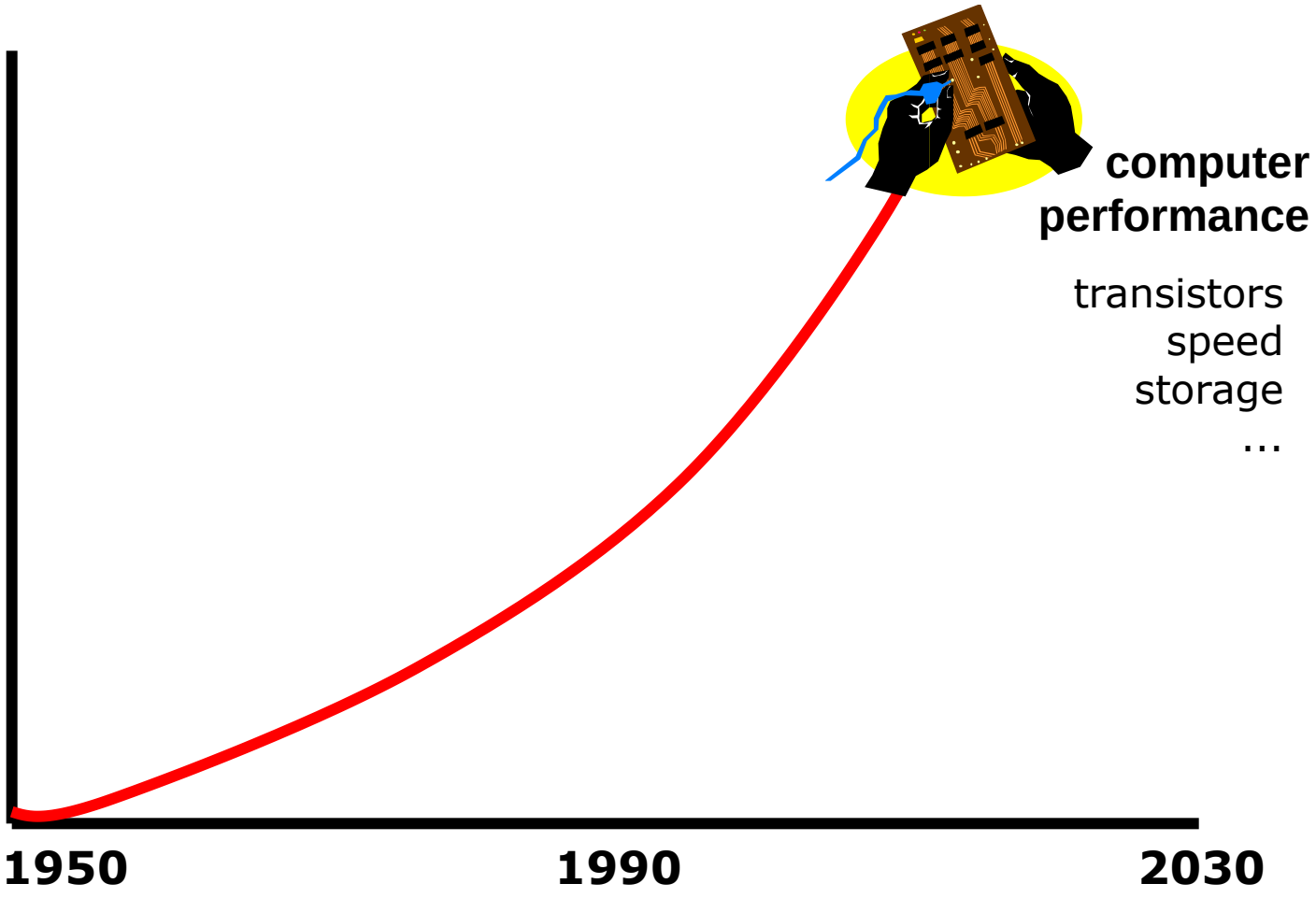
Interaction

LBSC 796/INFM 718R

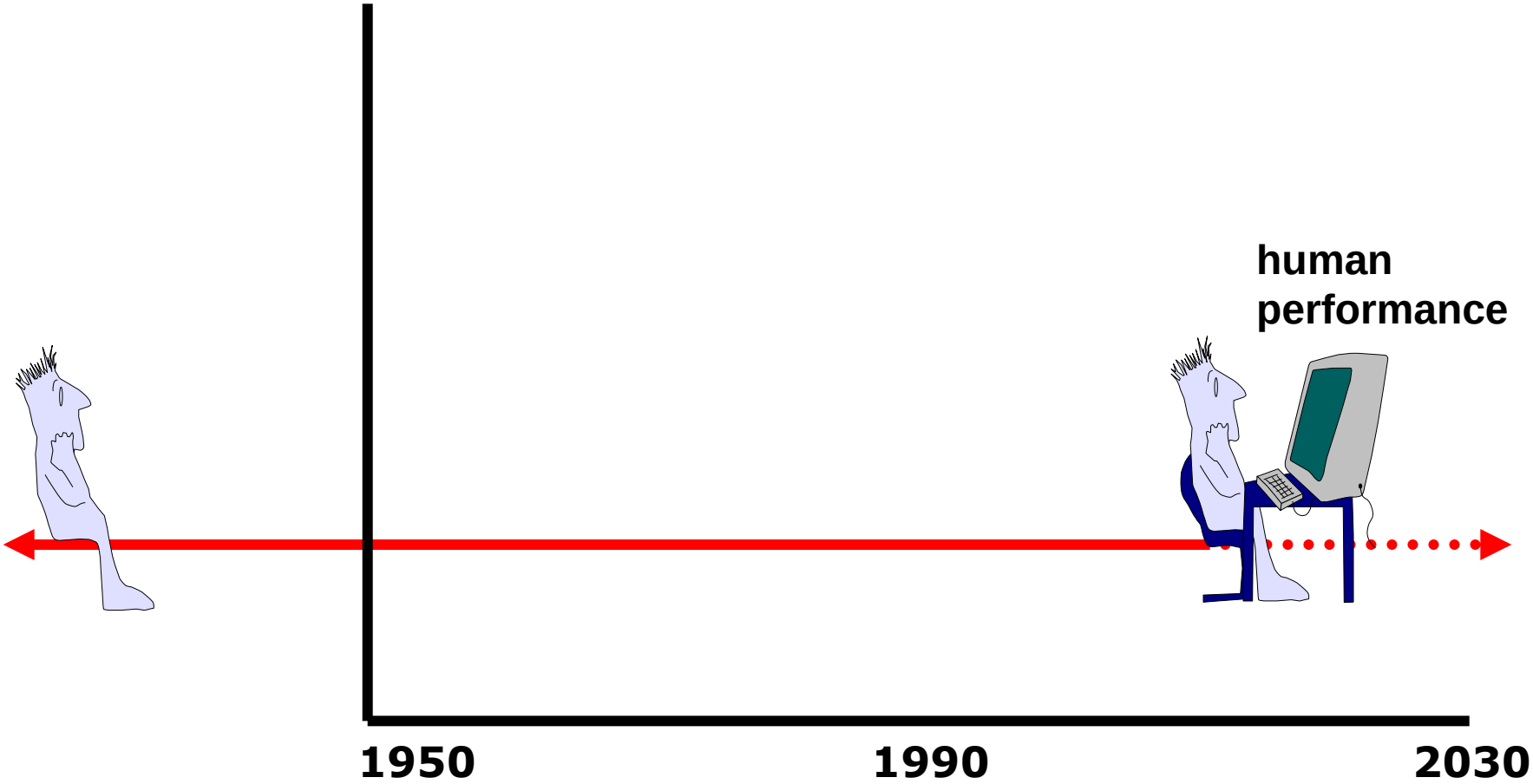
Douglas W. Oard

Week 4, February 23, 2011

Moore's Law

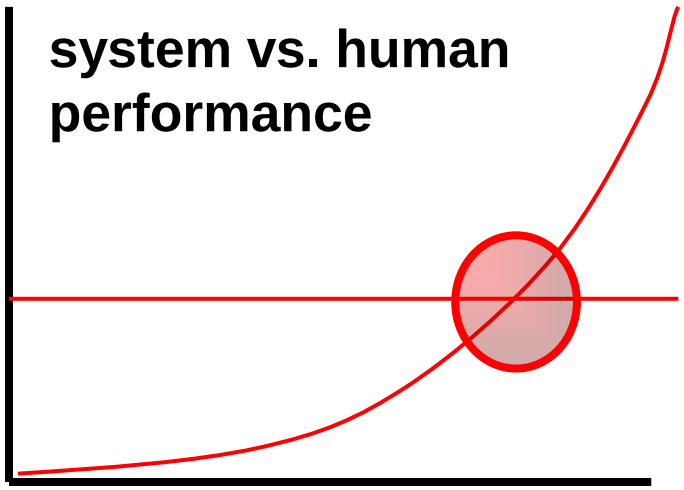

computer performance
transistors
speed
storage
...
1950
1990
2030

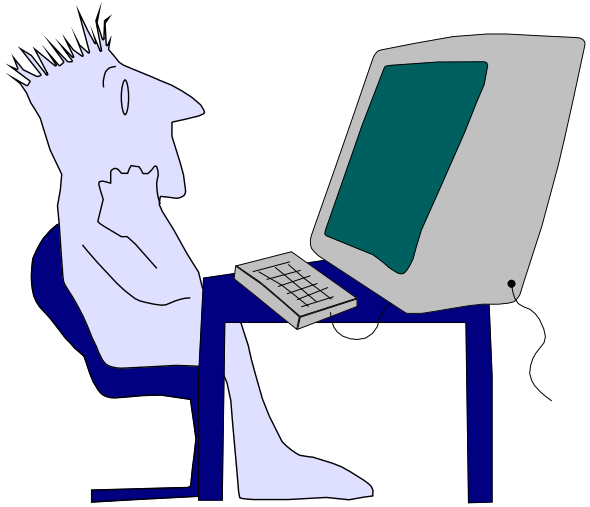
Human Cognition



1990

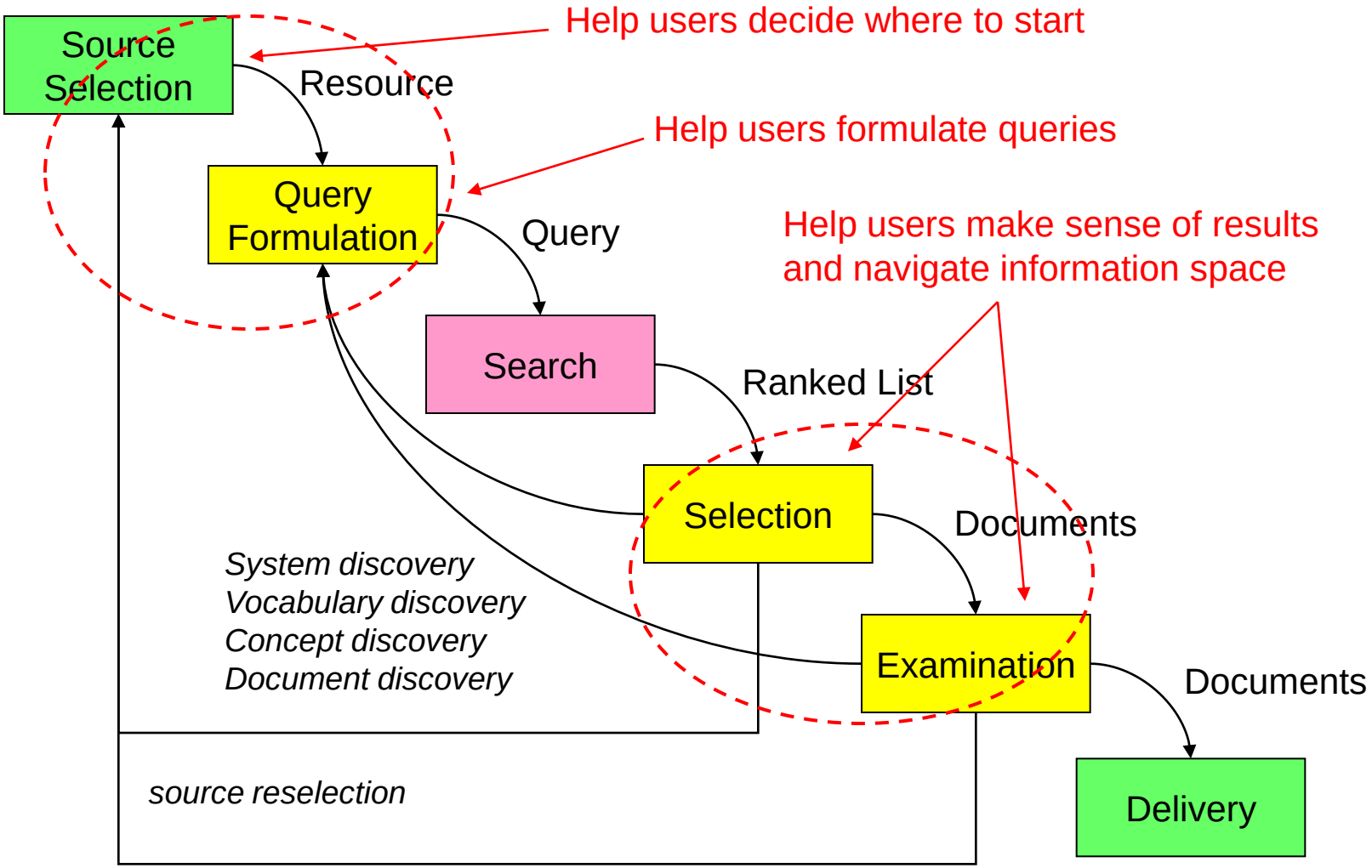
Where is the bottleneck?



