

Toward Human-Computer Information Retrieval

Gary Marchionini
University of North Carolina at Chapel Hill
march@ils.unc.edu

Samuel Lazerow Memorial Lecture
The Information School
University of Washington

October 17, 2005

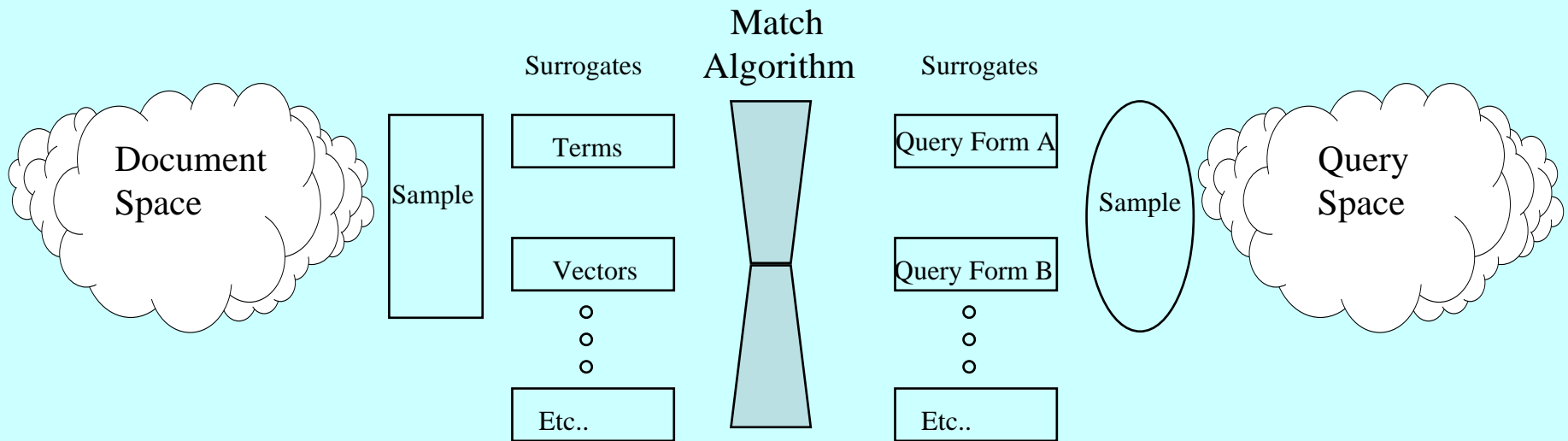
Outline

- HCIR motivation and vision
- Example 1: Relation Browser
- Example 2: Open Video
- Reflections and Implications

Message

- IR and HCI are related fields that have strong (staid?) traditions that have been energized (jolted?) by WWW.
- The intersection of these fields offers interesting new opportunities for high-impact IR R&D, especially in WebIR and DLs.
- Integrating the human and system interaction is the main design challenge: syminforosis—people continuously engaged with information

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).**

Things Change: 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)
 - 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?

Things Change: User Base Trend

- 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
- People want understanding rather than retrieval
- 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 (NetGen)

