

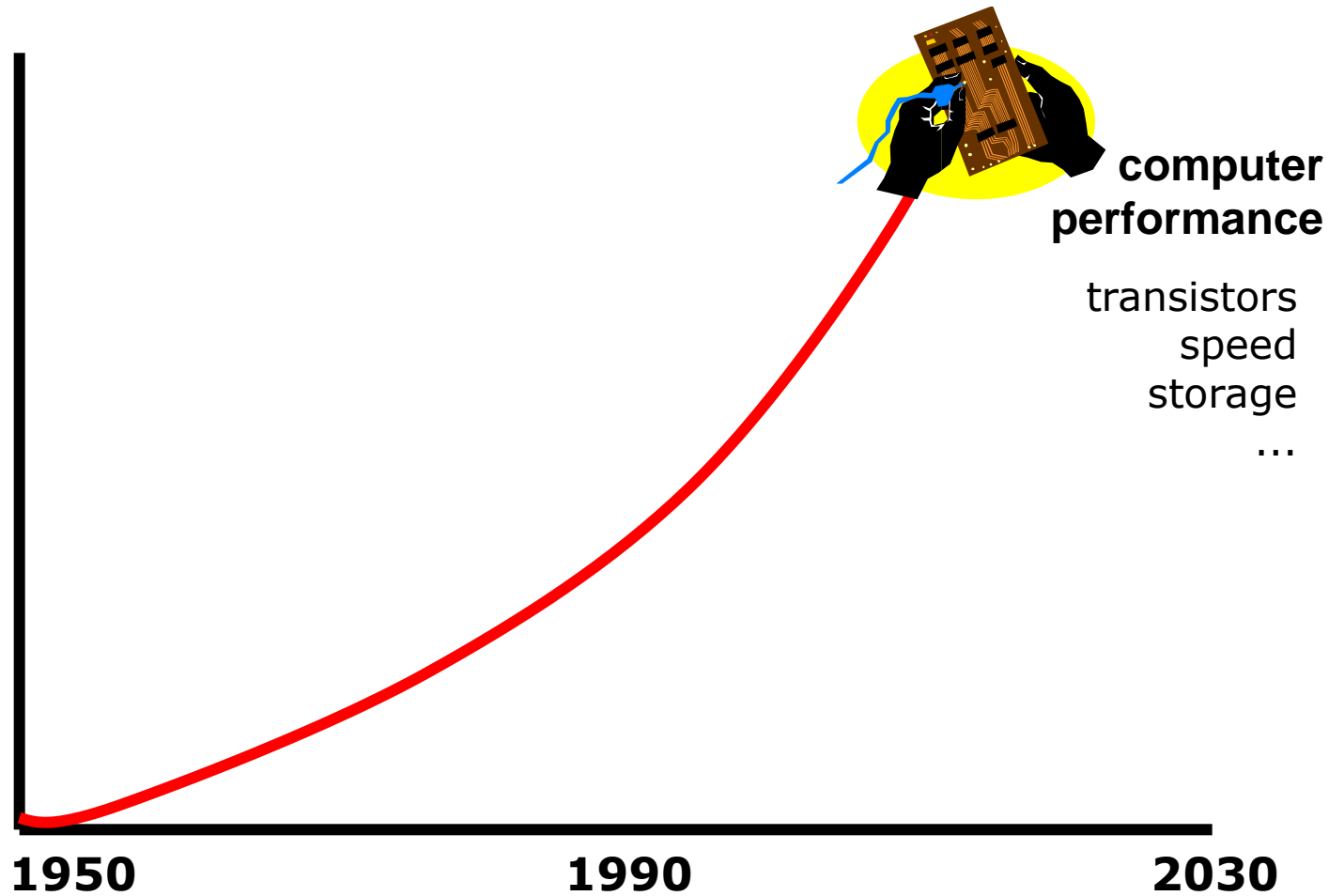
Interaction

LBSC 796/INFM 718R

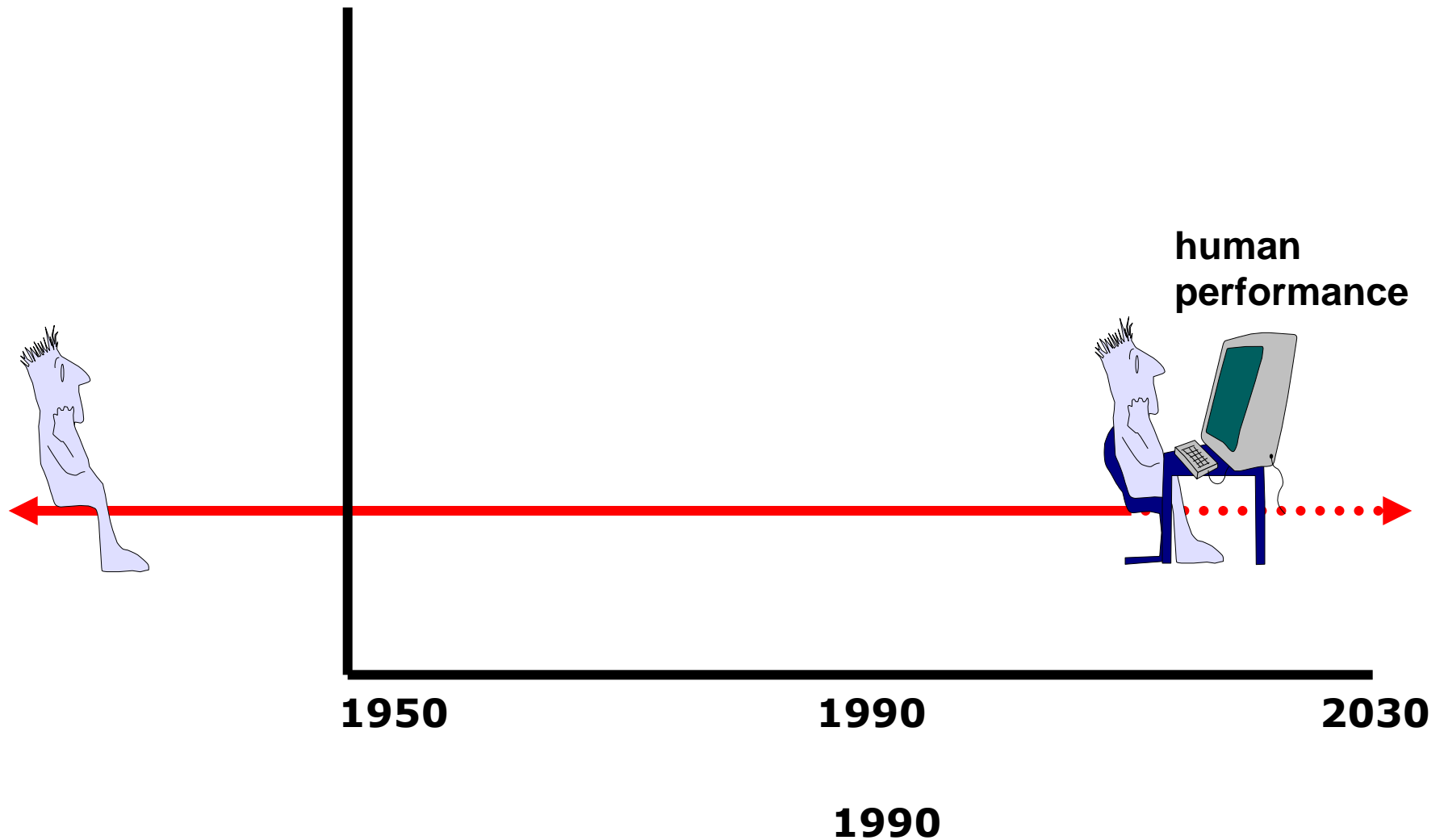
Douglas W. Oard