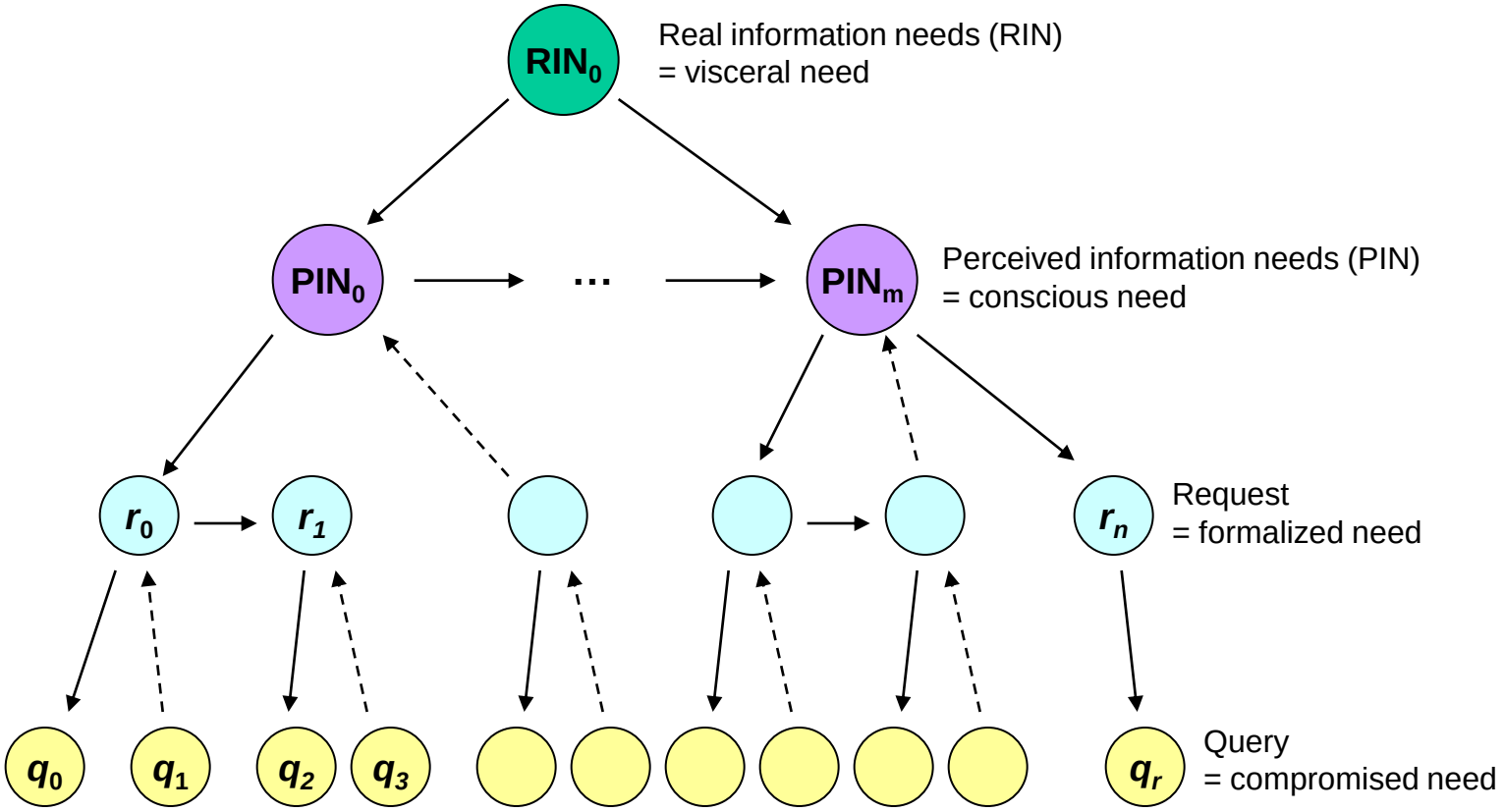


Slide idea by Bill Buxton

Interaction Points



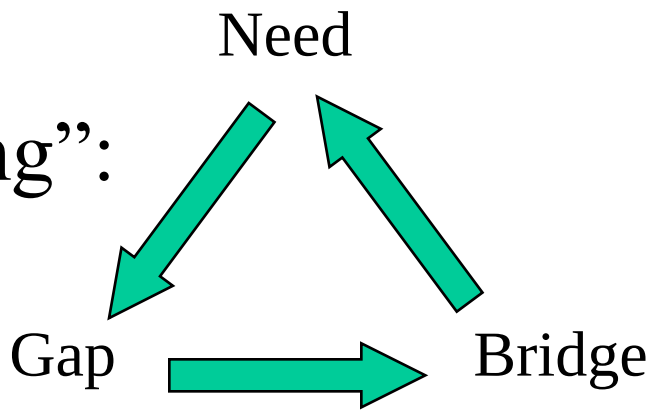
Information Needs



Stefano Mizzaro. (1999) How Many Relevances in Information Retrieval? *Interacting With Computers*, 10(3), 305-322.

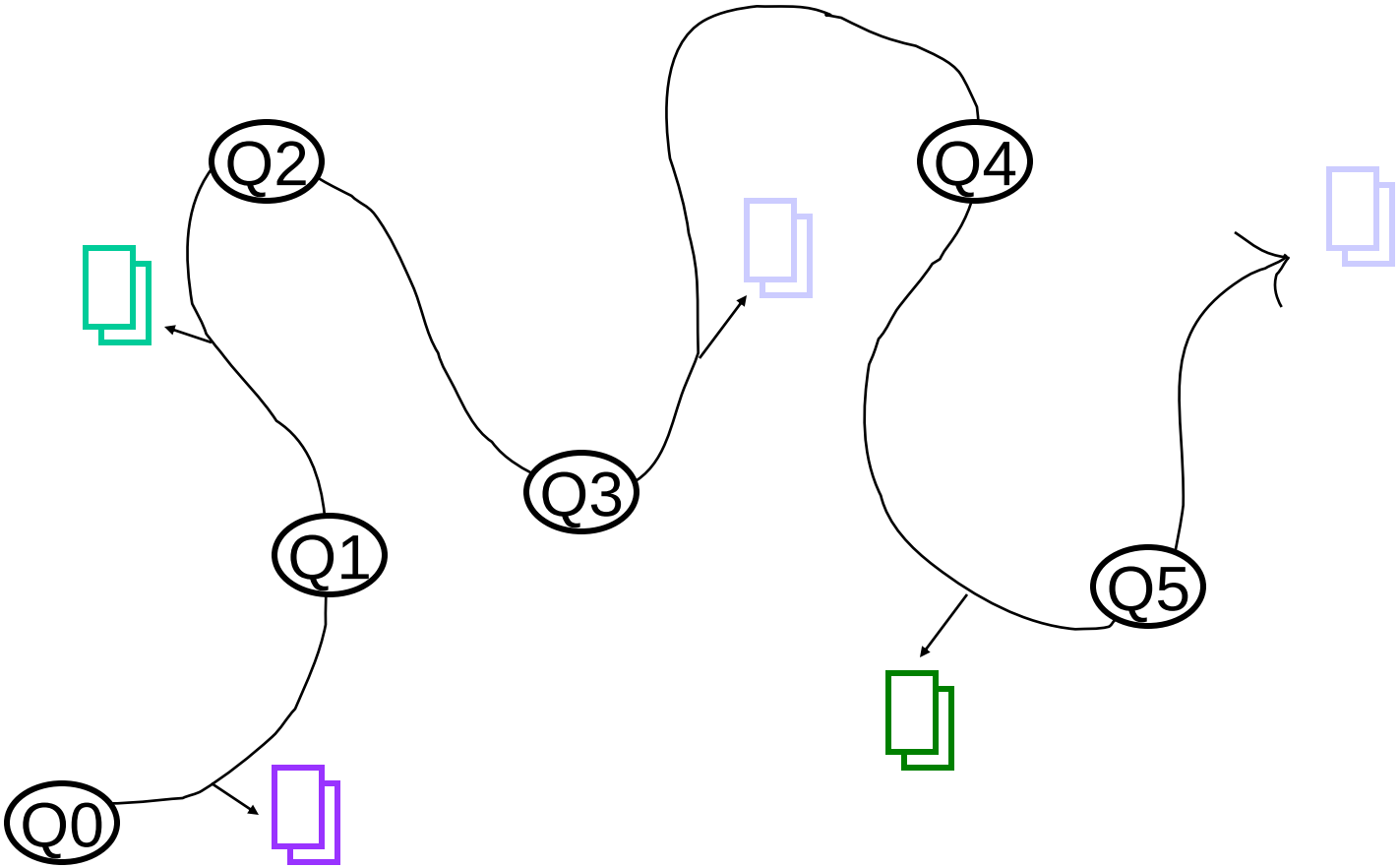
Anomalous State of Knowledge

- Belkin: Searchers do not clearly understand
  - The problem itself
  - What information is needed to solve the problem
- The query results from a clarification process
- Dervin's "sense making":



Bates' "Berry Picking" Model

A sketch of a searcher… "moving through many actions towards a general goal of satisfactory completion of research related to an information need."



Broder's Web Query Taxonomy

- Navigational (~20%)
  - Reach a particular site ("known item")
- Informational (~50%)
  - Acquire static information ("topical")
- Transactional (~30%)
  - Perform a Web-mediated activity ("service")

Andrei Broder, *SIGIR Forum*, Fall 2002

Some Desirable Features

- Make exploration easy
- Relate documents with why they are retrieved
- Highlight relationships between documents

Agenda

- Query formulation
- Selection
- Examination
- Source selection

Query Formulation

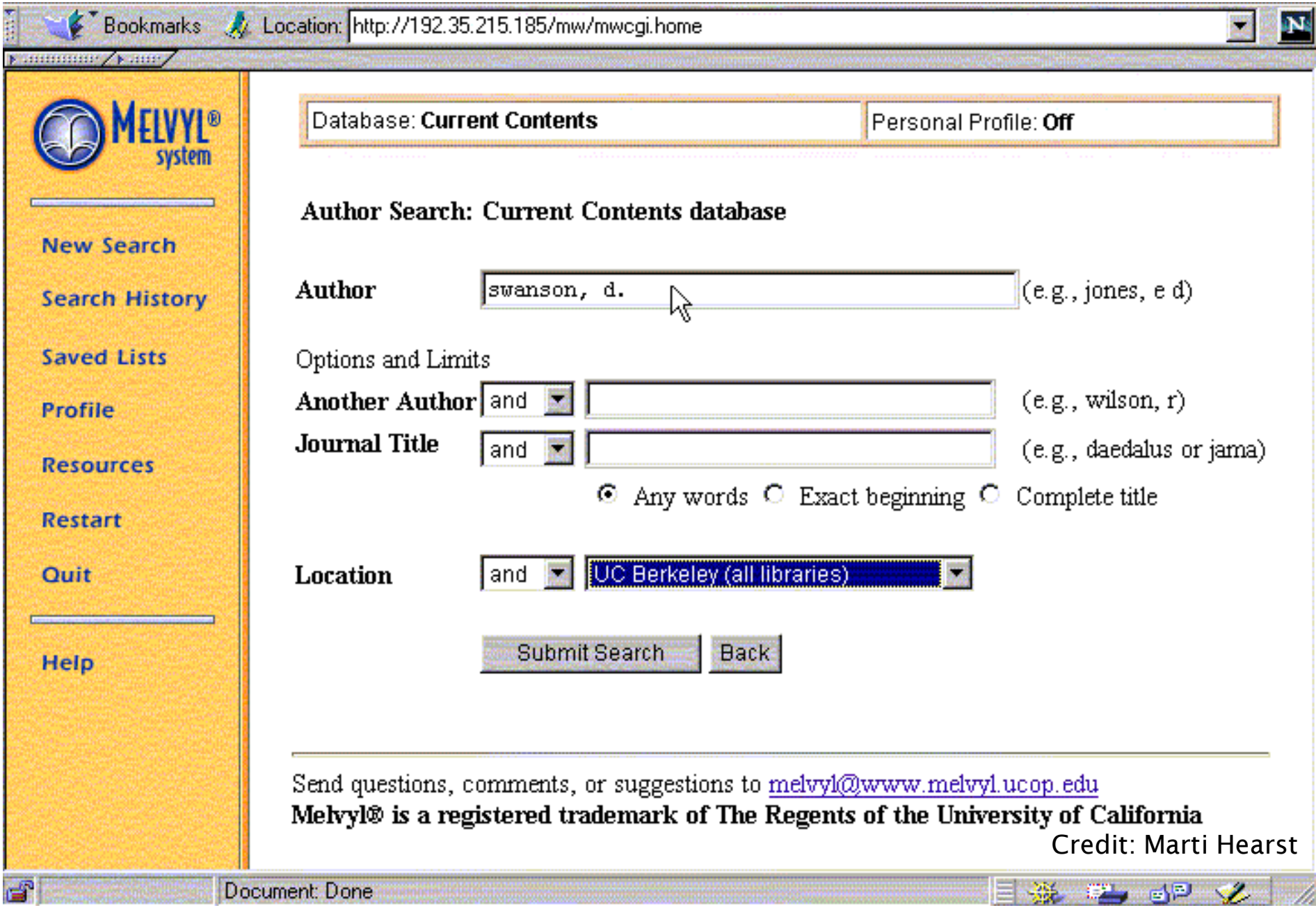
- Command Language
- Form Fill-in
- Menu Selection
- Direct Manipulation
- Natural Language