Responses to People Trend

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

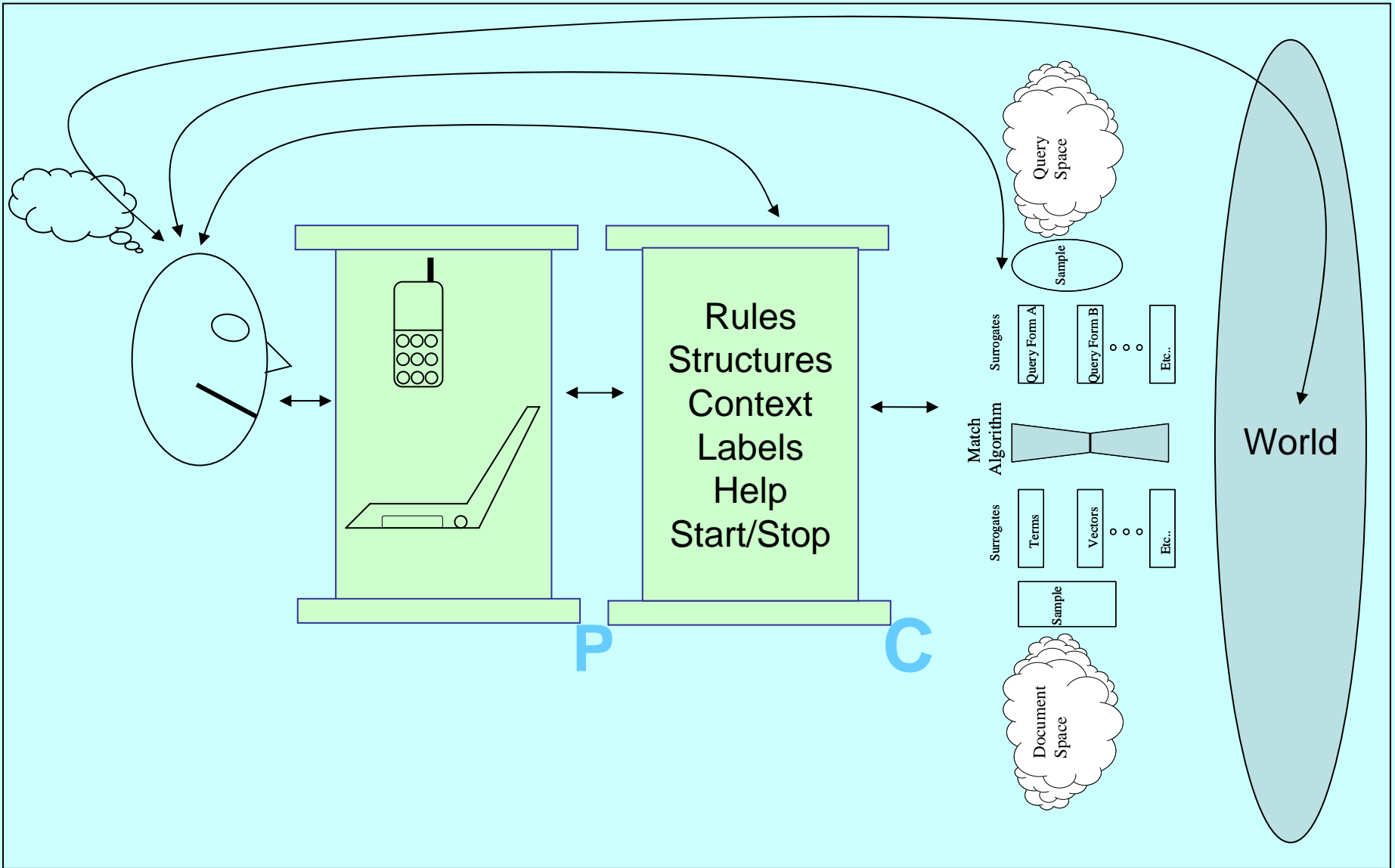
HCIR Model:

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

HCIR Features

- Get people closer to the information they need
 - Closer to the meaning
- End user increase in 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 information life cycle (from creation to preservation)
- Involve information professionals as integral to the IR system; importance of tuning

HCIR: Bringing User Closer to World



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
- New Evaluation Measures

Relation Browser Example with all EIA pages (Demo here)

EIA Web Collection

Fuel Type: Alternatives (126), Coal (905), Electricity (946), Natural Gas (2916), Nuclear (703), Petroleum (834), Renewable (334)

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)