Week 4, February 23, 2011

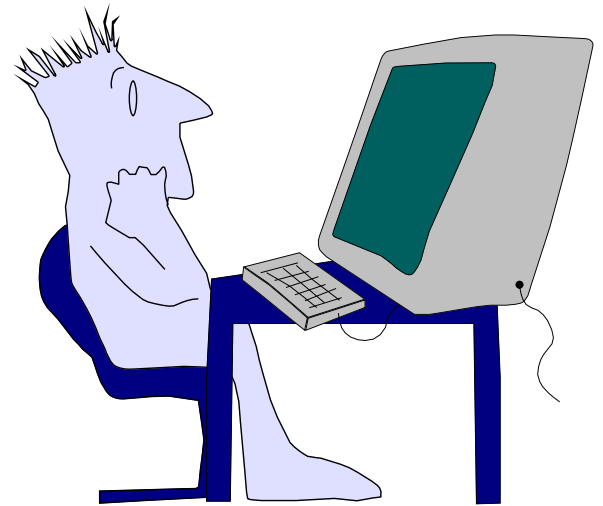
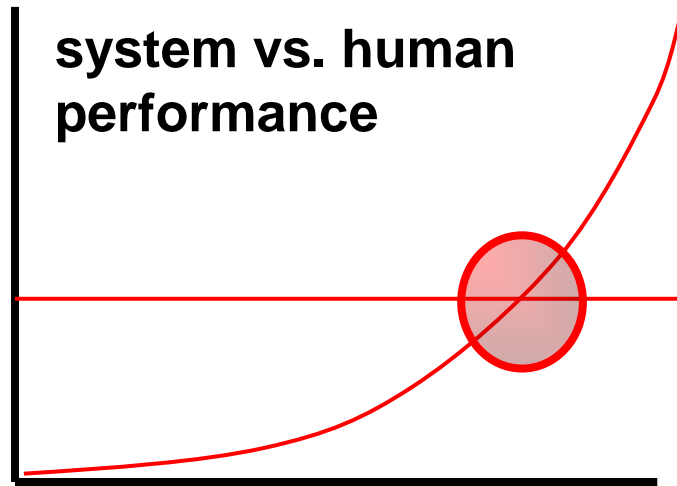
Moore's Law