Ben Shneiderman, 1997

WESTLAW® Query Examples

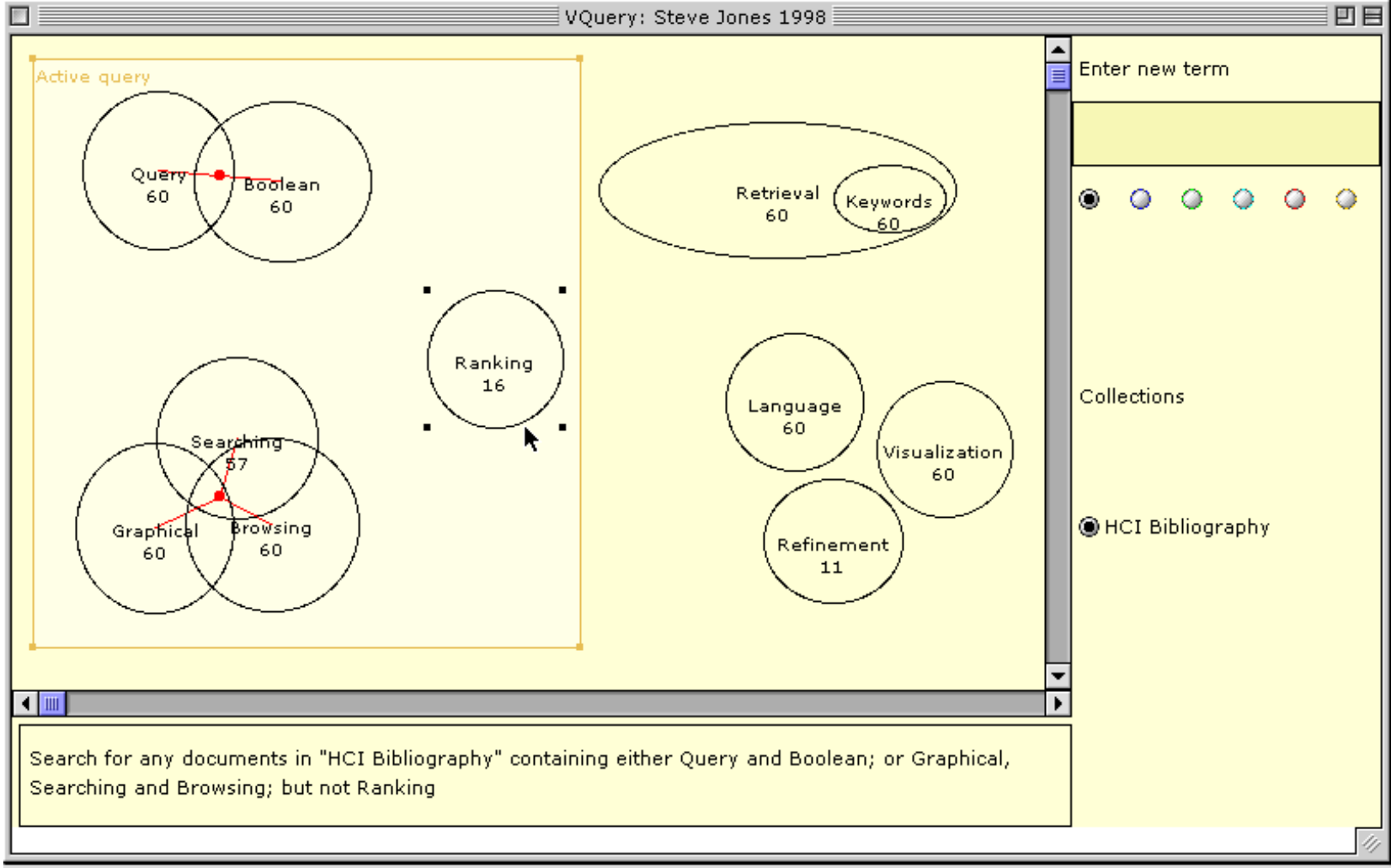
- What is the statute of limitations in cases involving the federal tort claims act?
  - LIMIT! /3 STATUTE ACTION /S FEDERAL /2 TORT /3 CLAIM
- What factors are important in determining what constitutes a vessel for purposes of determining liability of a vessel owner for injuries to a seaman under the "Jones Act" (46 USC 688)?
  - (741 +3 824) FACTOR ELEMENT STATUS FACT /P VESSEL SHIP BOAT /P (46 +3 688) "JONES ACT" /P INJUR! /S SEAMAN CREWMAN WORKER
- Are there any cases which discuss negligent maintenance or failure to maintain aids to navigation such as lights, buoys, or channel markers?
  - NOT NEGLECT! FAIL! NEGLIG! /5 MAINT! REPAIR! /P NAVIGAT! /5 AID EQUIP! LIGHT BUOY "CHANNEL MARKER"
- What cases have discussed the concept of excusable delay in the application of statutes of limitations or the doctrine of laches involving actions in admiralty or under the "Jones Act" or the "Death on the High Seas Act"?
  - EXCUS! /3 DELAY /P (LIMIT! /3 STATUTE ACTION) LACHES /P "JONES ACT" "DEATH ON THE HIGH SEAS ACT" (46 +3 761)

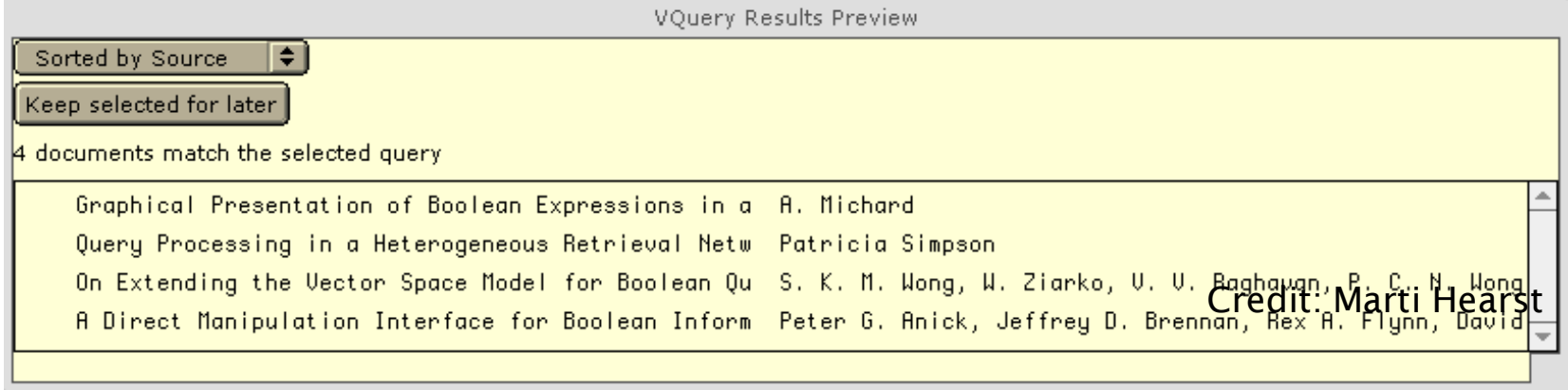
Form-Based Query Specification



Credit: Marti Hearst

Direct Manipulation Spec.
VQUERY (Jones 98)





Credit: Marti Hearst

Alternate Query Modalities

- Spoken queries
  - Used for telephone and hands-free applications
  - Reasonable performance with limited vocabularies
    - But some error correction method must be included
- Handwritten queries
  - Palm pilot graffiti, touch-screens, …
  - Fairly effective if some form of shorthand is used
    - Ordinary handwriting often has too much ambiguity

Agenda

- Query formulation
- ➢ Selection
- Examination
- Source selection

A Selection Interface Taxonomy

- One dimensional lists
  - Content: title, source, date, summary, ratings, ...
  - Order: retrieval status value, date, alphabetic, ...
  - Size: scrolling, specified number, score threshold
- Two dimensional displays
  - Construction: clustering, starfield, projection
  - Navigation: jump, pan, zoom
- Three dimensional displays
  - Contour maps, fishtank VR, immersive VR

Google: KeyWord In Context (KWIC)

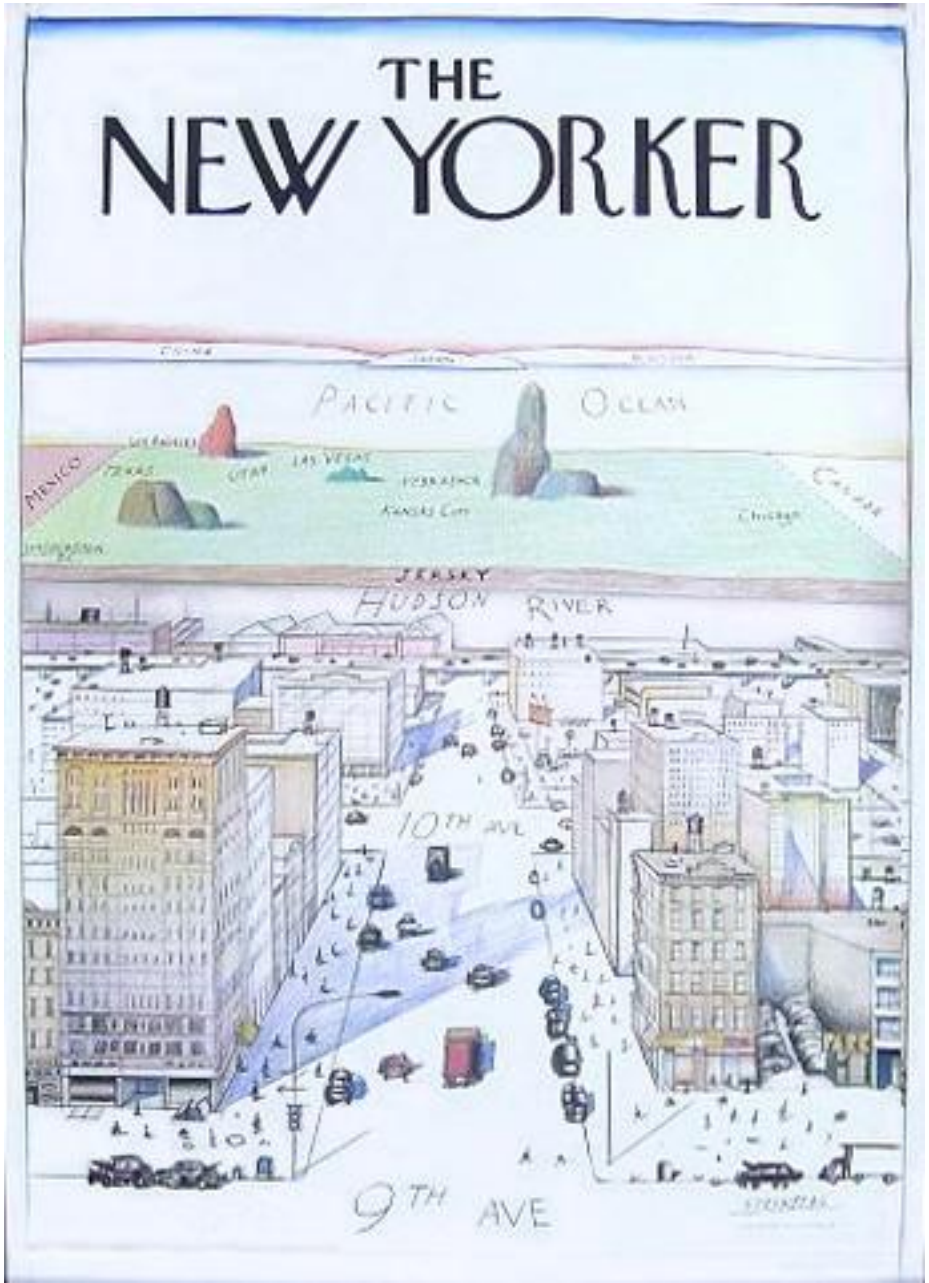
Query: University of Maryland College Park

The **University** of **Maryland**
... **University** of **Maryland**, **College Park**, MD 20742, USA 301.405.1000 Copyright 2003
University of **Maryland** Contact us with comments, questions and feedback.
Description: Official web site **of** the **University of Maryland** located in **College Park**, **Maryland**. Offers student,...
Category: Reference > Education > ... > College Park
www.umd.edu/ - 12k - Mar 5, 2004 - Cached - Similar pages

inforM @ **Maryland**
... Welcome to inforM@**Maryland**, one of the web hosting resources and services provided
to the **University** of **Maryland** by the Office of Information Technology (OIT). **...**
Description: The campus-wide information server for the **University of Maryland**. Offers a wide variety **of** information...
Category: Reference > Education > ... > College Park
www.inform.umd.edu/ - 8k - Mar 5, 2004 - Cached - Similar pages

Department of Computer Science
... at the **University** of **Maryland**, will be **...** of Google, has received the **university's** Outstanding
Young **...** Solutions, has received the **college's** Distinguished Alumnus **...**
Description: Offers BS, MS, and Ph.D. degrees in computer science. Research interests include artificial intellige...
Category: Computers > Computer Science > ... > United States > Maryland
www.cs.umd.edu/ - 34k - Mar 5, 2004 - Cached - Similar pages

Summarization



Indicative vs. Informative

- Terms often applied to document abstracts
  - Indicative abstracts support selection
    - They describe the contents of a document
  - Informative abstracts support understanding
    - They summarize the contents of a document
- Applies to any information presentation
  - Presented for indicative or informative purposes

Selection/Examination Tasks

- “Indicative” tasks
  - Recognizing what you are looking for
  - Determining that no answer exists in a source
  - Probing to refine mental models of system operation
- “Informative” tasks
  - Vocabulary acquisition
  - Concept learning
  - Information use

Generated Summaries

- Fluent summaries for a specific domain
- Define a knowledge structure for the domain
  - Frames are commonly used
- Analysis: process documents to fill the structure
  - Studied separately as “information extraction”
- Compression: select which facts to retain
- Generation: create fluent summaries
  - Templates for initial candidates
  - Use language model to select an alternative

Extraction-Based Summarization

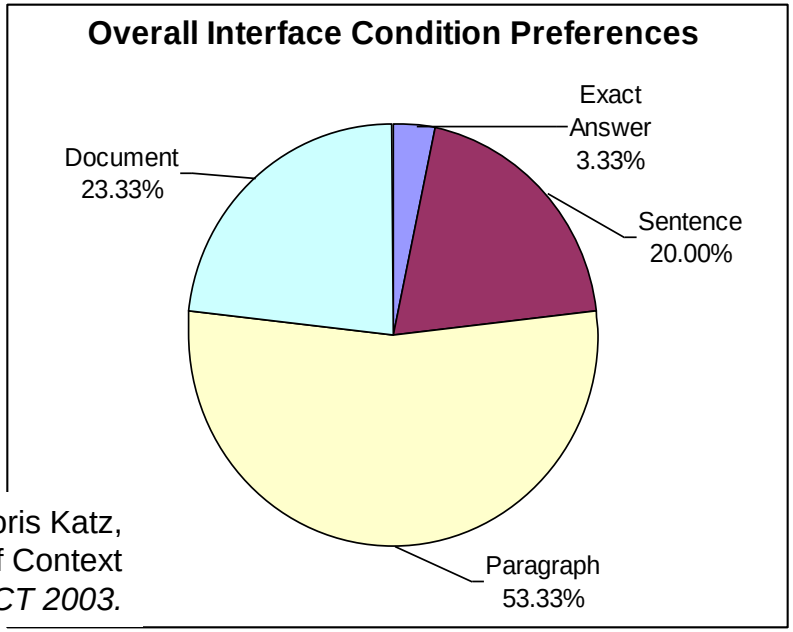
- Robust technique for making disfluent summaries
- Four broad types:
  - Query-biased vs. generic
  - Term-oriented vs. sentence-oriented
- Combine evidence for selection:
  - Salience: similarity to the query
  - Specificity: IDF or chi-squared
  - Emphasis: title, first sentence

Goldilocks and the Three Summaries…

- The entire document: too much!
- The exact answer: too little!