2916 result(s) [Restart] Fewer Categories << More Categories >>

Title	Page Size	URL
Shares of Foreign Direct Investment Position in US Petroleum...	4k	... Statistics on Foreign Direct Investment in the United States...
International Energy Outlook 2001 - Notes & Sources	26k	... intensive fossil fuel possible, coal, and the least carbon-intensive...
Gross Withdrawals From Gas and Oil Wells Natural Gas Statistics...	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 Natural...
Executive Summary	7k	... billion was spent for natural gas, \$1.4 billion was spent for...
19. Natural Gas Deliveries to Commercial Consumers by State...	null	... Natural Gas Deliveries to Commercial Consumers by State...
Projected Natural Gas Consumption for Electricity Generation...	2k	Slide 16 of 20.
- Natural Gas 1999 NewHampshire New Hampshire - Table...	null	... Summary Statistics for Natural Gas New Hampshire, 19...
Highlights Highlights	null	... Natural gas futures prices on the New York Mercantile...
US Natural Gas Plant Processing	12k	... Gas Processed and Liquids Extracted by State, PDF, OT...
15. Consumption of Natural Gas by State, 1993-2000 (Million...	null	... Consumption of Natural Gas by State, 1993-2000 (Million...
Table E6.2. End Uses of Fuel Consumption, 1998; Level: National...	null	... a Electricity Residual Fuel Oil Distillate Fuel Oil and Diesel...
- Natural Gas 1998 NewYork New York - Table 79	null	... Administration / Natural Gas Annual 1998 156 - Natural Gas...
August Natural Gas Monthly	null	... Average Price of Natural Gas Sold to Industrial Consumers...
EIA Environmental Page (Non-Java Version)	30k	... This page links to various US and international legislation...
Energy Policy Act Transportation Study: Interim Report on Natural...	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 companies' G&O 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

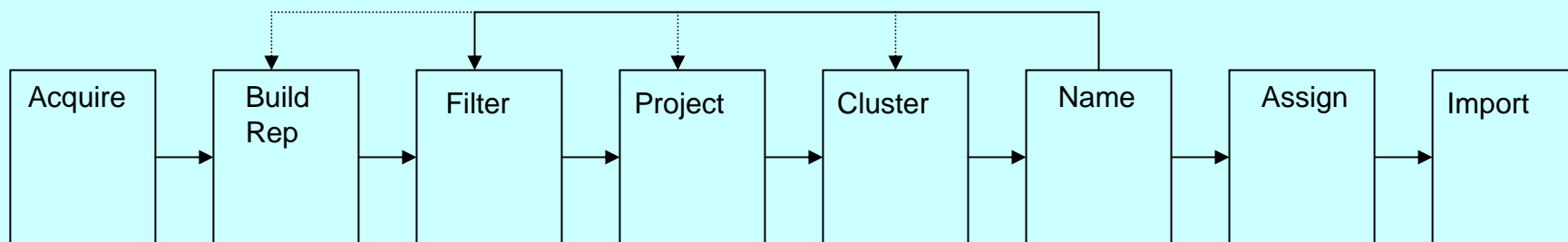
Relation Browser Principles

- 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 Limitations

- 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?