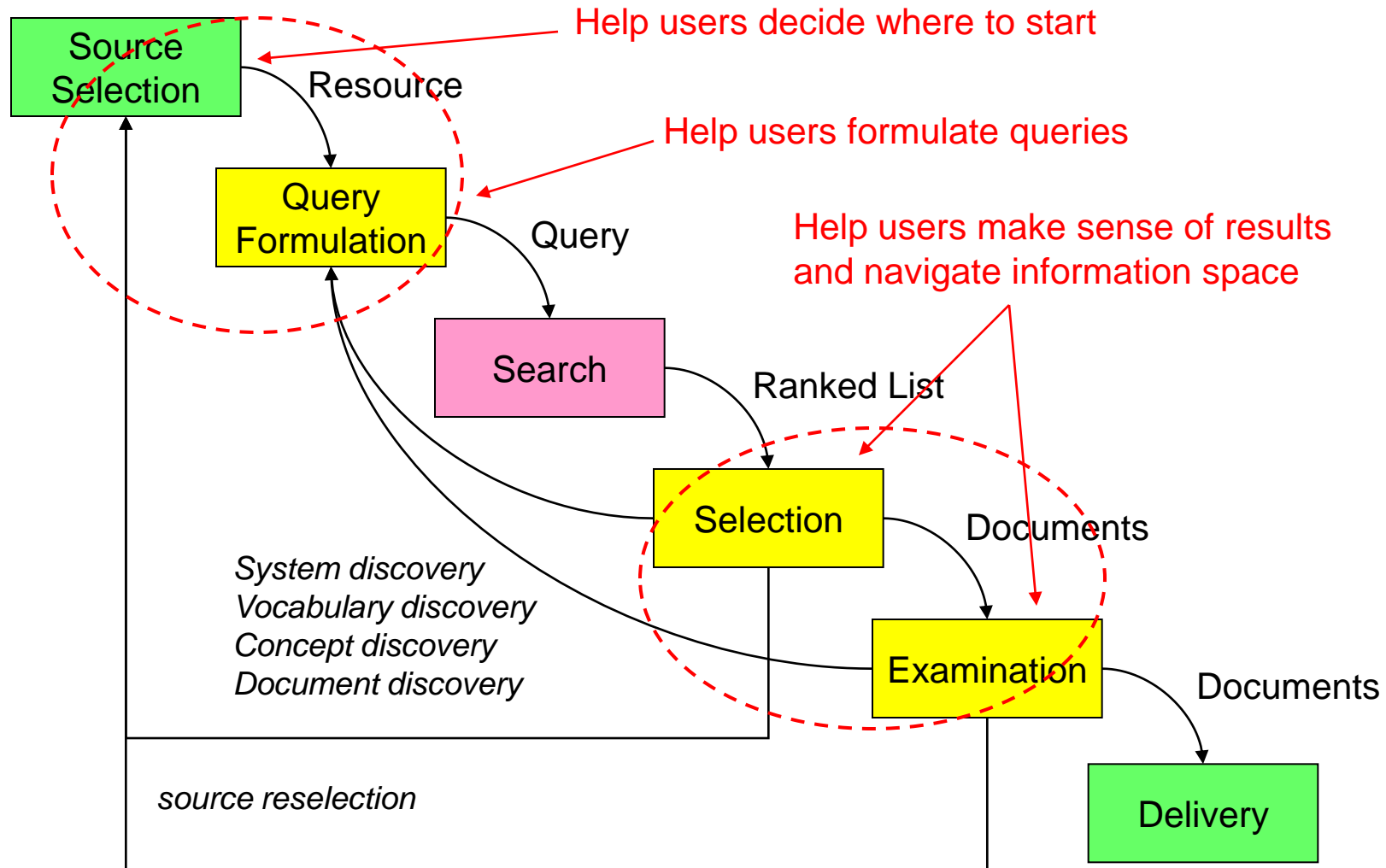
Human Cognition



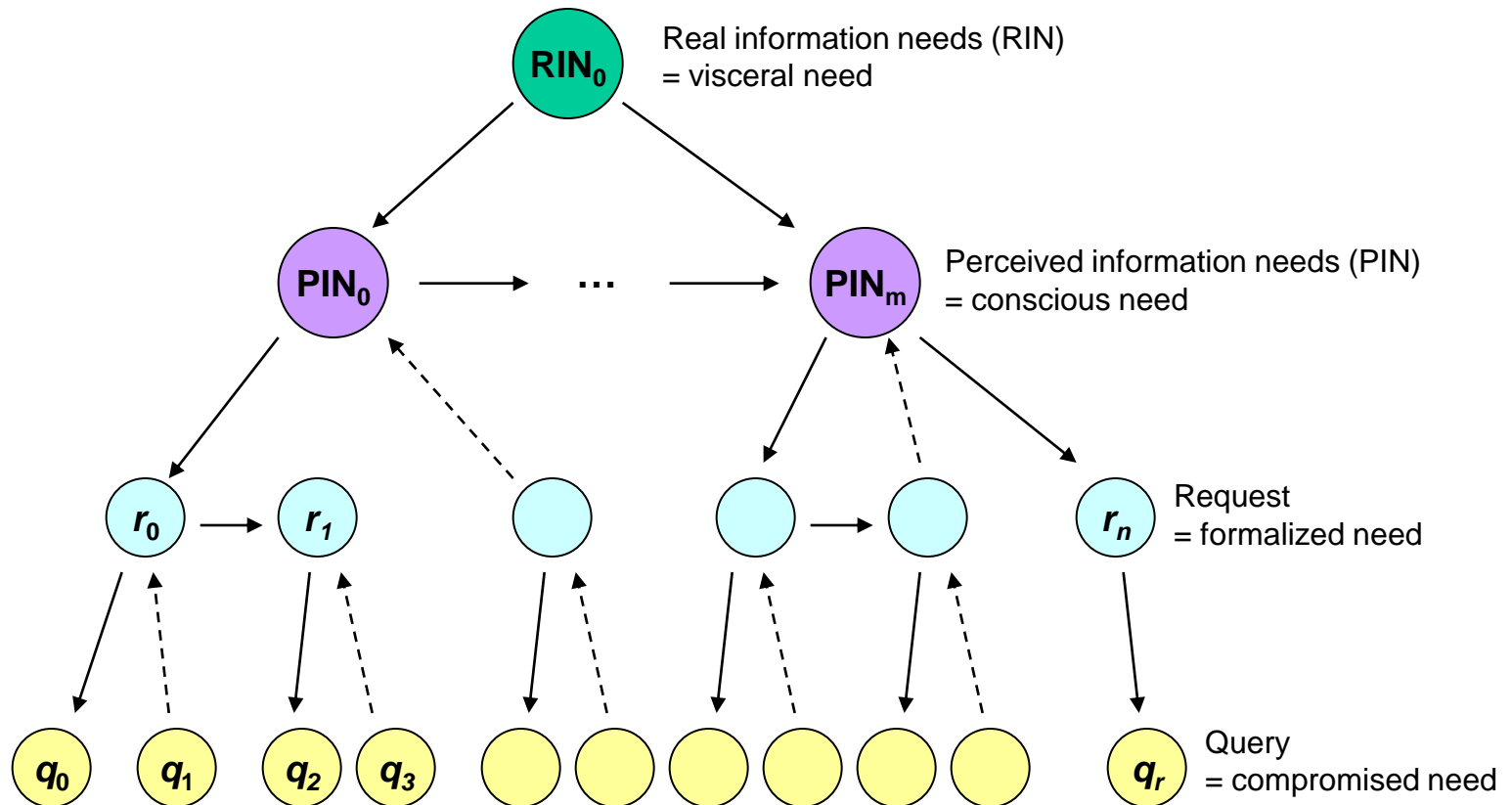
Where is the bottleneck?