  It occurred on July 4, 1776.
  What does this pronoun refer to?

- The surrounding paragraph: just right…



Jimmy Lin, Dennis Quan, Vineet Sinha, Karun Bakshi, David Huynh, Boris Katz, and David R. Karger. (2003) What Makes a Good Answer? The Role of Context in Question Answering. *Proceedings of INTERACT 2003.*

Open Directory Project

dmoz open directory project — In partnership with AOL search

home | feedback

Search: **solar**

Open Directory Categories (1-5 of 19)

1. **Business: Energy and Environment: Renewable** *(304 matches)*
2. **Science: Astronomy: Solar System** *(112)*
3. **Science: Technology: Energy: Renewable: Solar** *(101)*
4. **Kids and Teens: School Time: Science: Astronomy and Space: Solar System** *(41)*
5. **Business: Energy and Environment: Renewable: Solar** *(31)*

[more...]

Open Directory Sites (1-20 of 3121)

1. **Solar** Hydrogen Energy Corporation - Shec has developed a process to produce hydrogen which does not create green house gases such as co or co2, using only water and sunlight as the inputs into the process.
   -- *http://www.**solar**-h2.com/* *Business: Energy and Environment: Renewable (304)*

2. Views of the **Solar** System - Comprehensive multimedia tour with statistical data, detailed information, pictures and videos. Also includes educational resources.
   -- *http://www.solarviews.com/eng/homepage.htm* *Science: Astronomy: Solar System (112)*

3. American **Solar** Energy Society - Web site has information on membership and publications, as well as policy statements on topics related to renewable energy and electric utility restructuring.
   -- *http://www.ases.org* *Science: Technology: Energy: Renewable: Solar (101)*

4. Passport to the **Solar** System - Provides an overview of the sun and each of the nine planets. Also includes biographies of researchers.
   -- *http://www.passporttoknowledge.com/solarsystem/* *Kids and Teens: School Time: Science: Astronomy and Space: Solar System (41)*

5. Innovative Power Systems - Designs, installs, sells, and services renewable energy systems for homes and businesses in Minnesota and the surrounding region.
   -- *http://www.ips-**solar**.com/* *Business: Energy and Environment: Renewable: Solar (31)*

6. NASA Eclipse Home Page - Maps and tables for 7,000 years of lunar and **solar** eclipses, eclipse photographs, observing tips, and eye safety information.
   -- *http://sunearth.gsfc.nasa.gov/eclipse/eclipse.html* *Science: Astronomy: Eclipses, Occultations and Transits: Eclipses (33)*

7. **Solar** Data Analysis Center - **Solar** images, **solar** news, eclipse information, **solar** data, NASA **solar** physics programs.
   -- *http://umbra.nascom.nasa.gov/sdac.html* *Science: Technology: Space: NASA (31)*

8. Winter **Solar** Observatory - Located in Roeland Park, Kansas and contains images from a high resolution CCD camera that is captured directly into a computer or stored on 8mm video tape.
   -- *http://icstars.com/HTML/icstars/observ.htm* *Science: Astronomy: Amateur: Solar System Observing: Sun (6)*

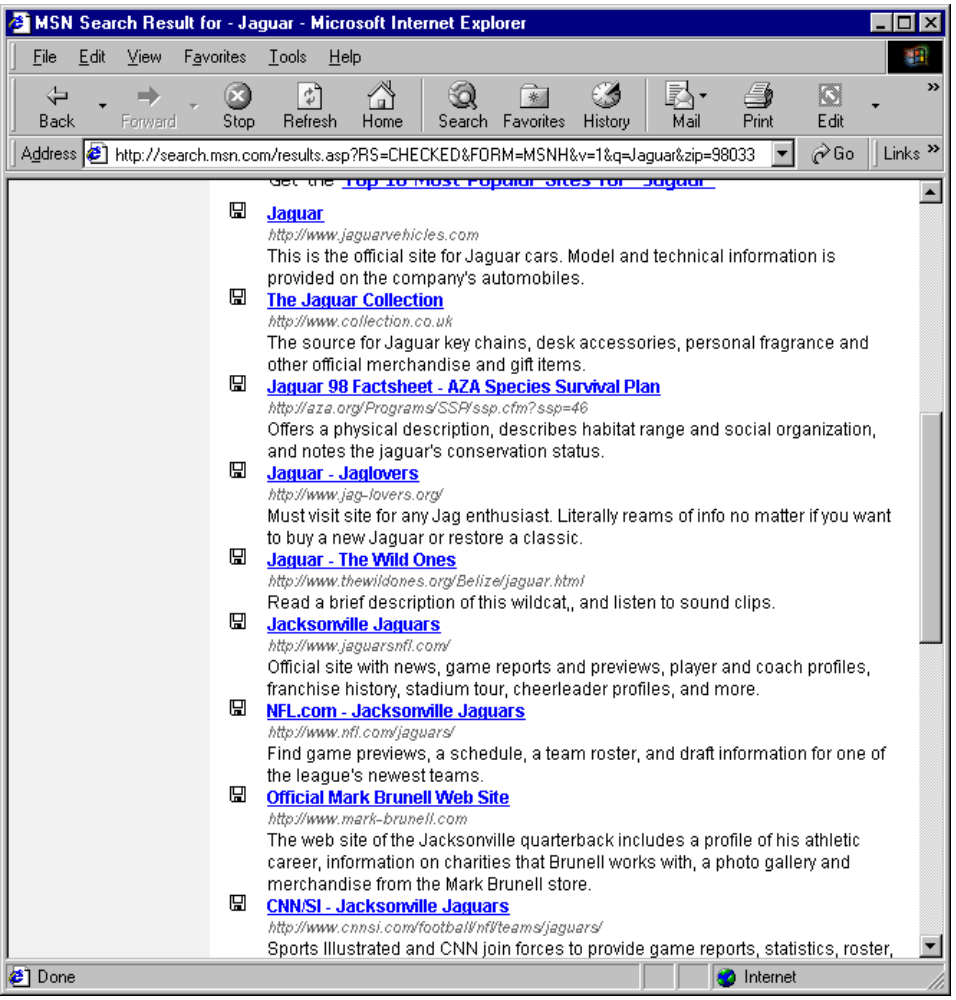
9. **Solar** Keratosis - Patient fact sheet explains causes, appearance, diagnosis, risks, and treatment.
   -- *http://www.patient.co.uk/showdoc/27000241/* *Health: Conditions and Diseases: Skin Disorders: Solar Keratosis (2)*

http://www.dmoz.org

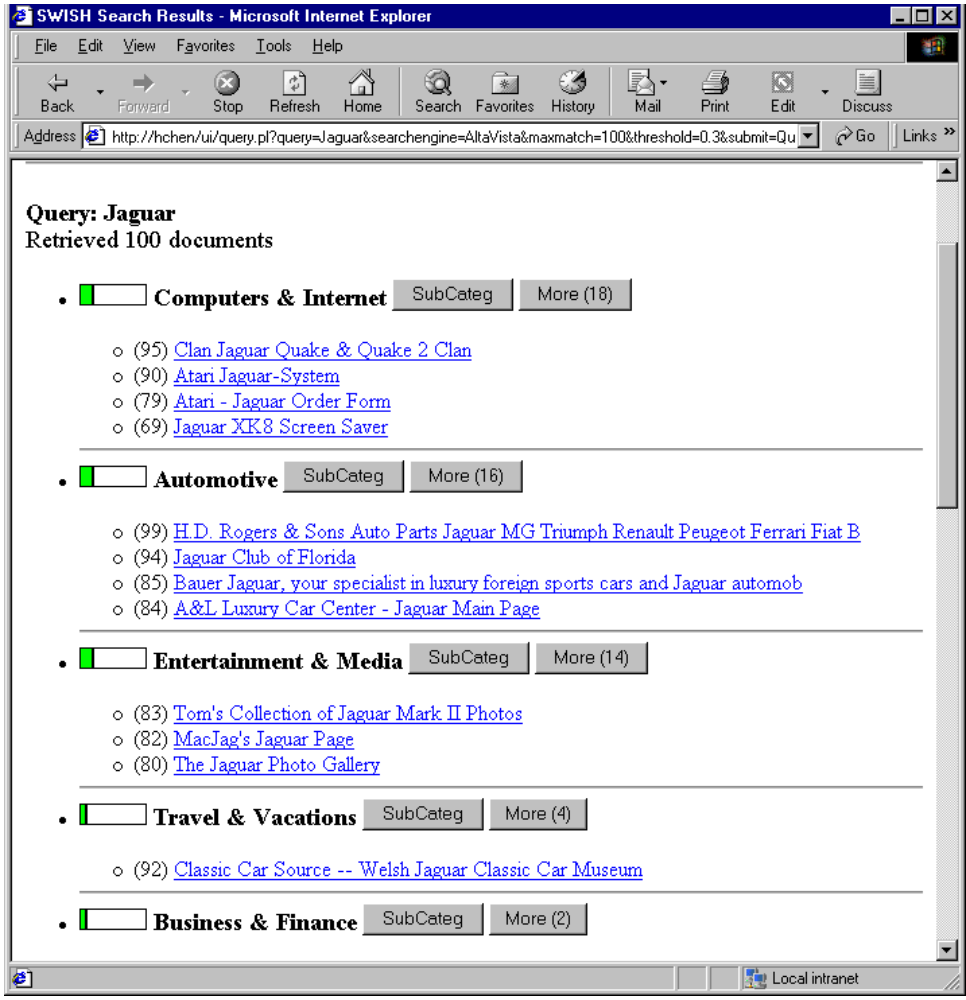
Query: jaguar

SWISH

List Interface



Category Interface

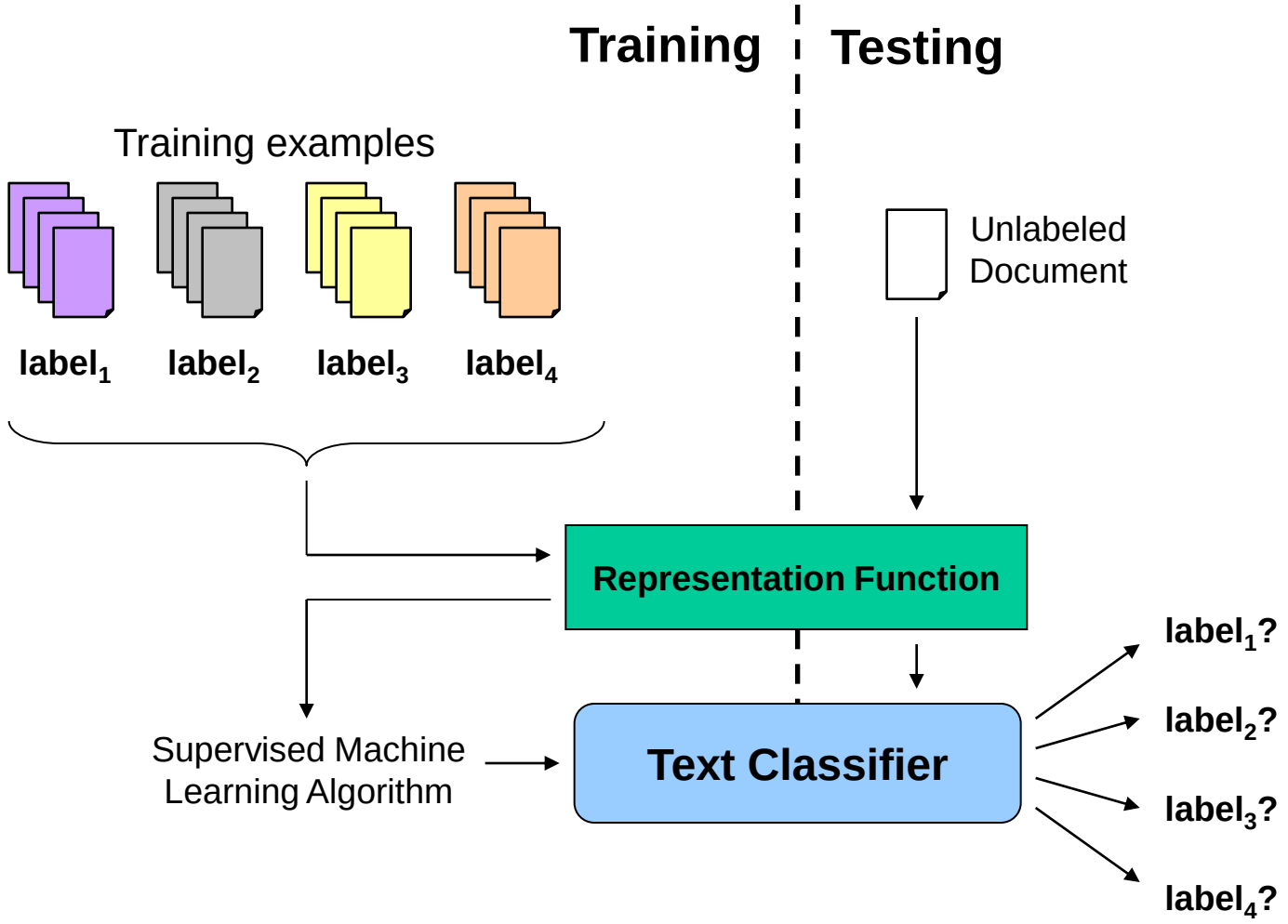


Hao Chen and Susan Dumais. (2000) Bringing Order to the Web: Automatically Categorizing Search Results. Proceedings of CHI 2000.