Behind the RB: Human-Machine Cooperation



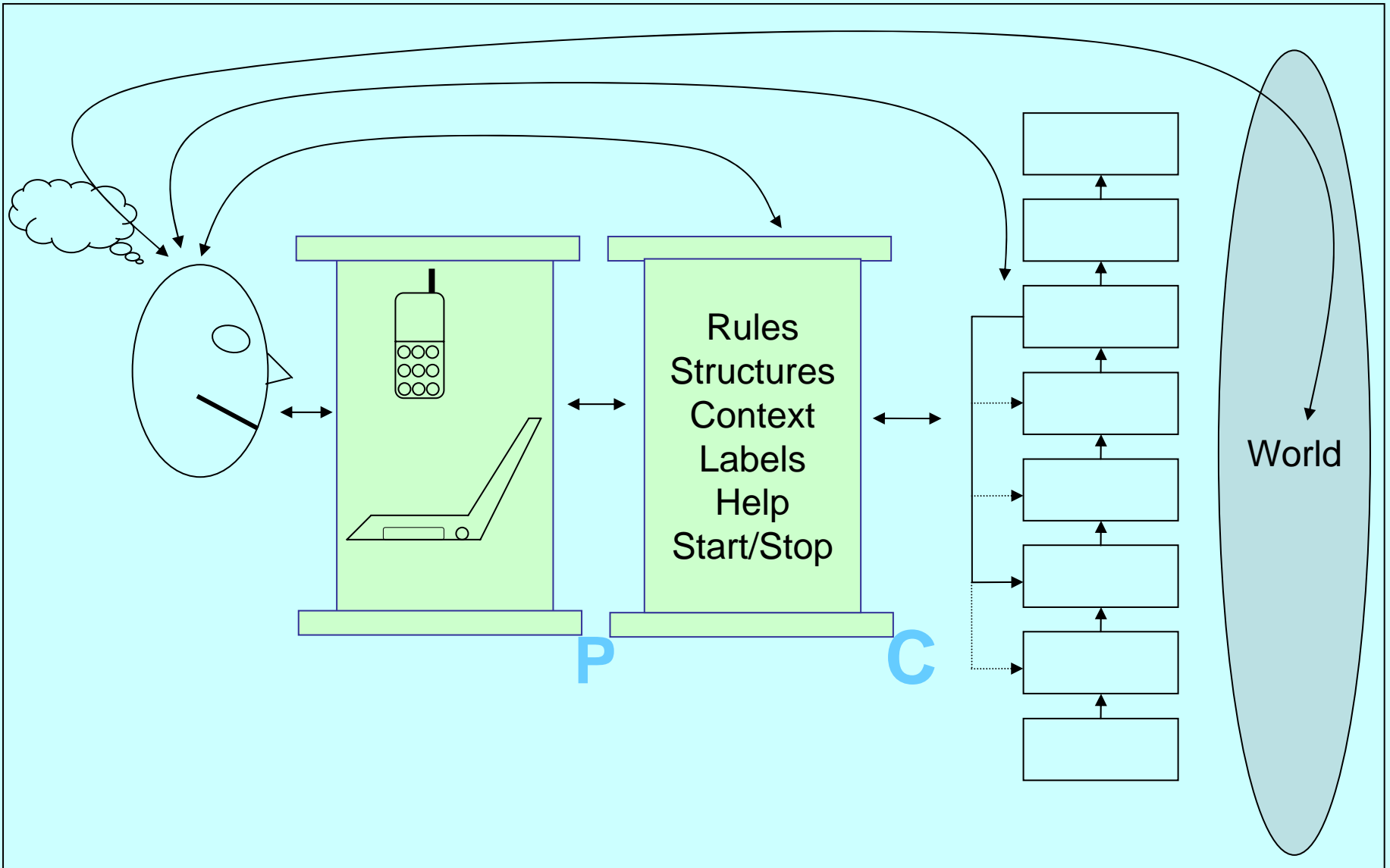
Crawl mirror [HTML]	Term/Doc matrix Titles, anchor text, metadata tags	Stop words infrequents	Reduce dimensionality to 50-100 dim PCA LSA ICA	K-means EM Yields prob model	Human effort Frequencies Log-odds	Cataloging (binning) based on model	Pipe to RB Add other facets
---------------------------	---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

A Metadata Mining Toolkit is Available

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

‘Automatic’ classification works best when its application is supported by humans with knowledge of the domain and the techniques at hand.

RB is Embedded in Larger Process of Information Seeking



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

Open Video Examples (Demo Here)

- Agile Views
- User-controlled slicing and dicing
- A beginning....