Interaction Points



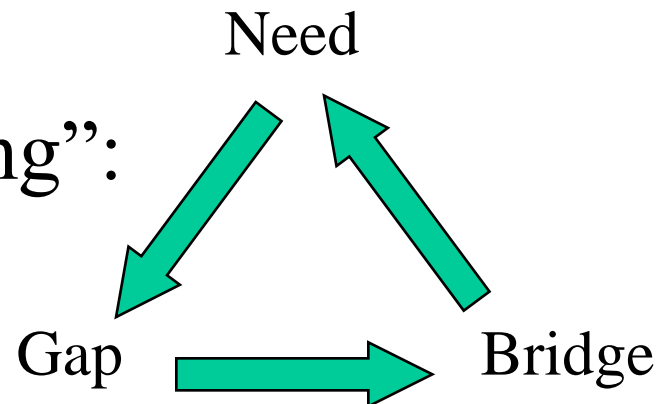
Information Needs



Anomalous State of Knowledge

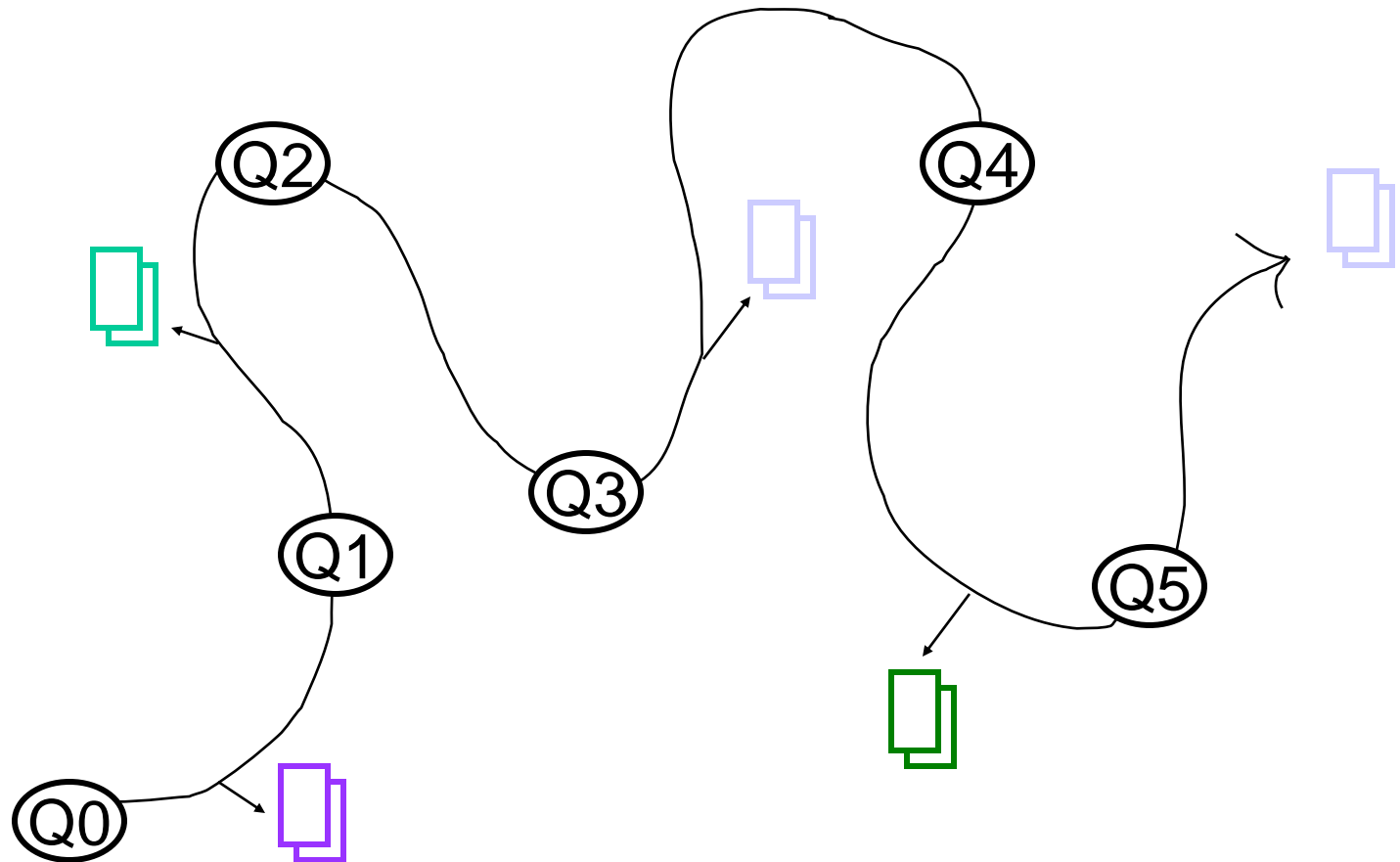
- Belkin: Searchers do not clearly understand
 - The problem itself
 - What information is needed to solve the problem
- The query results from a clarification process

- 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”)

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

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

Bookmarks Location: <http://192.35.215.185/mw/mwcgi.home>

 **MELVYL**[®]
system

[New Search](#)
[Search History](#)
[Saved Lists](#)
[Profile](#)
[Resources](#)
[Restart](#)
[Quit](#)
[Help](#)

Database: **Current Contents** Personal Profile: **Off**

Author Search: Current Contents database

Author (e.g., jones, e d)

Options and Limits

Another Author (e.g., wilson, r)

Journal Title (e.g., daedalus or jama)

☒ Any words ☐ Exact beginning ☐ Complete title

Location

Send questions, comments, or suggestions to melvyl@www.melvyl.ucop.edu
Melvyl[®] is a registered trademark of The Regents of the University of California

Credit: Marti Hearst

Document: Done

Direct Manipulation Spec. VQUERY (Jones 98)