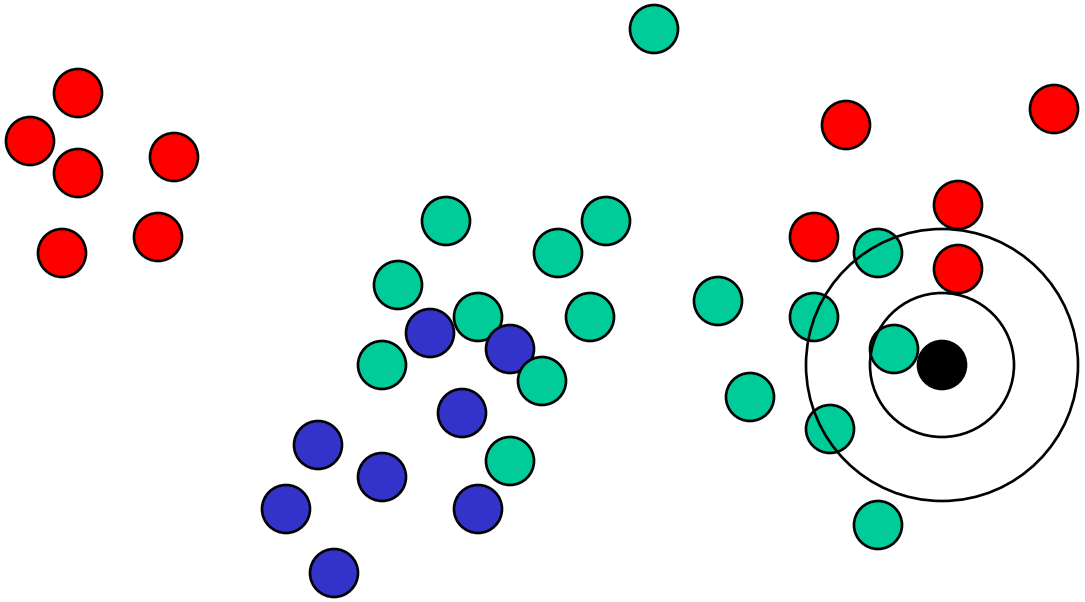
Text Classification

- Problem: automatically sort items into bins
- Machine learning approach
  - Obtain a training set with ground truth labels
  - Use a machine learning algorithm to “train” a classifier
    - kNN, Bayesian classifier, SVMs, decision trees, etc.
  - Apply classifier to new documents
    - System assigns labels according to patterns learned in the training set

Machine Learning



k Nearest Neighbor (*k*NN) Classifier



*k*NN Algorithm

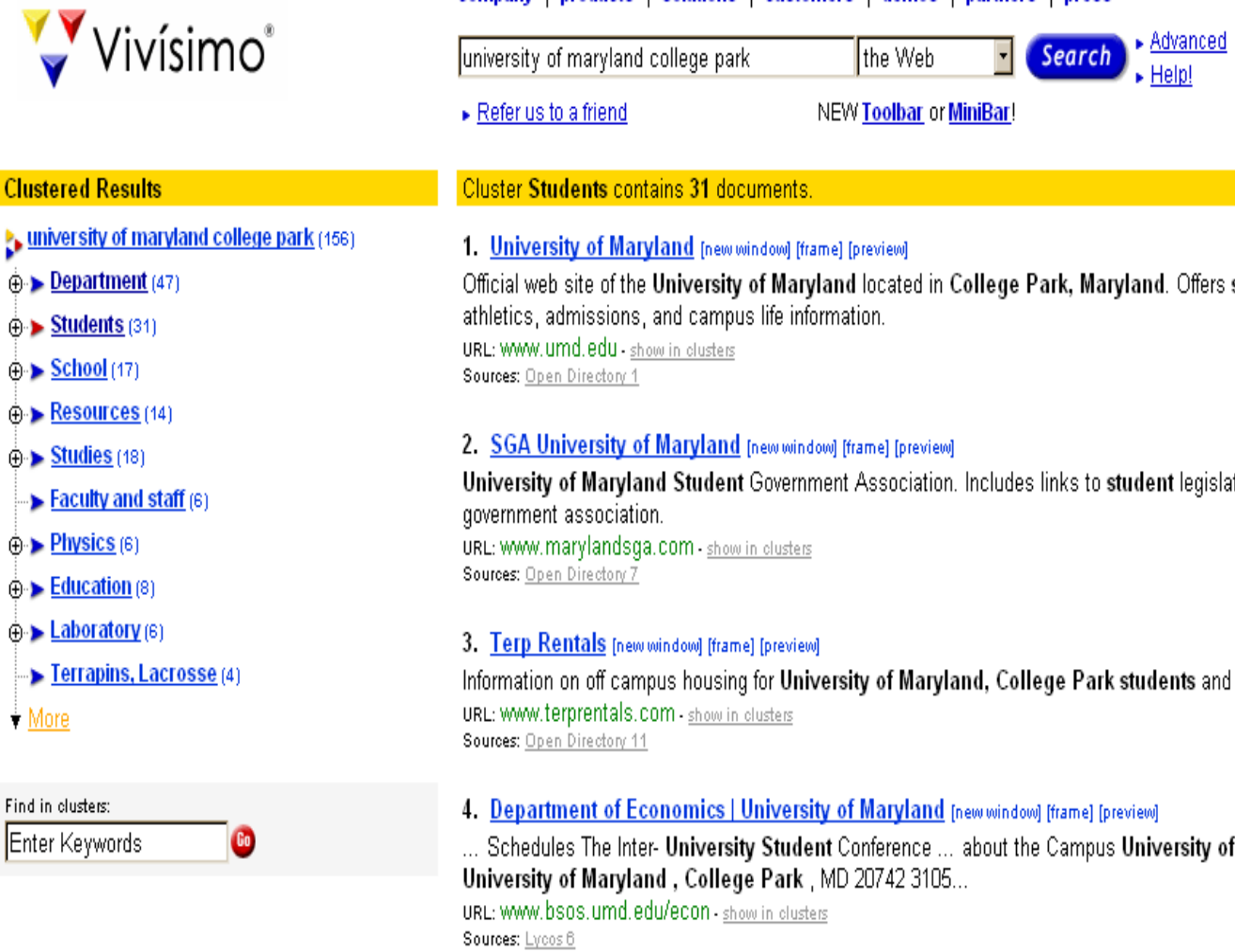
- Select *k* most similar labeled documents
- Have them "vote" on the best label:
  - Each document gets one vote, or
  - More similar documents get a larger vote
- How can similarity be defined?

The Cluster Hypothesis

"Closely associated documents tend to be relevant to the same requests."

van Rijsbergen 1979

Vivisimo: Clustered Results



http://www.vivisimo.com

Kartoo's Cluster Visualization

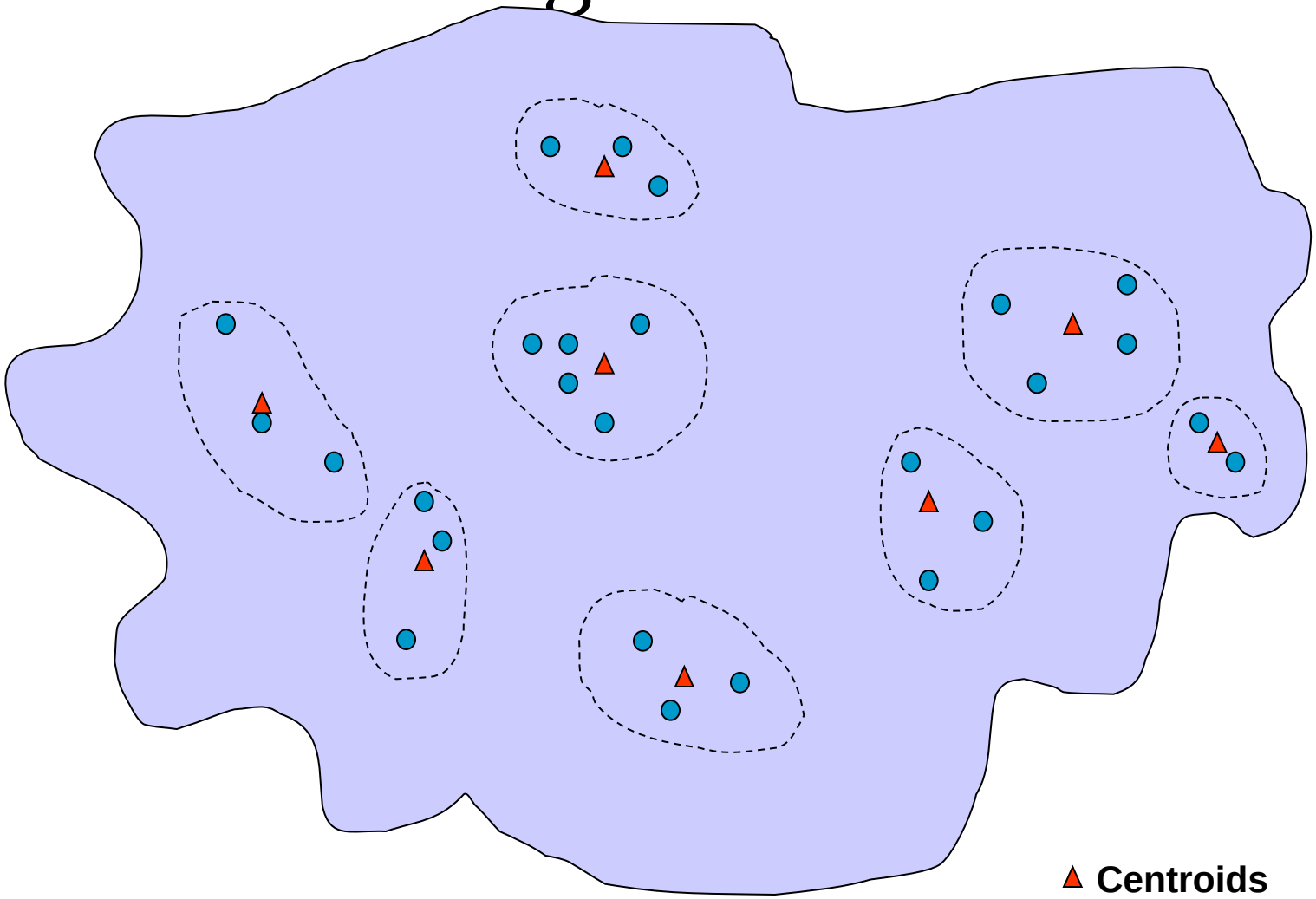
http://www.kartoo.com/

Clustering Result Sets

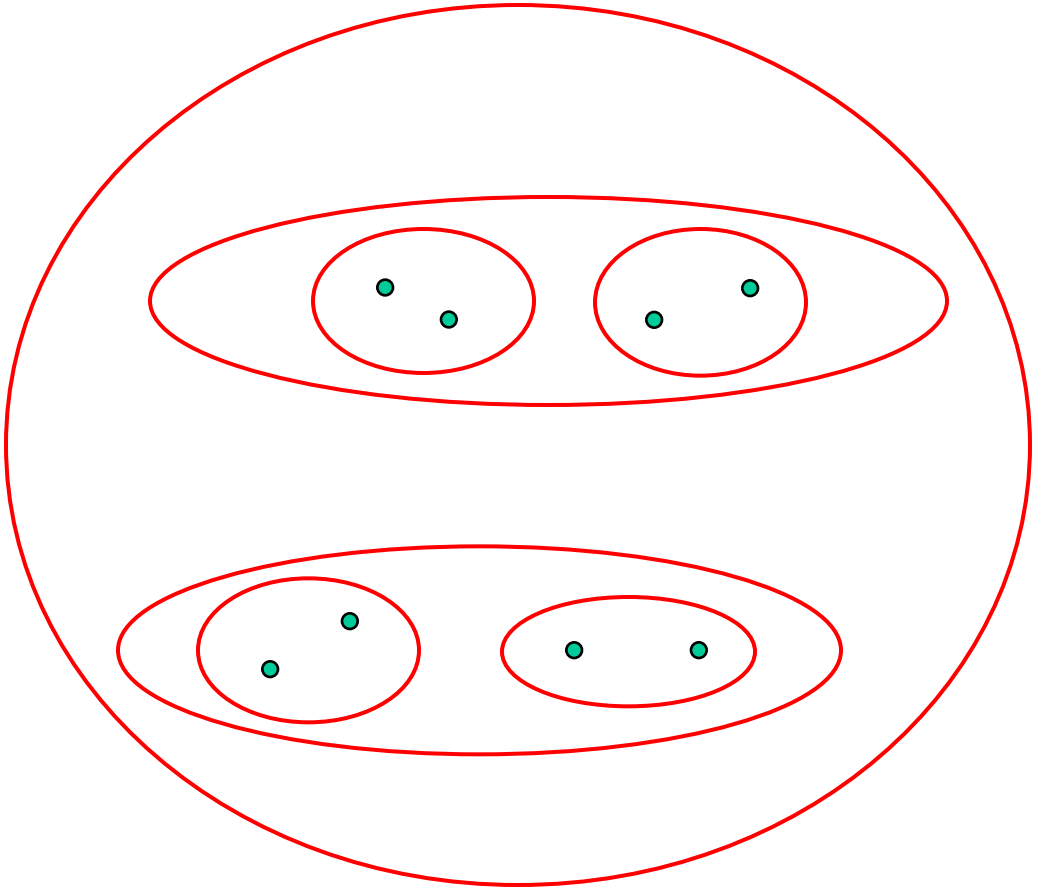
- Advantages:
  - Topically coherent document sets are presented together
  - User gets a sense for the themes in the result set
  - Supports browsing retrieved hits