Search

ex. "water" or "space shuttle"

[Detailed Search](#)

Browse

Genre

- Documentary [494]
- Educational [668]
- Ephemeral [1139]
- Historical [187]
- Lecture [33]
- Other [5]
- Public Service [17]

Duration

- Less than 1 minute [248]
- 1 to 2 minutes [319]
- 2 to 5 minutes [567]
- 5 to 10 minutes [401]
- More than 10 minutes [989]

Color

- In color [1504]
- In black & white [1039]

Sound

- With sound [2158]
- Silent [385]

Collections

- University of Maryland HCIL Open House Video Reports
- The Informedia Project at Carnegie Mellon University
- Internet Moving Images Archive
- 2001 TREC Video Retrieval Test Collection
- CHI Video Retrospective
- Digital Himalaya Project
- NASA K-16 Science Education Programs

Special Collection Spotlight



Digital Himalaya Project

Digital Himalaya is a pilot project to develop digital collection, archiving and distribution strategies for multimedia anthropological information from the Himalayan region.

This collection contains 34 videos.

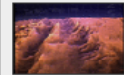
[other special collections...](#)

Project News

- Teachers using video?
- CHI videos added
- Redesigned Project Web Site
- Understanding Video Symposium [more...](#)

Featured Video

New



NASAWhy?Files - The Red Planet

Other new videos

- NASAWhy?Files - The Case of the Inhabitable Habitat
- NASAWhy?Files - The Case of the Electrical Mystery
- NASAWhy?Files - Terrarium [more...](#)

Popular



Atom Bomb - Joe Bonica's Movie of the Month

Other popular videos

- Classic Television Commercials (Part II)
- NASA 25th Anniversary Show, Segment 01
- Bunjee [more...](#)

- Home
- Contribute
- About

Modify Search

Search: All fields
for salton

Genre: - Any Genre -
Duration: - Any Duration -
Format: - Any Format -


Color: Color B&W Either
Sound: Sound Silent Either


Detailed Search


Page 1 Search Results (16 videos found)


Layout: 


Sort by: Relevance
Results per page: 10

- 

Salton Honored - A SMART Celebration: Professor Gerard Salton (1995)
Talk by professor Gerard Salton about interesting issues in information retrieval.
Genre: Lecture
Keywords: Information Retrieval; IR; SMART; Cornell; 30 Years; Computer Science; Gerard Salton
Duration: 15:07
Popularity (downloads): 222
- 

Salton Honored - A SMART Celebration: Wrapup (1995)
Robert Constable closed the conference by acknowledgement to Gerard Salton for the accomplishment of 30 years of IR through SMART at the Department of Computer Science, Cornell University.
Genre: Lecture
Keywords: Information Retrieval; IR; SMART; Cornell; Gerard Salton; Robert Constable
Duration: 00:56
Popularity (downloads): 139
- 

Salton Honored - A SMART Celebration: General Discussion (1995)
Chris Buckley led the general discussion about some of the problems and issues people need to think about when thinking of the future of information retrieval.
Genre: Lecture
Keywords: Information Retrieval; IR; SMART; Cornell; Gerard Salton; Chris Buckley; Future of Information Retrieval
Duration: 28:35
Popularity (downloads): 143
- 

Salton Honored - A SMART Celebration: Tape #3 Full (1995)
On April 22, 1995-01-01, information retrieval researchers from Europe and the United States gathered to honor Gerard Salton and the achievements of the SMART group over the past 30 years...
Genre: Lecture
Keywords: Information Retrieval; IR; SMART; Cornell; Gerard Salton; Chris Buckley; Robert Constable; Future of Information Retrieval
Duration: 29:58
Popularity (downloads): 70
- 

Salton Honored - A SMART Celebration: IR other than Statistics? (1995)
The second speaker, William Cooper, gave the talk - Can IR founded on any science other than statistics?
Genre: Lecture
Keywords: Information Retrieval; IR; SMART; Cornell; 30 Years; Computer Science; William Cooper; Science; Statistics
Duration: 30:47
Popularity (downloads): 175

Modify Search

Search: All fields

for salton

Genre: - Any Genre -

Duration: - Any Duration -

Format: - Any Format -

Color: Color B&W EitherSound: Sound Silent Either

Search

Detailed Search
















Page 1

Search Results (16 videos found)