VQuery: Steve Jones 1998

Active query

Query 60 Boolean 60

Searching 57

Graphical 60 Browsing 60

Ranking 16

Retrieval 60 Keywords 60

Language 60

Visualization 60

Refinement 11

Enter new term

Collections

☒ HCI Bibliography

Search for any documents in "HCI Bibliography" containing either Query and Boolean; or Graphical, Searching and Browsing; but not Ranking

VQuery Results Preview

Sorted by Source

Keep selected for later

4 documents match the selected query

Graphical Presentation of Boolean Expressions in a	A. Michard
Query Processing in a Heterogeneous Retrieval Netw	Patricia Simpson
On Extending the Vector Space Model for Boolean Qu	S. K. M. Wong, W. Ziarko, U. U. Baghawan, P. C. N. Wong
A Direct Manipulation Interface for Boolean Inform	Peter G. Anick, Jeffrey D. Brennan, Rex A. Flynn, David

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

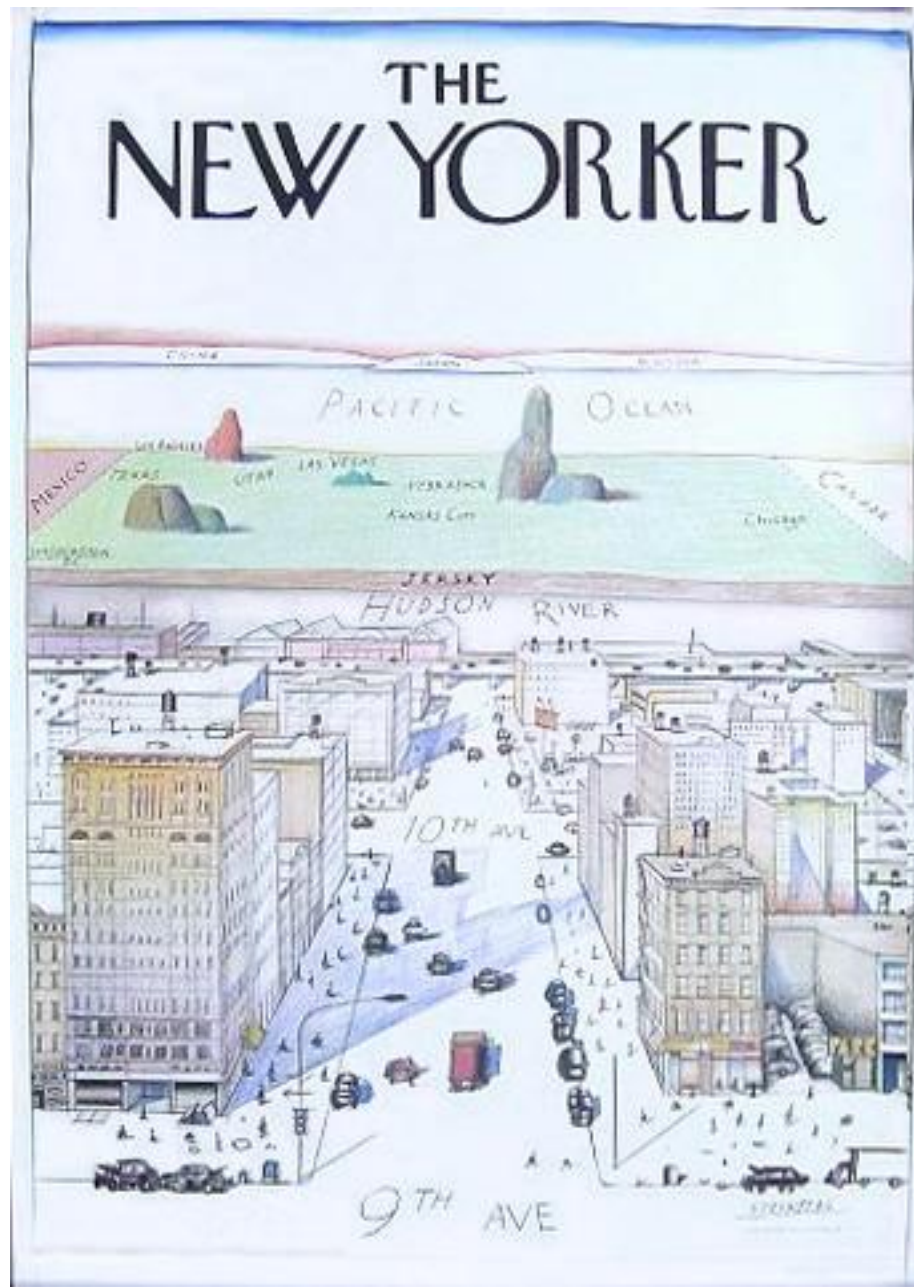
Young ... Solutions, has received the **college's** Distinguished Alumnus ...

Description: Offers BS, MS, and Ph.D. degrees in computer science. Research interests include artificial intelligence...

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

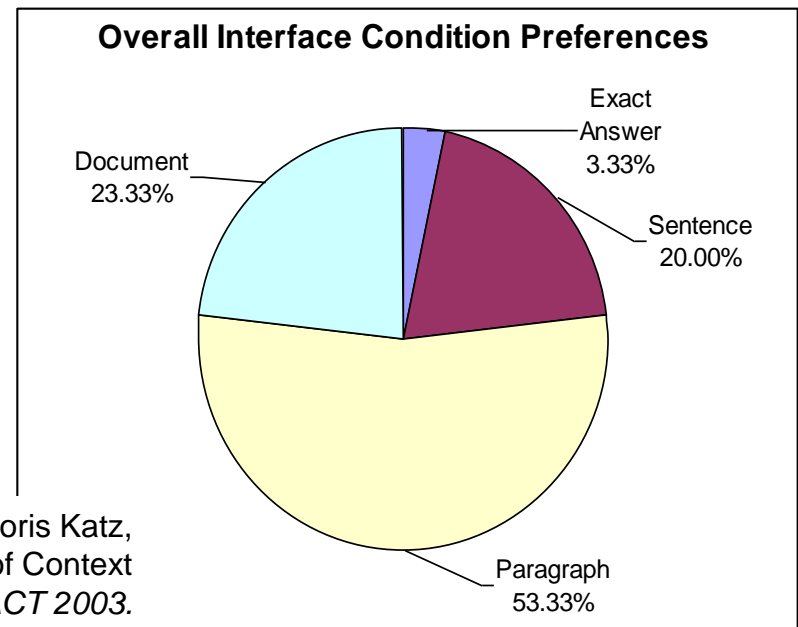
Goldilocks and the Three Summaries...