- Disadvantages:
  - May be difficult to understand the theme of a cluster based on summary terms
  - Clusters themselves might not “make sense”
  - Computational cost

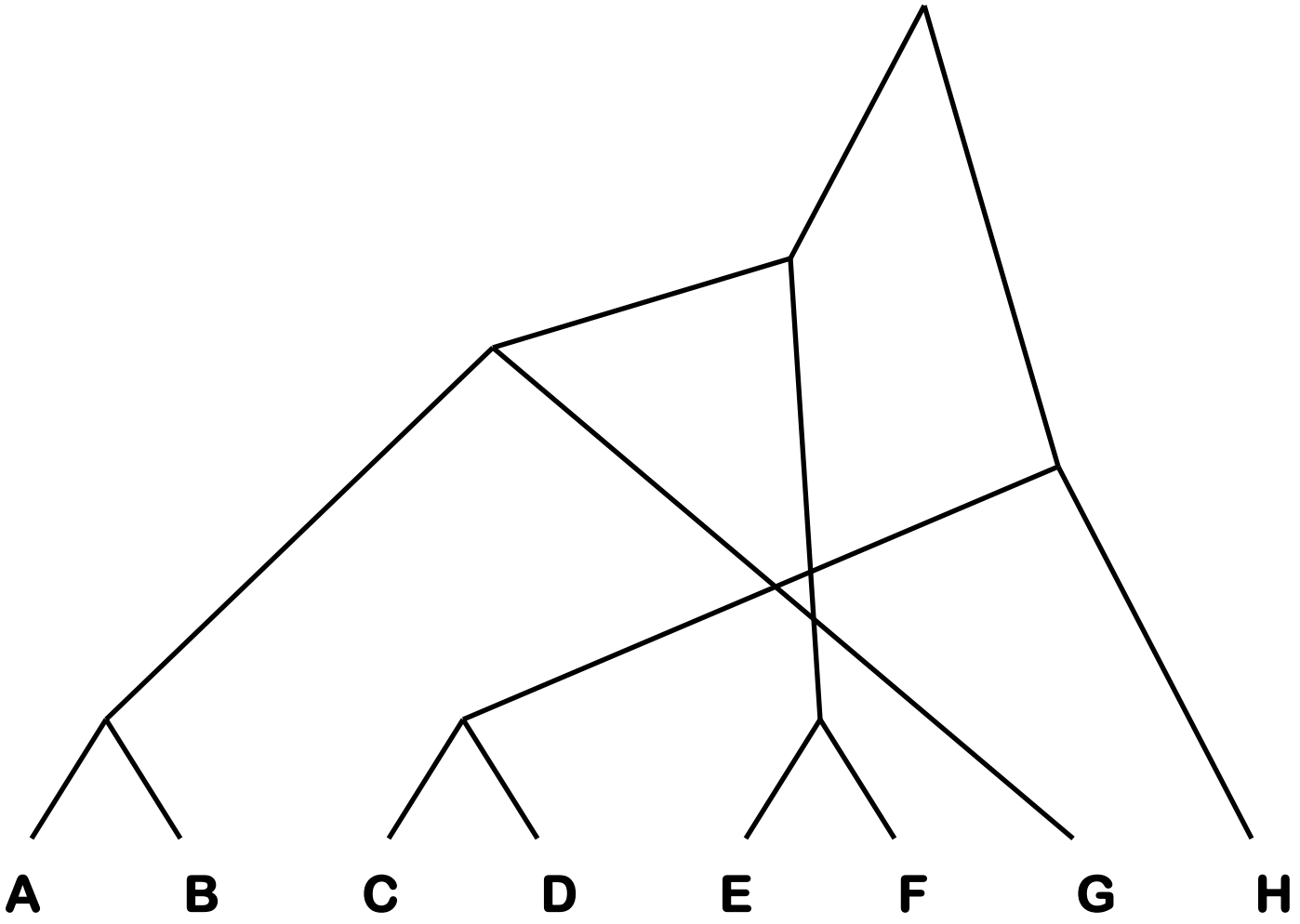
Visualizing Clusters



Hierarchical Agglomerative Clustering



Another Way to Look at H.A.C.



The H.A.C. Algorithm

- Start with each document in its own cluster
- Until there is only one cluster:
  - Determine the two most similar clusters c_i and c_j
  - Replace c_i and c_j with a single cluster $c_i \cup c_j$
- The history of merging forms the hierarchy

Cluster Similarity

- Assume a similarity function that determines the similarity of two instances: sim(x,y)
  - What's appropriate for documents?

- What's the similarity between two clusters?
  - Single Link: similarity of two most similar members
  - Complete Link: similarity of two least similar members
  - Group Average: average similarity between members

Scatter/Gather

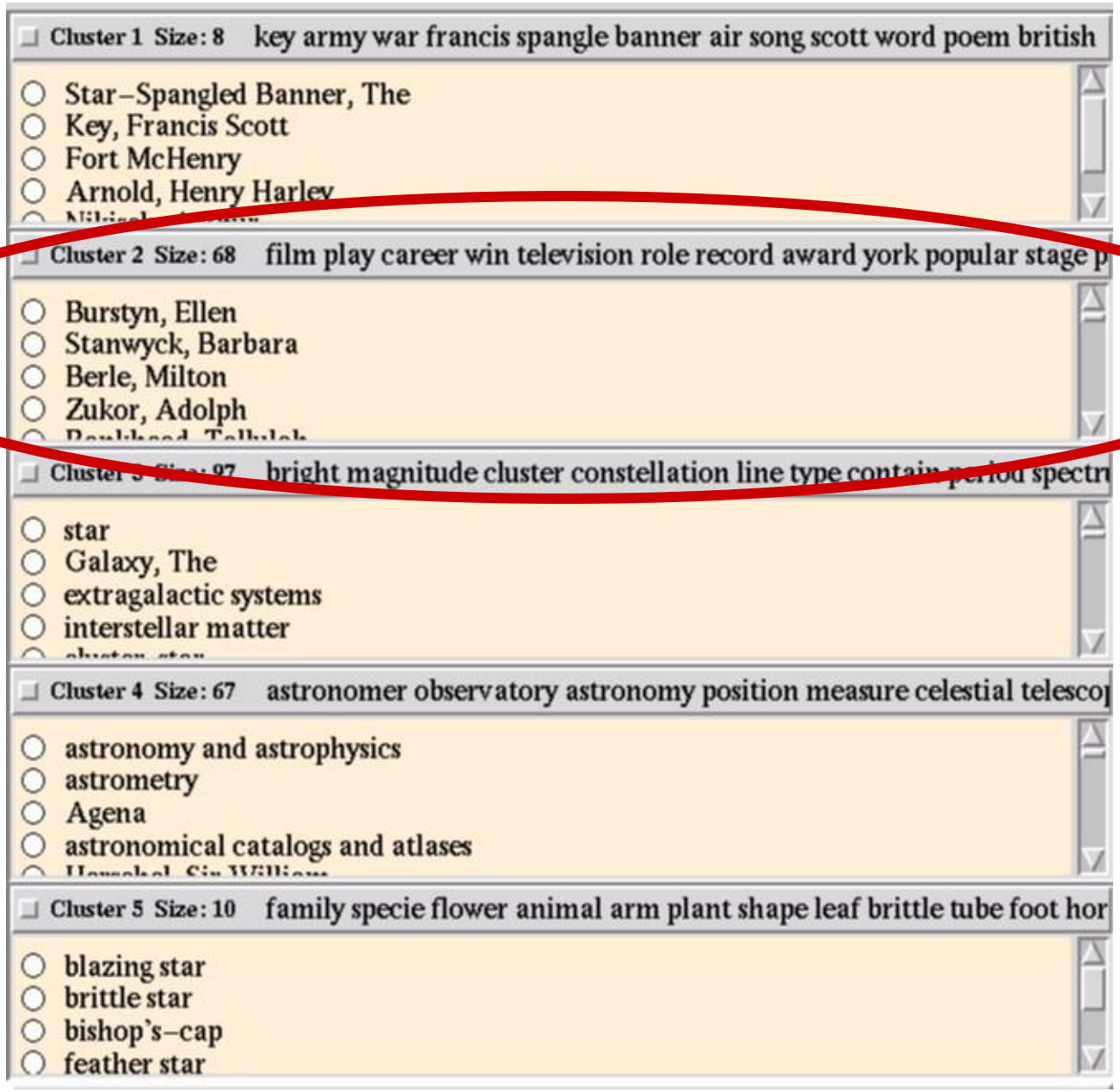
Query = “star” on encyclopedic text

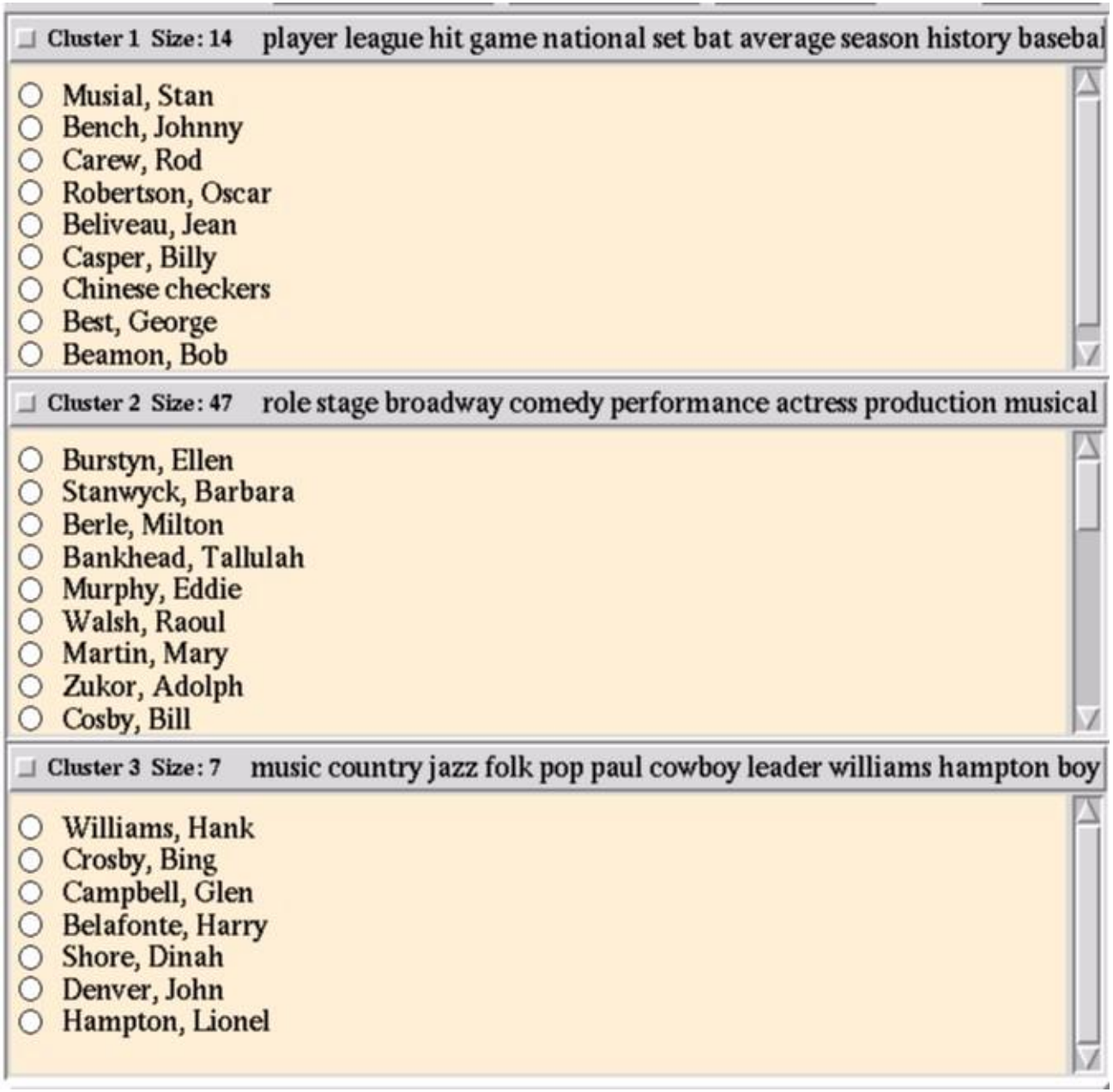
Clustering and re-clustering is entirely automated

Scatter/Gather

- Sustem clusters documents into “themes”
  - Displays clusters by showing:
    - Topical terms
    - Typical titles
- User chooses a subset of the clusters
- System re-clusters documents in selected cluster
  - New clusters have different, more refined, “themes”

Marti A. Hearst and Jan O. Pedersen. (1996) Reexaming the Cluster Hypothesis: Scatter/Gather on Retrieval Results. Proceedings of SIGIR 1996.


