

Human-Computer Information Retrieval: Finding and Understanding What We Need

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

Department of Computer Science Distinguished
Lecture Series
University of Toronto
September 30, 2004

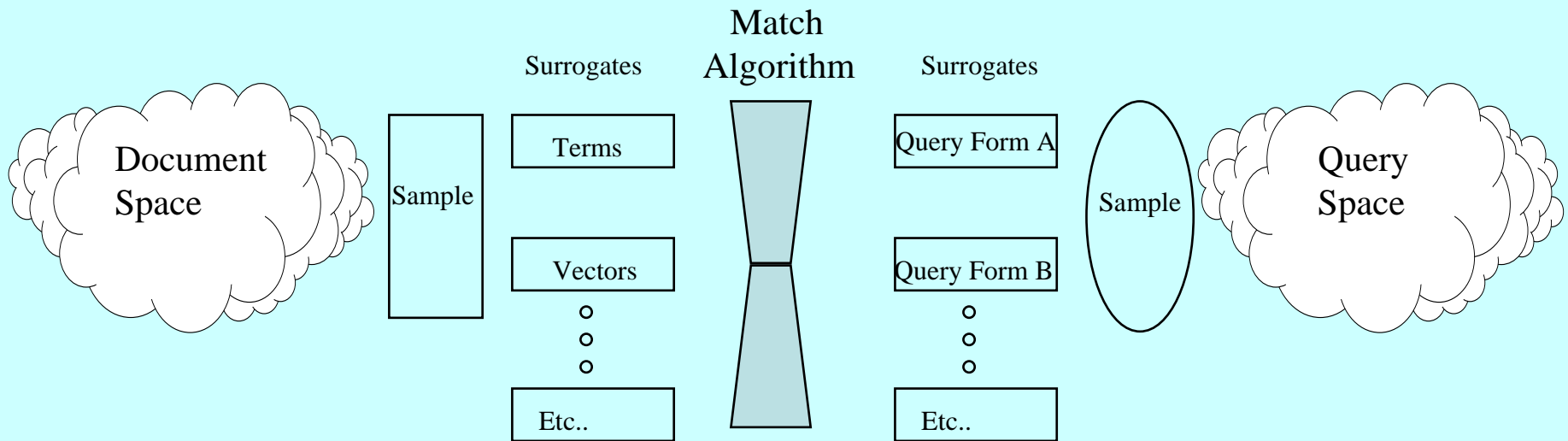
Message

- HCI and IR 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 R&D
- Integrating the human and system interaction is the main design challenge: syminforosis—people continuously engaged with meaningful information

Outline

- IR and the WWW
- HCI and the WWW
- HCIR rooted in system development
- Examples
 - Open Video
 - Relation Browser
- Challenges and Opportunities

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, personal collections
- Content acquires history

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?