- 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

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)

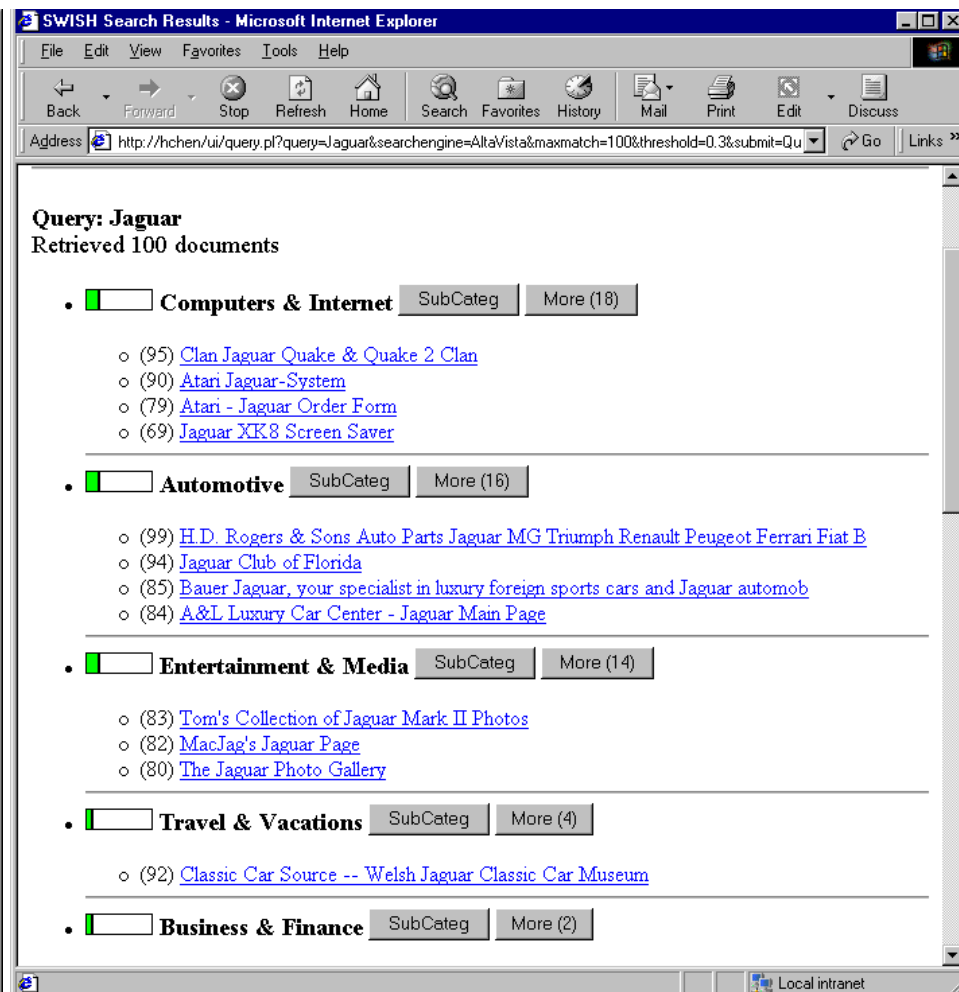
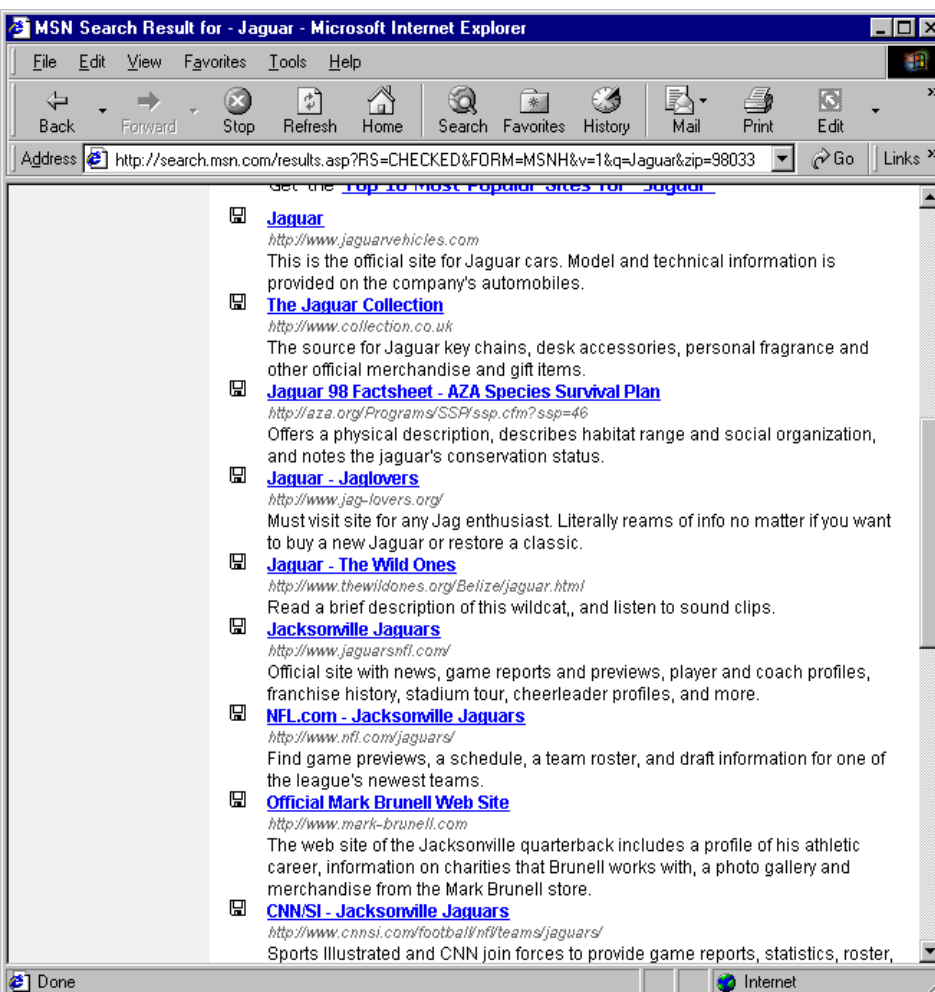
1. [Solar Hydrogen Energy Corporation](#) - Shec has developed a process to produce hydrogen which does not create green house gases such as co or co2, using only water and sunlight as the inputs into the process.
-- <http://www.solar-h2.com/> [Business: Energy and Environment: Renewable](#) (304)
2. [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)