Cluster 1 Size: 8 key army war francis spangle banner air song scott word poem british
Star–Spangled Banner, The
Key, Francis Scott
Fort McHenry
Arnold, Henry Harley
Cluster 2 Size: 68 film play career win television role record award york popular stage p
Burstyn, Ellen
Stanwyck, Barbara
Berle, Milton
Zukor, Adolph
bright magnitude cluster constellation line type
star
Galaxy, The
extragalactic systems
interstellar matter
Cluster 4 Size: 67 astronomer observatory astronomy position measure celestial telesco
astronomy and astrophysics
astrometry
Agena
astronomical catalogs and atlases
Cluster 5 Size: 10 family specie flower animal arm plant shape leaf brittle tube foot hor
blazing star
brittle star
bishop's–cap
feather star


Cluster 1 Size: 14 player league hit game national set bat average season history baseba
Musial, Stan
Bench, Johnny
Carew, Rod
Robertson, Oscar
Beliveau, Jean
Casper, Billy
Chinese checkers
Best, George
Beamon, Bob
Cluster 2 Size: 47 role stage broadway comedy performance actress production musical
Burstyn, Ellen
Stanwyck, Barbara
Berle, Milton
Bankhead, Tallulah
Murphy, Eddie
Walsh, Raoul
Martin, Mary
Zukor, Adolph
Cosby, Bill
Cluster 3 Size: 7 music country jazz folk pop paul cowboy leader williams hampton boy
Williams, Hank
Crosby, Bing
Campbell, Glen
Belafonte, Harry
Shore, Dinah
Denver, John
Hampton, Lionel

Summary: Clustering

- Advantages:
  - Provides an overview of main themes in search results
  - Helps overcome polysemy

- Disadvantages:
  - Documents can be clustered in many ways
  - Not always easy to understand the theme of a cluster
  - What is the correct level of granularity?
  - More information to present

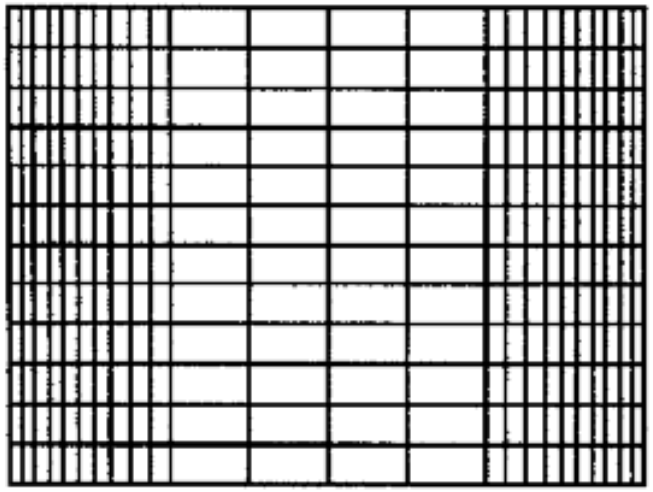
Recap

- Clustering
  - Automatically group documents into clusters
- Classification
  - Automatically assign labels to documents