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

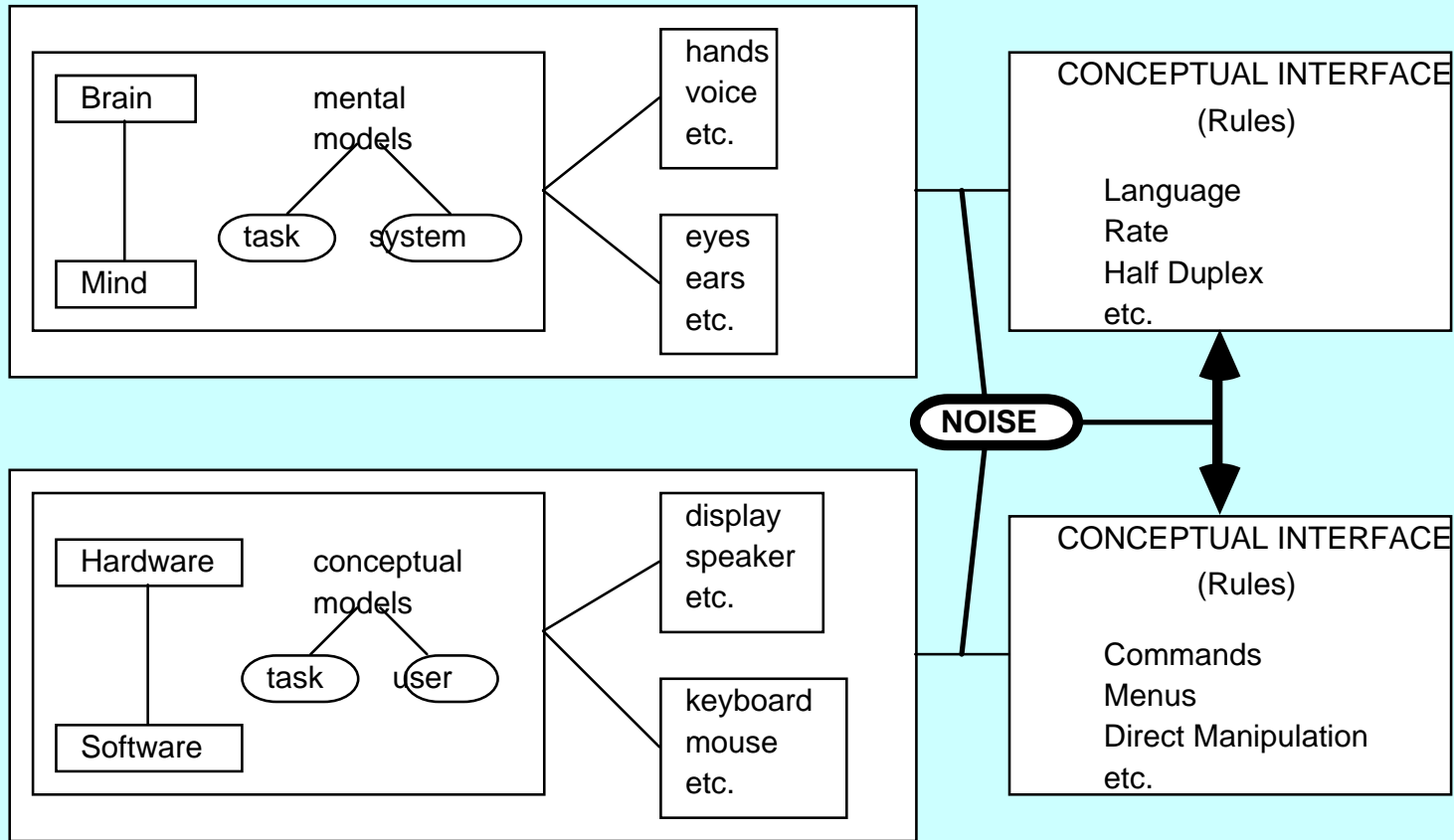
Responses to People Trend

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

An Expanded Model:

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

Human-Computer Communication Model of HCI



A user-oriented model that has driven R&D. Evaluation based on user time, accuracy, and satisfaction.