Query: jaguar

SWISH

List Interface

Category Interface

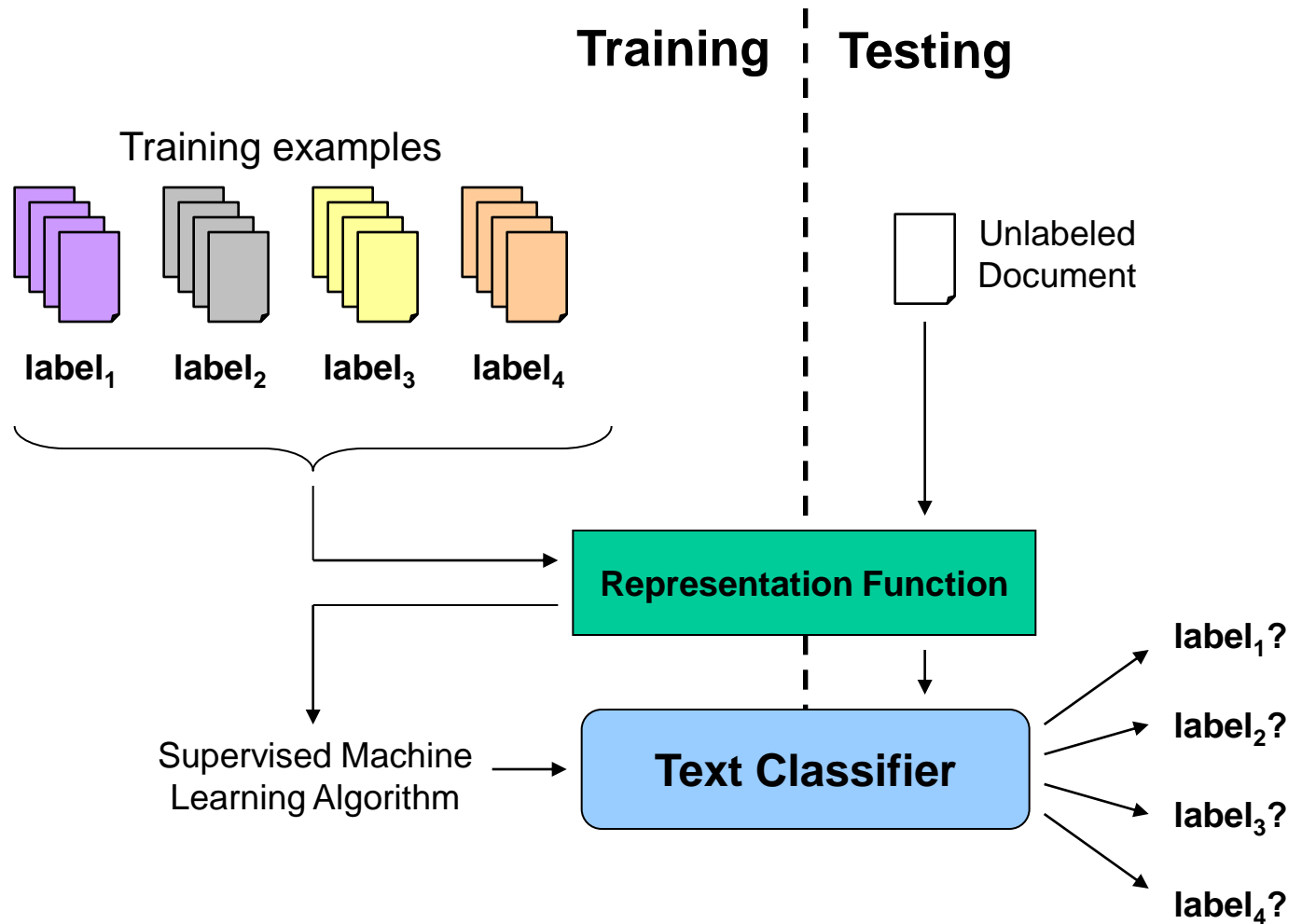


Hao Chen and Susan Dumais. (2000) Bringing Order to the Web: Automatically Categorizing Search Results. Proceedings of CHI 2000.