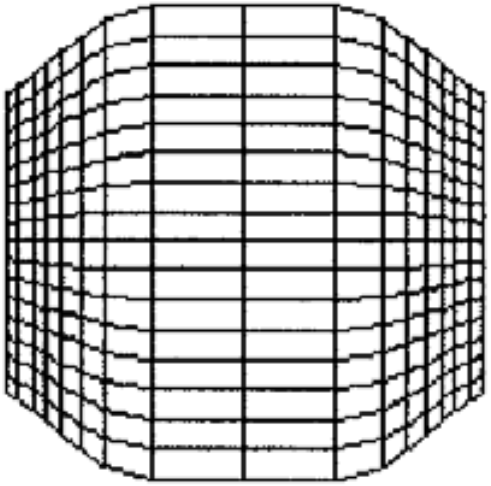
Agenda

- Query formulation
- Selection
- ➢ Examination
- Source selection

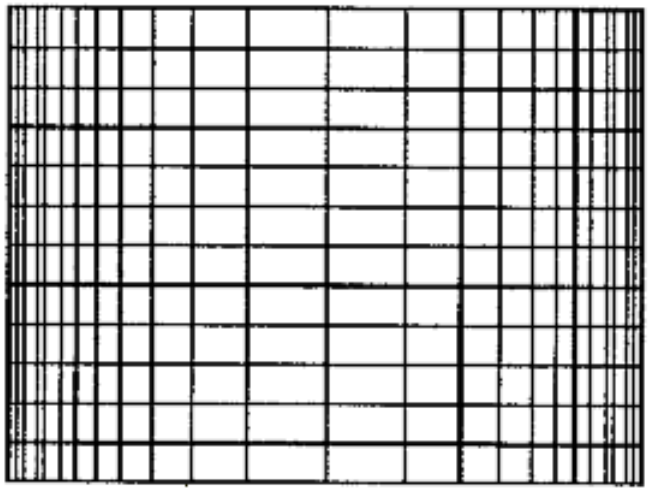
Distorting Reality



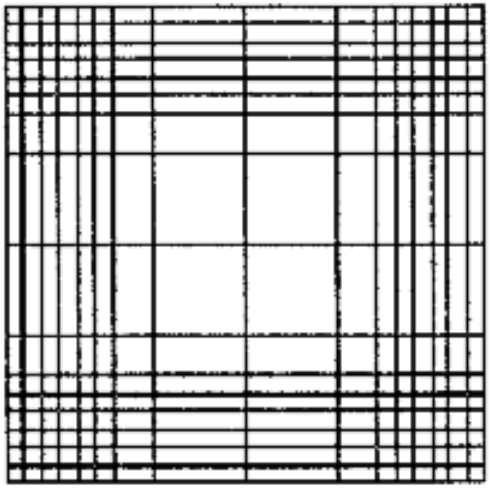
Bifocal

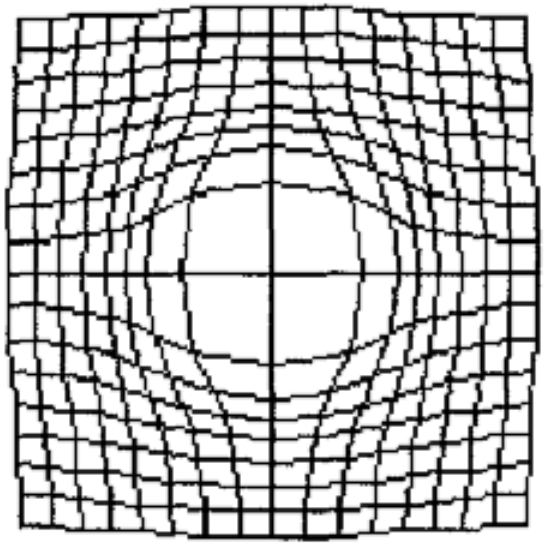


Perspective Wall

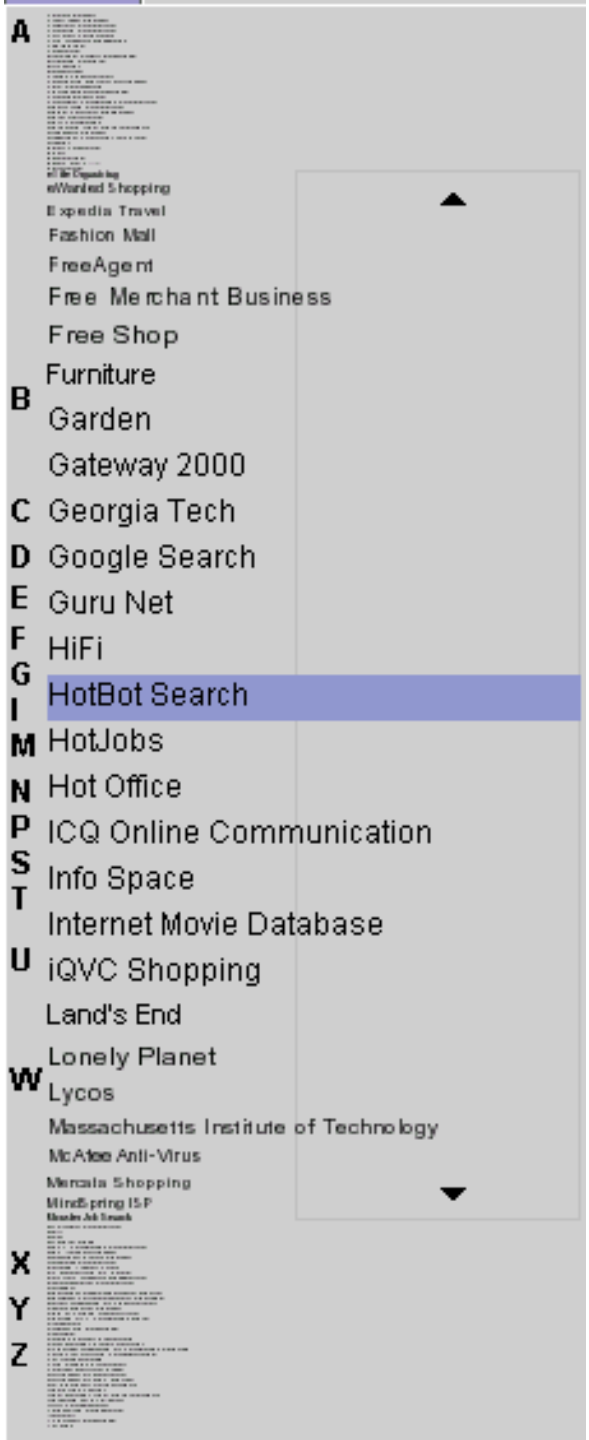


Fisheye



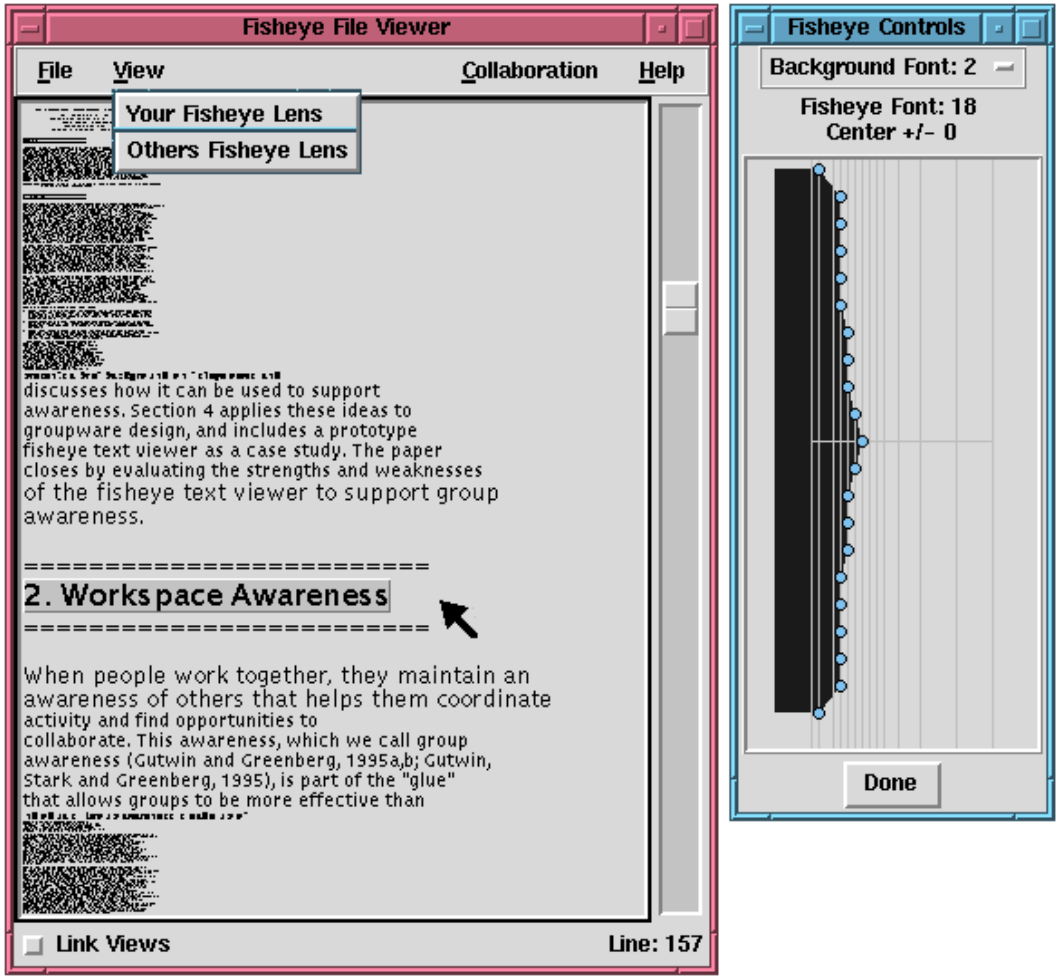


1-D Fisheye Menu



http://www.cs.umd.edu/hcil/fisheyemenu/fisheyemenu-demo.shtml

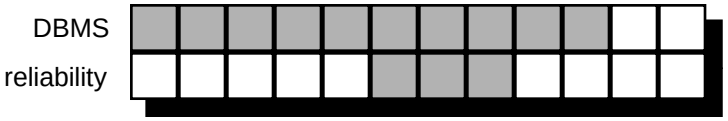
1-D Fisheye Document Viewer



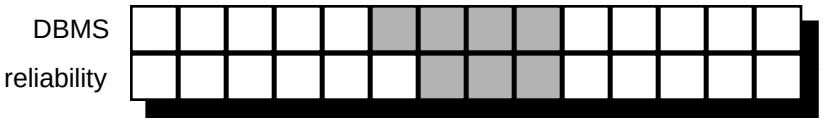
TileBars

Topic: reliability of DBMS (database systems)
Query terms: DBMS, reliability

Mainly about both DBMS and reliability



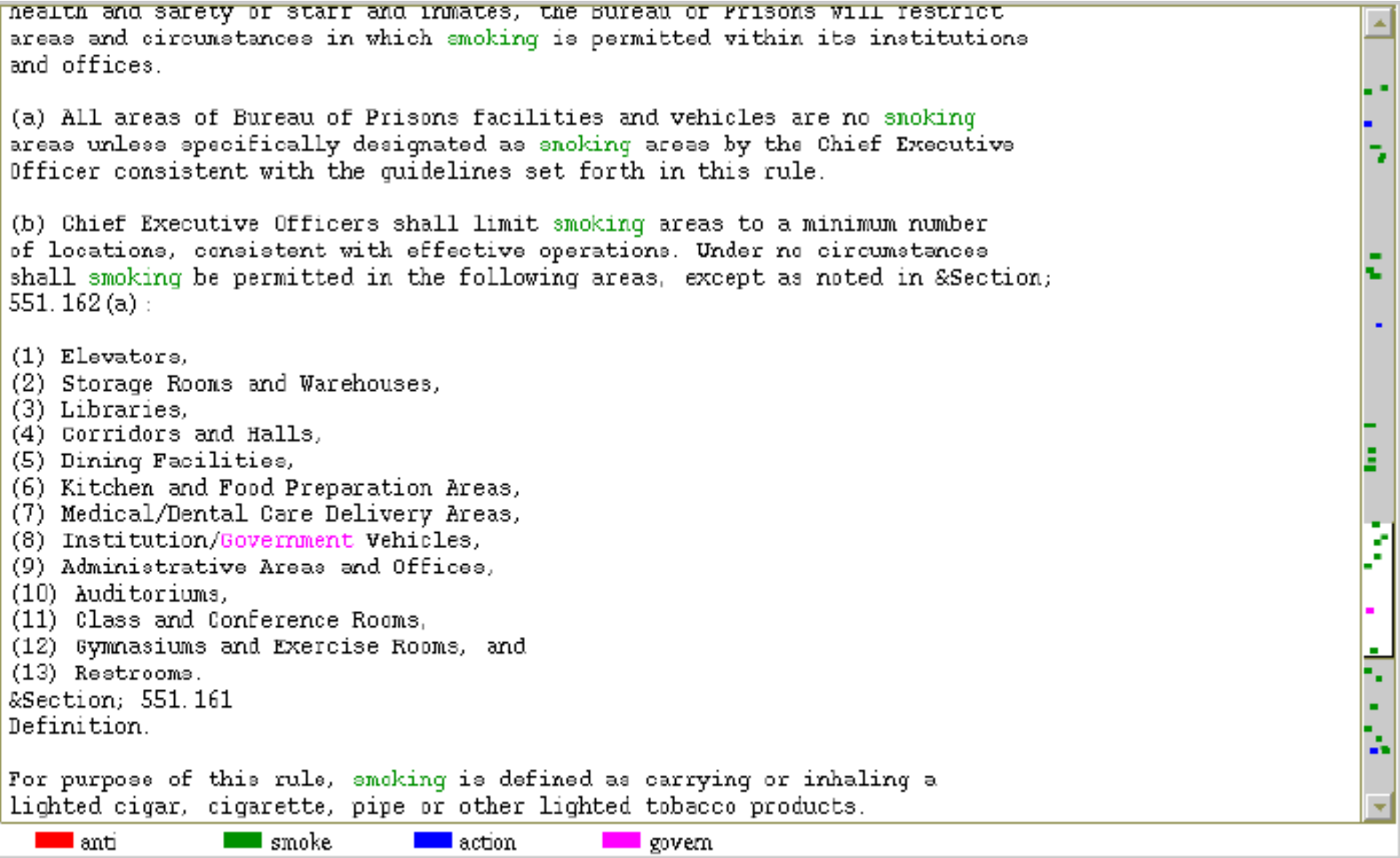
Mainly about DBMS, discusses reliability



Mainly about, say, banking, with a subtopic discussion on DBMS/Reliability

Mainly about high-tech layoffs

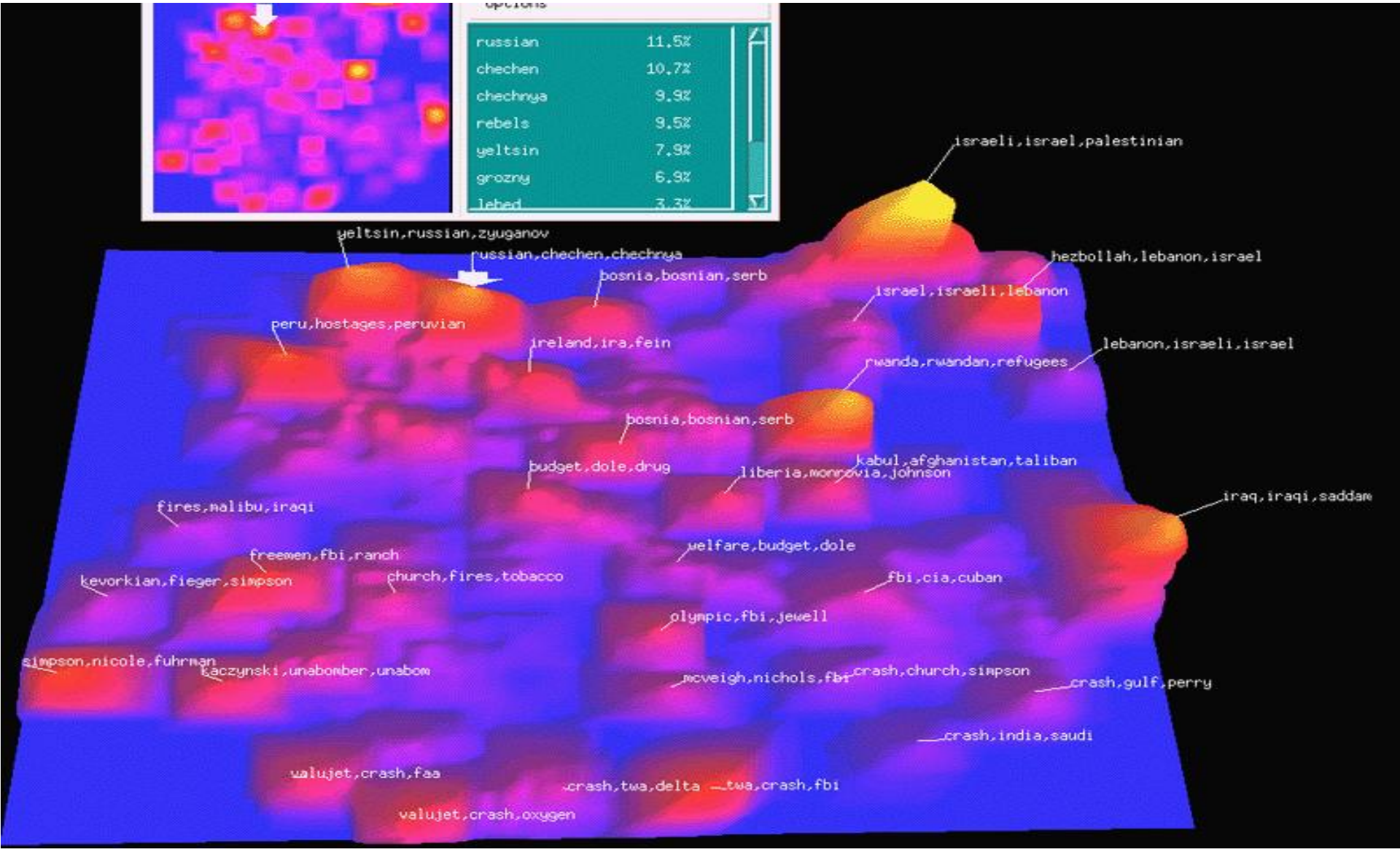
U Mass: Scrollbar-Tilebar



Agenda

- Query formulation
- Selection
- Examination
- ➢ Source selection

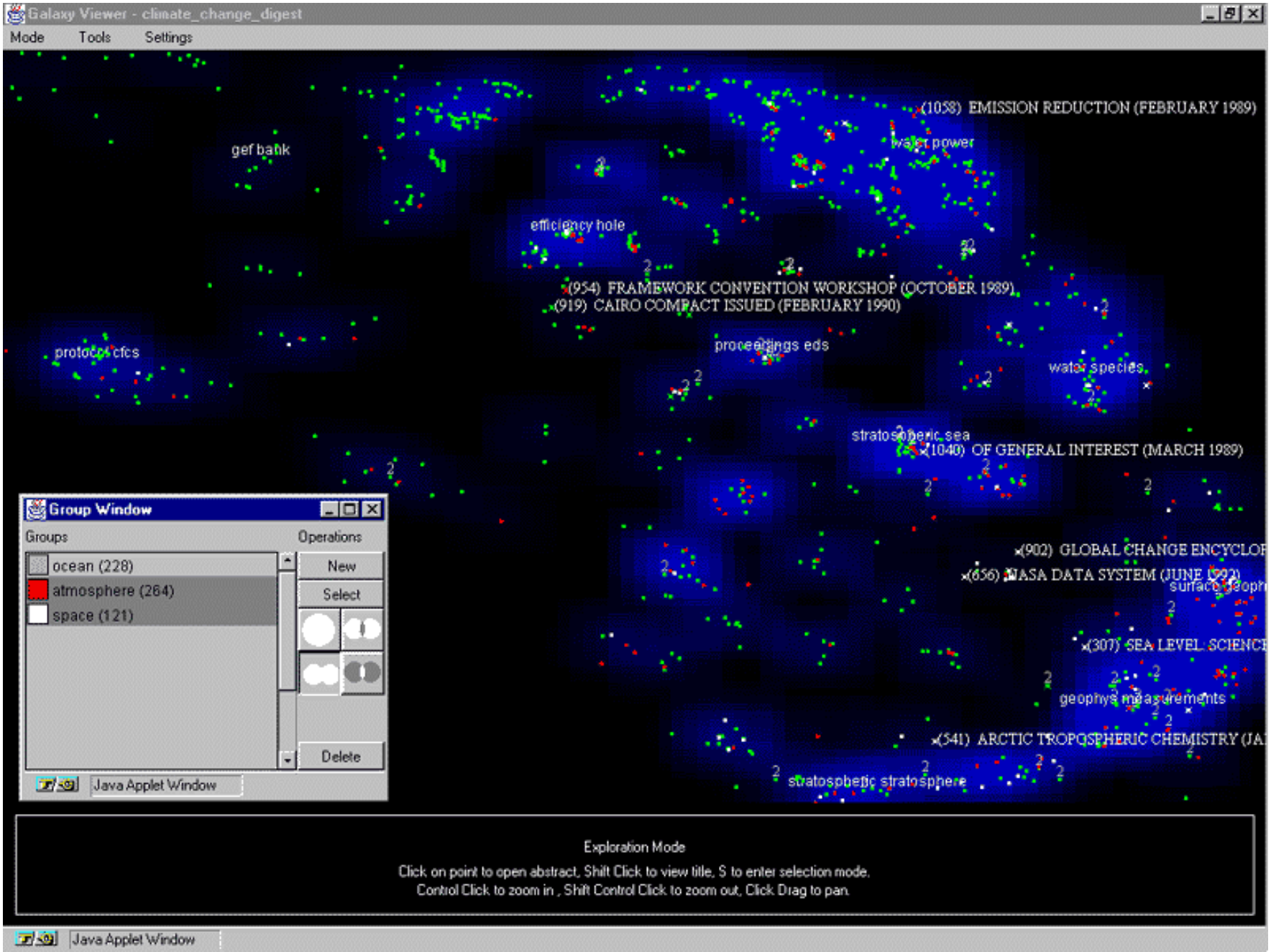
ThemeView



http://www.pnl.gov/infoviz/technologies.html

Pacific Northwest National Laboratory

WebTheme


Galaxy Viewer - climate_change_digest
Mode
Tools
Settings
(1058) EMISSION REDUCTION (FEBRUARY 1989)
water power
gef bank
efficiency hole
(954) FRAMEWORK CONVENTION WORKSHOP (OCTOBER 1989)
(919) CAIRO COMPACT ISSUED (FEBRUARY 1990)
proceedings eds
water species
stratospheric sea
(1040) OF GENERAL INTEREST (MARCH 1989)
(902) GLOBAL CHANGE ENCYCLOP
(656) NASA DATA SYSTEM (JUNE
surface geopr
(307) SEA LEVEL SCIENCE
geophys measurements
(541) ARCTIC TROPOSPHERIC CHEMISTRY (JA
stratospheric stratosphere
Group Window
Groups
Operations
ocean (228)
atmosphere (264)
space (121)
New
Select
Delete
Java Applet Window
Exploration Mode
Click on point to open abstract, Shift Click to view title, S to enter selection mode.
Control Click to zoom in , Shift Control Click to zoom out, Click Drag to pan.
Java Applet Window

Ben S' 'Seamless Interface' Principles

- Informative feedback
- Easy reversal
- User in control
  - Anticipatable outcomes
  - Explainable results
  - Browsable content
- Limited working memory load
  - Query context
  - Path suspension
- Alternatives for novices and experts
    - Scaffolding

My ‘Synergistic Interaction’ Principles

- Interdependence with process (“interaction models”)
  - Co-design with search strategy
  - Speed
- System initiative
  - Guided process
  - Exposing the structure of knowledge
- Support for reasoning
  - Representation of uncertainty
  - Meaningful dimensions
- Synergy with features used for search
  - Weakness of similarity, Strength of language
- Easily learned
  - Familiar metaphors (timelines, ranked lists, maps)

Some Good Ideas

- Show the query in the selection interface
  - It provides context for the display
- Suggest options to the user
  - Query refinements, for example
- Explain what the system has done
  - Highlight query terms in the results, for example
- Complement what the system has done
  - Users add value by doing things the system can't
  - Expose the information users need to judge utility

One Minute Paper

- When examining documents in the selection and examination interfaces, which type of information need (visceral, conscious, formalized, or compromised) guides the user's decisions? Please justify your answer.