Layout

Sort by: Popularity

Results per page: 20

Title	Year	Duration	Genre	Popularity
 Salton Honored - A SMART Celebration: A Unified Model for IR?	1995	40:58	Lecture	756
 The Story of Hoover Dam, segment 02 of 12	1996	02:46	Documentary	405
 Salton Honored - A SMART Celebration: Professor Gerard Salton	1995	15:07	Lecture	222
 Salton Honored - A SMART Celebration: IR other than Statistics?	1995	30:47	Lecture	175
 The Miracle of Water, segment 05 of 11	1996	01:23	Documentary	152
 Salton Honored - A SMART Celebration: General Discussion	1995	28:35	Lecture	143
 Salton Honored - A SMART Celebration: Wrapup	1995	00:56	Lecture	139
 Salton Honored - A SMART Celebration: Research Issues for Industry	1995	35:31	Lecture	121
 Salton Honored - A SMART Celebration: Welcome and Introduction	1995	19:55	Lecture	109
 Salton Honored - A SMART Celebration: Testing Predictions	1995	40:45	Lecture	85
 Salton Honored - A SMART Celebration: Influence of SMART on IR Systems #1.2	1995	06:18	Lecture	82
 Salton Honored - A SMART Celebration: Tape #3 Full	1995	29:58	Lecture	70
 Salton Honored - A SMART Celebration: Influence of SMART on IR Systems #2	1995	12:26	Lecture	68
 Salton Honored - A SMART Celebration: IR hearted in Digital Libraries	1995	25:17	Lecture	56
 Salton Honored - A SMART Celebration: Influence of SMART on IR Systems #1.1	1995	12:44	Lecture	51
 Salton Honored - A SMART Celebration: Influence of SMART on IR Systems #3	1995	21:33	Lecture	44

Display option four—no words

The Open Video Project :: Video Results - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://open-video.org/results.php

ARIC Public Use Data Best of the Web Channel Guide Customize Links DEMO CD Free Hotmail Internet Explorer Ne... Internet Start Microsoft Product News RealPlayer Today's Links Web Gallery >>

OV THE OPEN VIDEO PROJECT
a shared digital video collection

- Home
- Contribute
- About

Modify Search

Search: All fields
for: salton

Genre: - Any Genre -
Duration: - Any Duration -
Format: - Any Format -

Color: Color B&W Either
Sound: Sound Silent Either


Search

Detailed Search

Page 1 Search Results (16 videos found)

Layout: [Icons for layout options]

Sort by: Popularity
Results per page: 20



Home • Search • Collections • Contribute • About • Project News

The Open Video Project is managed at the [Interaction Design Laboratory](#),
at the School of Information and Library Science, University of North Carolina at Chapel Hill

Done


Search

ex. "water" or "space shuttle"

Detailed Search

Related Video

• **Video Grab Bag**

 **NASASciFiles - The Case of the Unknown Stink**

Other random videos

- Apollo, Segment 6005
- NASADestinationTomorrow - Episode 4
- NASA Connect - EOM - Planetary Observer Activity

[more...](#)

• **Related keyword searches**

- Information Retrieval
- IR
- SMART
- Cornell
- 30 Years
- Computer Science

Video Details

Salton Honored - A SMART Celebration: Professor Gerard Salton



Storyboard

Talk by professor Gerard Salton about interesting issues in information retrieval.

Download:

- MPEG-1 • 150.10 MB
- MPEG-2 • 566.10 MB
- MPEG-4 • 96.80 MB

[This video is one segment of a larger video title. See all 5 parts >>](#)

Video Information

Year:	1995
Genre:	Lecture
Keywords:	Information Retrieval; IR; SMART; Cornell; 30 Years; Computer Science; Gerard Salton
Duration:	00:15:07
Color:	Yes
Sound:	Yes
Amount of Motion:	Medium
Language:	English
Sponsor:	N/A
Contributing Organization:	Computer Science Dept., Cornell Univ.
Transcript Available:	No
Copyright Statement:	Use for research purposes only