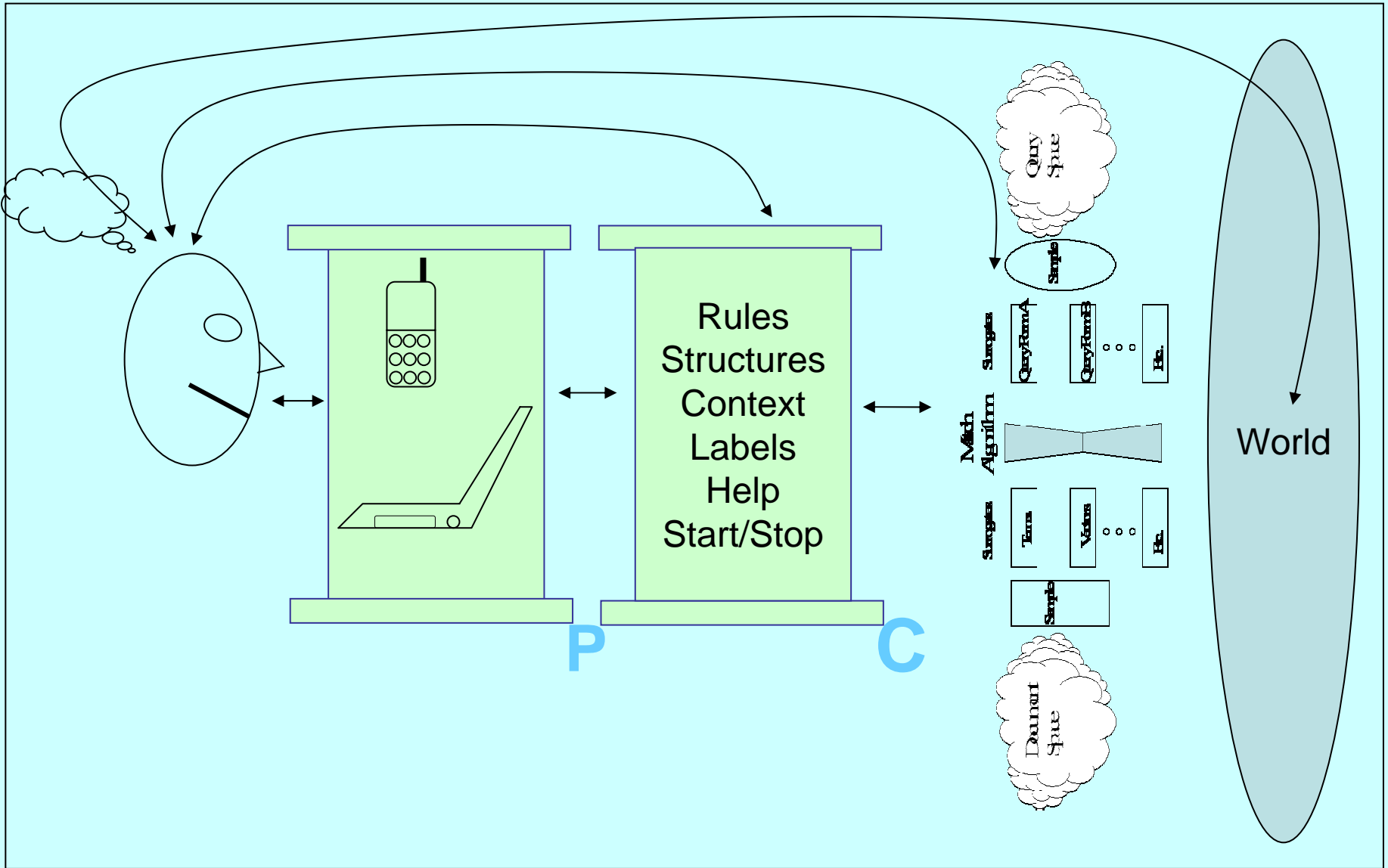
HCI WWW Trends

- First decade of WWW as great equalizer (we all get impoverished, but we admit MANY more people)
- Universal access
- Platform independence (lots of devices)
- Enhanced browsers, specialized browsers
- Interface Servers
- Social awareness (user is not alone)

HCIR

- Trend toward getting people closer to the information they need
 - Closer to the backend
 - Closer to the meaning
- Increasing responsibility as well as control
- More demanding and knowledgeable installed base
- Ubiquity, digital libraries, e-commerce as extended memories and tools (personal and shared)

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
- Moving beyond retrieval to understanding
 - context

Two examples of getting people involved in continuous decision making and interaction with information resources: dynamic queries and the agile views interaction framework instantiated in Open Video and Relation Browser


Open Video Example








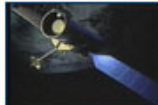
www.open-video.org

- Open access digital library of digital video for education and research
- 2000+ 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









Page 1 Search Results (111 videos found)

Layout:  Sort by: Relevance Results per page: 10

 <p>Space Works 6, complete video 1986 • Documentary • Popularity (downloads): 168</p>	 <p>Space Works 5, complete video 1986 • Documentary • Popularity (downloads): 188</p>	 <p>Space Works 7a, complete video 1986 • Documentary • Popularity (downloads): 55</p>	 <p>Cheerios/V-8 "Space Offer" Television Commercial 1960 • Ephemeral • Popularity (downloads): 522</p>
 <p>STS-48 Earth Views with In-Cabin and FCR Activities, segment 07 of 9 Documentary • Popularity (downloads): 302</p>	 <p>ambientROOM: Integrating Ambient Media with Architectural Space 1998 • Educational • Popularity (downloads): 116</p>	 <p>Space Works 8, complete video 1986 • Documentary • Popularity (downloads): 116</p>	 <p>The Four Great Observatories Educational • Popularity (downloads): 223</p>

Page 1 Search Results (111 videos found)

