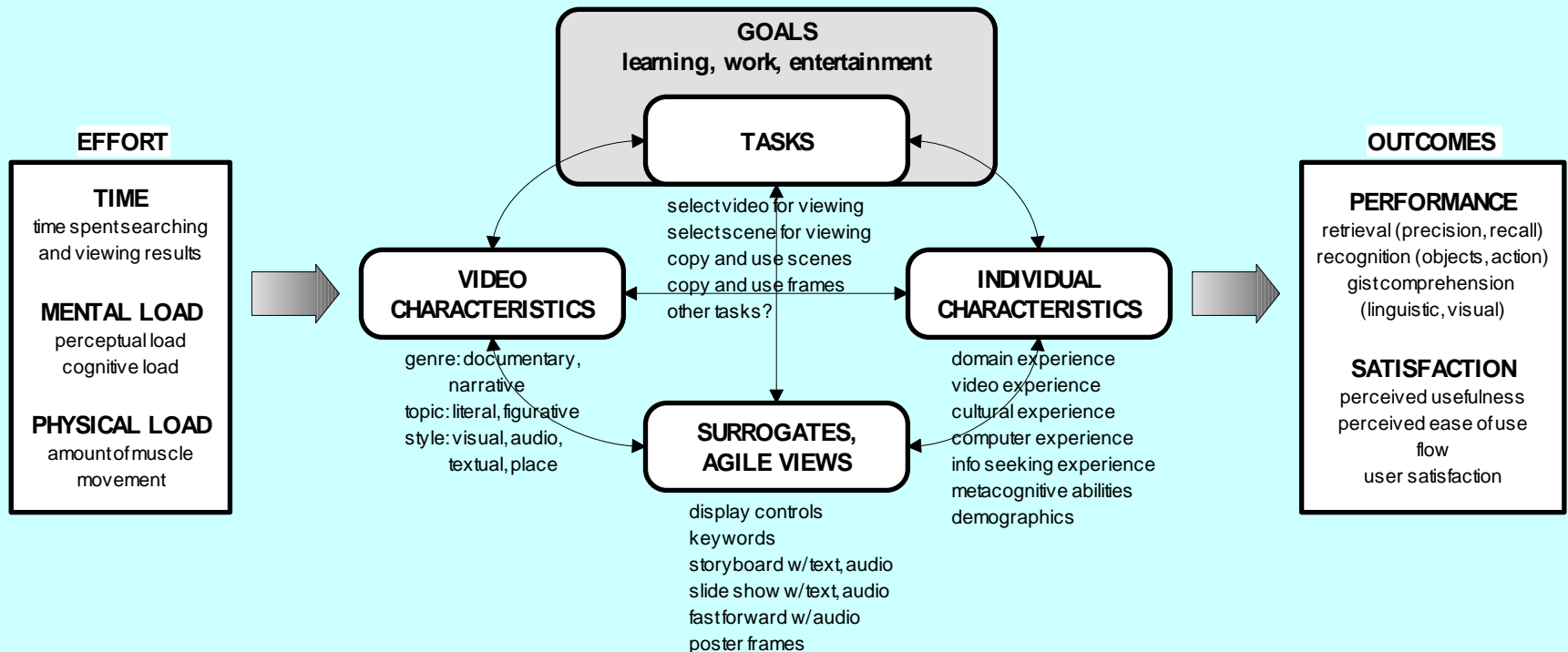
The screenshot shows a Microsoft Internet Explorer browser window displaying the Open Video Project website. The address bar shows the URL: <http://www.open-vids.org/details.php?video=692>. The website header features the 'OV' logo and the text 'THE OPEN VIDEO PROJECT a shared digital video collection'. Navigation links for 'Home', 'Contribute', and 'About' are visible in the top right.

The main content area is divided into several sections:

- Search:** A search box with a 'Search' button and a 'Detailed Search' link below it.
- Related Video:** A section titled 'Video Grab Bag' featuring a video thumbnail and the text 'A New Horizon, segment 06 of 13'. Below this are 'Other random videos' and 'Related keyword searches'.
- Video Details:** The main content area, titled 'Browsing and annotating digital photographs with Photofinder'. It includes a video player thumbnail, three buttons ('Fast except', 'Storyboard', 'Fastforward'), and a description: 'Software tools for personal photo collection management are proliferating, but they usually have limited searching and browsing functions. PhotoFinder enable non-technical users of personal photo collections to search and browse easily. Direct annotation allows users to drag labels, such as personal names and drop them on a photo.'
- Download:** A section showing 'MP4G-1 • 26.20 MB'.
- Video Information:** A table providing metadata for the video.

Video Information	
Year:	2000
Genre:	Educational
Keywords:	HCI;
Duration:	00:02:40
Color:	Yes
Sound:	Yes
Amount of Motion:	Low
Language:	English
Sponsor:	University of Maryland, HCI
Contributing Organization:	University of Maryland, Human-Computer Interaction Lab (HCI)
Transcript Available:	No

User Study Framework



Video Surrogate Studies

- A dozen studies over 6 years (ACM CHI 07, JCDL 04, ASIST, AVI, MM 06, others)
 - Story boards, slide shows, fast forwards, excerpts, spoken keywords, spoken descriptions, combinations
 - Multiple tasks (gist, vist)
 - Multiple measures (accuracy, time, satisfaction)
 - Within, between, ethnographic
- Results
 - Words matter
 - Visual adds value (conceptual and affective)
 - People able to infer from few cues, tolerate high rates
 - Coordination of multiple channels?

Thank You!

Questions and Discussion
march@ils.unc.edu