Digitization Information

Digitization Date:	2004
Digitizing Organization:	Open Video


Search

ex. "water" or "space shuttle"

[Detailed Search](#)

Related Video

♦ **Video Grab Bag**

 **A New Horizon, segment 06 of 13**

Other random videos

- Oceanfloor Legacy, segment 11 of 14
- Winning: Aerospace, Segment 09
- NASAConnect - Geometry of Exploration: Water Below the Surface of Mars

[more...](#)

♦ **Related keyword searches**

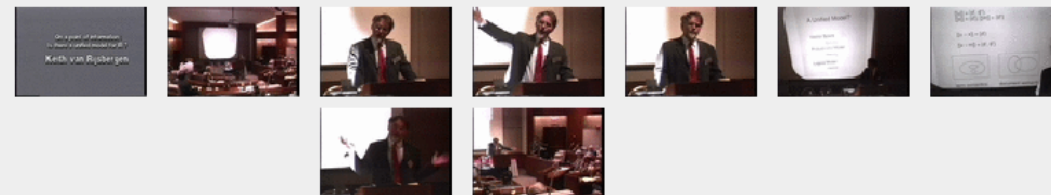
- Information Retrieval
- IR
- SMART
- Cornell
- 30 Years
- Computer Science
- Keith van Rijsbergen

Video Details

Salton Honored - A SMART Celebration: A Unified Model for IR?



The first speaker of the conference, Keith van Rijsbergen, presented the talk - On a point of information: Is there a unified model for IR?



Download:

- MPEG-1 • 386.10 MB
- MPEG-2 • 1473.40 MB
- MPEG-4 • 262.70 MB

[This video is one segment of a larger video title. See all 5 parts »](#)

Video Information

Year:	1995
Genre:	Lecture
Keywords:	Information Retrieval; IR; SMART; Cornell; 30 Years; Computer Science; Keith van Rijsbergen; Unified Model
Duration:	00:40:58
Color:	Yes
Sound:	Yes
Amount of Motion:	Medium
Language:	English
Sponsor:	N/A
Contributing Organization:	Computer Science Dept., Cornell Univ.
Transcript Available:	No
Copyright Statement:	Use for research purposes only

Digitization Information


Search

ex. "water" or "space shuttle"

Detailed Search

Related Video

• **Video Grab Bag**

 **Fights of nations**

Other random videos

- Lancaster, Pa., high school
- NASA Connect - BHFSTE - Exercise In Space
- CoMedi: Using Computer Vision to Support Awareness and Privacy in Mediaspaces

[more...](#)

• **Related keyword searches**

- Sex education
- Teenagers
- Social guidance

Video Details

How Much Affection?



7-sec excerpt

Storyboard

FastForward

How far can young people go in petting and still stay within the bounds of personal standards and social mores?

- Download:**
- MPEG-1 • 206.90 MB
 - MPEG-2 • 525.00 MB
 - MPEG-4 • 58.40 MB

Video Information

Year:	1958
Genre:	Ephemeral
Keywords:	Sex education; Teenagers; Social guidance;
Duration:	00:19:48
Color:	No
Sound:	Yes
Amount of Motion:	n/a
Language:	English
Sponsor:	N/A
Contributing Organization:	Internet Archive
Transcript Available:	No
Copyright Statement:	Unrestricted use except for resell or conversion to formats other than open-source MPEG-4 format. See http://www.archive.org/about/terms.php for more information.