Layout:  Sort by: Relevance Results per page: 10

Title	Year	Duration	Genre	Popularity
 Space Works 6, complete video	1986	29:09	Documentary	168
 Space Works 5, complete video	1986	29:49	Documentary	188
 Space Works 7a, complete video	1986	29:03	Documentary	55
 Cheerios/V-8 "Space Offer" Television Commercial	1960	01:00	Ephemeral	522
 STS-48 Earth Views with In-Cabin and FCR Activities, segment 07 of 9		14:22	Documentary	302
 ambientROOM: Integrating Ambient Media with Architectural Space	1998	05:30	Educational	116
 Space Works 8, complete video	1986	27:41	Documentary	116
 The Four Great Observatories		05:26	Educational	223

Alternative Previews for a Specific Video Segment

The screenshot shows a web browser window titled "The Open Video Project - Video Details - Microsoft Internet Explorer". The address bar shows the URL "http://www.open-video.org/details.php?id=vid-012". The page 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 in the top right.

The main content area is divided into several sections:

- Search:** A search box with a "Search" button and a "Default 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" with a list of links: "An Animated Direct Manipulation Interface to Digital Library Services", "Classic Television Commercials (Part II)", and "Brazil: South American Medley". A "Related keyword searches" section lists "HCL".
- Video Details:** A section titled "Browsing and annotating digital photographs with Photofinder" with a video thumbnail. To the right of the thumbnail are three buttons: "Free excerpt", "Storyboard", and "Fullframe". A descriptive paragraph follows: "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 with a "Download" button and a file icon, labeled "MP4-1 • 29.29 MB".
- Video Information:** A table listing video metadata.

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

Relation Browser Example

www.idl.ils.unc.edu/rave

- A general purpose dynamic query interface for databases with a small number of facets (~10) and a small number of categories in each facet (~10).
- Easy to look ahead (overviews and previews)
- Couples interactive partitioning/exploration with string query
- Semi-automatic category generation and webpage classification

Relation Browser Start State for Energy Information Admin Website

The screenshot shows a Java Applet Window titled "EIA Web Collection". It features a grid of categories with the following data:

Fuel Type	Geography	Sector	Process
514 Alternatives	3043 State	2328 Commercial	845 Delivery
3860 Coal	2275 Region	2024 Electric Utility	1852 Imports/exports
4135 Electricity	3114 U.S.	2431 Industrial	2950 Price/Cost
2916 Natural Gas	2009 International	2328 Residential	2446 Production
3166 Nuclear			2287 Resources/reserves
3766 Petroleum			2762 Usage
1873 Renewable			

Below the grid is a search bar with "0 result" and a "Search" button. There are also buttons for "Fewer Categories <<" and "More Categories >>".

At the bottom, there is a table with columns for "Title", "Page Size", and "URL". The table is currently empty.

Java Applet Window

Mousing over “Coal” under the “Fuel type” category reveals the distribution of coal related web pages to other categories

The screenshot shows a Java Applet Window titled "EIA Web Collection". It features a search interface with four main categories: Fuel Type, Geography, Sector, and Process. The "Coal" option under "Fuel Type" is highlighted in blue. Below the search results, there are fields for "Title:", "Page Size:", and "URL:", and a table with columns for "Title", "Page Size", and "Description".

Fuel Type	Count	Geography	Sector	Count	Process	Count
Alternatives	1219	State	Commercial	435	Delivery	
Coal	1218	Region	Electric Utility	947	Imports/exports	1022
Electricity	1419	U.S.	Industrial	1339	Price/Cost	1483
Natural Gas	1006	International	Residential	1176	Production	1252
Nuclear					Resources/reserves	1191
Petroleum					Usage	1508
Renewable						

0 result Search Fewer Categories << More Categories >>

Title: Page Size: URL:

Title	Page Size	Description
-------	-----------	-------------

Java Applet Window

Click on Natural Gas and Mouse over Residential Sector

EIA Web Collection

Fuel Type: [Dropdown] Geography: [Dropdown] Sector: [Dropdown] Process: [Dropdown]

84	Alternatives	540	State	434	Commercial	193	Delivery
365	Coal	226	Region	177	Electric Utility	205	Imports/exports
404	Electricity	252	U.S.	377	Industrial	407	Price/Cost
902	Natural Gas	189	International	902	Residential	314	Production
315	Nuclear					206	Resources/reserves
280	Petroleum					331	Usage
157	Renewable						

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

Title	Page Size	URL	Description
Energy Glossary, R page	49k	...	It includes the reconfiguration of the vertically-integrated ...
The Northeast Heating Fuel Market: Assessment and Optio...	14k and Pennsylvania. This conforms to Census region 1, co...
- Natural Gas 1996 District of Columbia District of Columbia	null Summary Statistics for Natural Gas District of Columbia, ...
Figure 6.5 Natural Gas Consumption by Sector	null Administration/Annual Energy Review 2001 Figure 6.5 N...
- Natural Gas 1996 Washington Washington	null Vehicle Fuel: Deliveries to Consumers: Electric Resident...
- Natural Gas 1999 NewHampshire New Hampshire -Tabl...	null Summary Statistics for Natural Gas New Hampshire, 19...
- Natural Gas 1998 NewYork New York -Table 79	null Administration / Natural Gas Annual 1998 156 - Natural ...
Ukraine	17k of the country-s carbon emissions, with the remainder c...
- Natural Gas 1999 Tennessee Tennessee -Table 83	null Summary Statistics for Natural Gas Tennessee, 1995-19...
Residential Sector Channel	39k Washington, DC 20585. - Home - Petroleum - Gasoline ...
The National Energy Modeling System: An Overview 2000 - ...	13k Energy Information Administration, NEMS International E...
Voluntary Reporting of Greenhouse Gases 1997 - 3. Reduc...	43k or, in the case of CLE Resources, a subsidiary of an elec...
Brunei Country Analysis Brief	26k A second consortium, this one between TotalFinaElf and ...
Retail Unbundling - West Virginia	11k	...	Retail Unbundling - West Virginia. Status: All residential cu...
Energy Plug: Winter Fuels Outlook: 2001-2002	8k Base Forecast. Natural Gas (Midwest). ... Expenditures a...
Energy Savings	11k the savings vary according to the climate in the Census ...
Next Generation * Natural Gas Information Requirements	null For example, the State of Georgia has mandated retail s...
- Natural Gas 1998 Mountain Mountain -Table 42	null Summary Statistics for Natural Gas Mountain, 1994-1998...
Natural Gas Use in American Households	9k Census Region: More than three-fourths of households i...
- Natural Gas 2000 Michigan Michigan -Table 64	null Summary Statistics for Natural Gas Michigan, 1996-2000...

(Fuel Type=Natural Gas)****(Sector=Residential)

RB++ showing 'hous' typed in title field

The screenshot shows the RB++ search interface with the following filters applied:

- Fuel Type:** 59 results (Natural Gas selected)
- Geography:** 10 results (U.S. selected)
- Sector:** 26 results (Residential selected)
- Process:** 36 results (Usage selected)

Search results for "hous" (59 results):

Title	Page Size	Page Size	URL	Description
Greenhouse Gases 1987-1994, Text Boxes	11k			... EIA-s natural gas "venting and flaring" data ar...
Table 1. Dollars Saved per household for a 1-; F Lower	null			... in the Northeast Census Region, 1997 (Continued) RSE ...
Voluntary Reporting of Greenhouse Gases 1997 - 2.	45k			... the renewable energy displaced SCEGÇOs marginal nat...
Emissions of Greenhouse Gases in the United States 1997	21k			US Nitrous Oxide Emissions from Stationary Combustion ...
Table 3. Dollars Saved per household for a 1-; F Lower	null			... Division in the South Census Region, 1997 (Continued) ...
housing Characteristics 1993 -- Total US Data	3k			null
Voluntary Reporting of Greenhouse Gases 2000 - Chapter	31k			... For example, a single DSM program reported by an electr...
Voluntary Reporting of Greenhouse Gases 1999 - 3.	34k			... Two projects involved fuels other than natural gas and el...
Table 2.5 household Energy Consumption by Census	null			(Quadrillion Btu, Except as Noted). Census Region 1. 1978...
Table 2. Dollars Saved per household for a 1-; F Lower	19k			... Midwest Census Region - - -
Table 4. Total Energy Expenditures in US households by	null			... Table 4. Total Energy Expenditures in US Households by...
Voluntary Reporting of Greenhouse Gases 1997 - 1.	37k			... Peabody in the coal mining industry, BP America in the p...
Voluntary Reporting of Greenhouse Gases 2000 - Chapter	34k			... Btu per standard cubic foot to calculate the amount of nat...
Voluntary Reporting of Greenhouse Gases 1997 - 3.	43k			... or, in the case of CLE Resources, a subsidiary of an elec...
Table 1. Total Energy Consumption in US households by	null			... Table 1. Total Energy Consumption in US Households b...
Emissions of Greenhouse Gases in the United States 1997	22k			US Nitrous Oxide Emissions from Stationary Combustion ...
Table 3. Dollars Saved per household for a 1-; F Lower	21k			... South Census Region - - -
Space-Heating energy used by households in the	12k			... Natural Gas and LPG, per Household.Table 5.28. The f...
Figure 7. Electricity and Natural Gas Prices: All US	3k			Figure 7. Electricity and Natural Gas Prices: All US Househ...
Voluntary Reporting of Greenhouse Gases 1997	56k			... emissions if hydropower is used to meet baseload dem...