Digitization Information

Digitization Date:	2001
Digitizing Organization:	Internet Archive

Some Interaction Principles and Caveats in These Examples

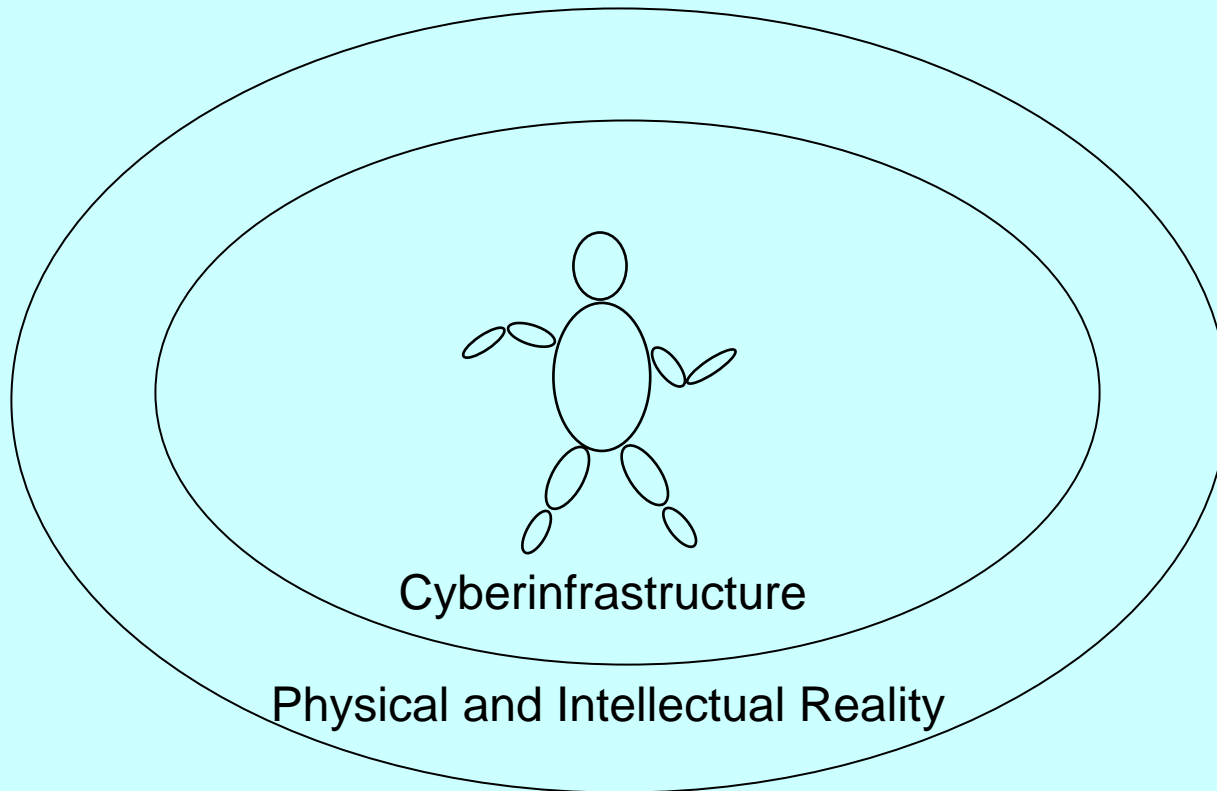
- Principles
 - Look ahead without penalty
 - Minimize scrolling and clicking
 - Alternative ways to slice and dice
 - Closely couple search, browse, and examine
 - Continuous engagement—useful attractors
 - Treasures to surface
- Caveats
 - Scalability (getting metadata to client side)
 - Metadata crucial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

Evaluation

- Beyond recall and precision
- Time???
 - > time better for learning
 - < time better for work (what about when learning is your work, e.g., research)
- Actions
 - Click, buy, print, save, read, forward
- Understand
- Change behavior/attitude

Long Term Paradigm: Information Interaction as Core Life Process

Examples represent early ways 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.



Coda

- Our hopes that we can create systems (solutions) that 'do' IR for us are unreasonable
- Our expectations that people can find and understand information without thinking and investing effort are unreasonable.
- We aim to develop 'systems' that involve people and machines continuously learning and changing together. Google would not work as well next month if there were not a large group of employees tuning the system, adding new spam filters, and crawlers checking out pages and links continuously.

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

NOTE: ACM/IEEE JCDL June 11-15, 2006

www.jcdl2006.org