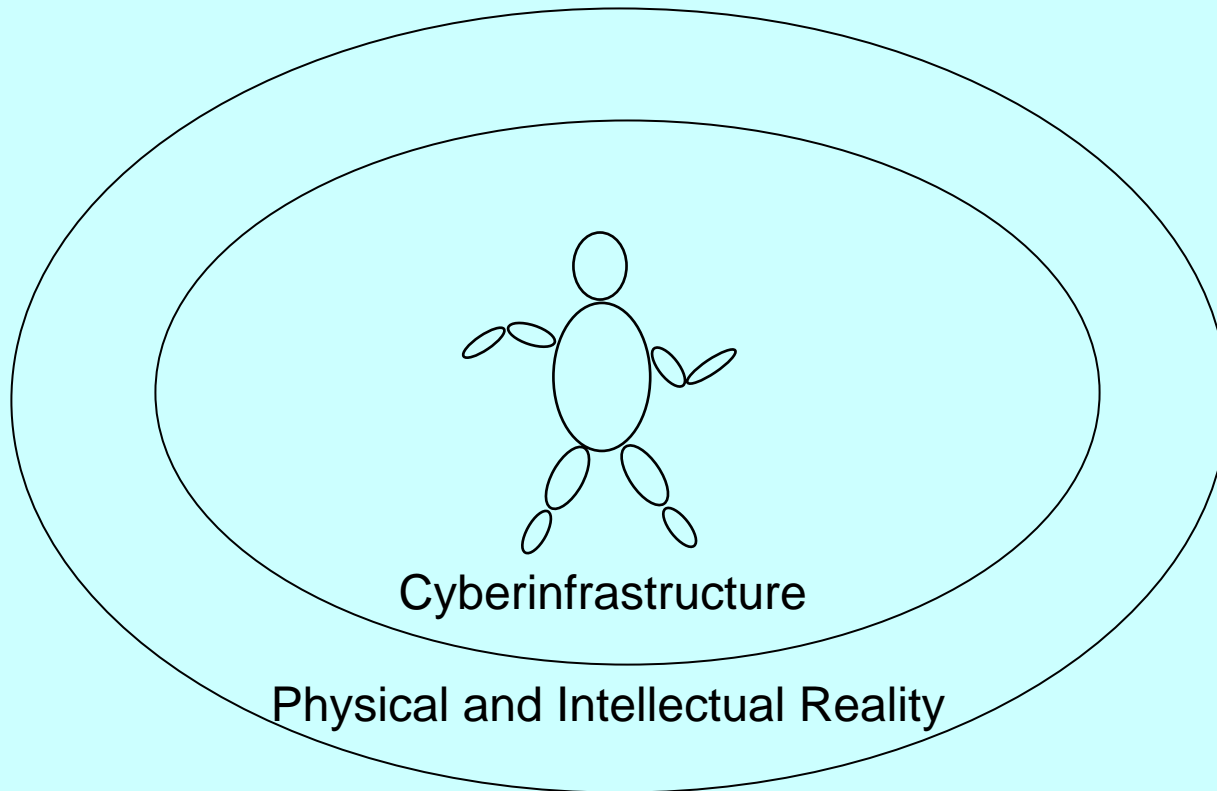
(Fuel Type=Natural Gas)****(Title=hous)

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!)

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