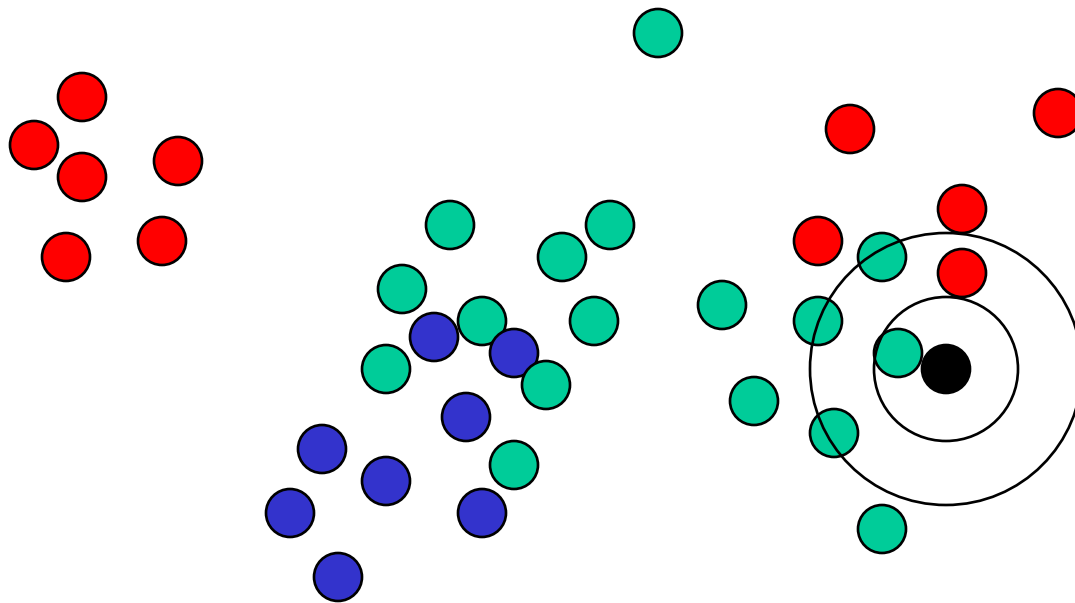
Text Classification

- 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



[company](#) | [products](#) | [solutions](#) | [customers](#) | [demos](#) | [partners](#) | [press](#)

[Advanced](#)
[Help!](#)

[Refer us to a friend](#)

NEW [Toolbar](#) or [MiniBar](#)!

Clustered Results

[university of maryland college park](#) (156)

- [Department](#) (47)
- [Students](#) (31)
- [School](#) (17)
- [Resources](#) (14)
- [Studies](#) (18)
- [Faculty and staff](#) (6)
- [Physics](#) (6)
- [Education](#) (8)
- [Laboratory](#) (6)
- [Terrapins, Lacrosse](#) (4)
- [More](#)

Find in clusters:

Cluster **Students** contains 31 documents.

1. [University of Maryland](#) [\[new window\]](#) [\[frame\]](#) [\[preview\]](#)

Official web site of the **University of Maryland** located in **College Park, Maryland**. Offers **student**, faculty and staff directories, departments and programs, athletics, admissions, and campus life information.

URL: www.umd.edu - [show in clusters](#)

Sources: [Open Directory 1](#)

2. [SGA University of Maryland](#) [\[new window\]](#) [\[frame\]](#) [\[preview\]](#)

University of Maryland Student Government Association. Includes links to **student** legislators, voting records, and general information about the UM **student** government association.

URL: www.marylandsga.com - [show in clusters](#)

Sources: [Open Directory 7](#)

3. [Terp Rentals](#) [\[new window\]](#) [\[frame\]](#) [\[preview\]](#)

Information on off campus housing for **University of Maryland, College Park students** and teachers.

URL: www.terprentals.com - [show in clusters](#)

Sources: [Open Directory 11](#)

4. [Department of Economics | University of Maryland](#) [\[new window\]](#) [\[frame\]](#) [\[preview\]](#)

... Schedules The Inter- **University Student** Conference ... about the Campus **University of Maryland** at **College Park** ... Department of Economics, **University of Maryland**, College Park, MD 20742 3105...

URL: www.bsos.umd.edu/econ - [show in clusters](#)

Sources: [Lycos 6](#)

<http://www.vivisimo.com>

Kartoo's Cluster Visualization

The screenshot displays the Kartoo website interface. At the top, the search bar contains the text "university of maryland college park". The main area features a cluster visualization of websites, with nodes labeled with terms like "states", "copyright", "served", "technology", "professor", "department", "distinguished", "science", "comments", "umterps.fansonly.com", "umterps.ocsn.com", "www.umuc.edu", "www.getcited.org", "www.inform.umd.edu", "www.enfp.umd.edu", "www.umpd.umd.edu", "www.math.umd.edu", "www.lib.umd.edu", "www.cs.umd.edu", and "www.bsos.umd.edu". The nodes are connected by lines, forming a complex network. The left sidebar contains a "Topics" list with categories like "copyright 2003 university", "provided to the university", "school of business", "student advantage", "copyright", "comments", "department", "science", "distinguished", "professor", "technology", "served", "states", "terrapins", "athletic", "sports", and "network". The right sidebar shows a list of "Found sites" including "www.worldwidelearn.com", "www.umd.edu", "www.cs.umd.edu", "www.inform.umd.edu", "www.lib.umd.edu", "www.math.umd.edu", "www.bsos.umd.edu", "www.enfp.umd.edu", "www.umpd.umd.edu", "www.umuc.edu", "umterps.ocsn.com", and "www.getcited.org". Below this is a "Sponsor" section for "University of Maryland Online Degrees". The bottom of the page features a "Click here to start search" button and a "Watch" button.

Kartoo

help << university of maryland college park Search english pages options Products

Found sites :

- www.worldwidelearn.com
- www.umd.edu
- www.cs.umd.edu
- www.inform.umd.edu
- www.lib.umd.edu
- www.math.umd.edu
- www.bsos.umd.edu
- www.enfp.umd.edu
- www.umpd.umd.edu
- www.umuc.edu
- umterps.ocsn.com
- www.getcited.org

Sponsor

University of Maryland Online Degrees WorldWideLearn presents our education partner University of Maryland. UMUC offers 34 degrees and 55 certificates - entirely online.

Map : university of maryland colleg

- print the map
- Send a map
- Add a site
- Add a Topic
- save the map...

next map

Watch

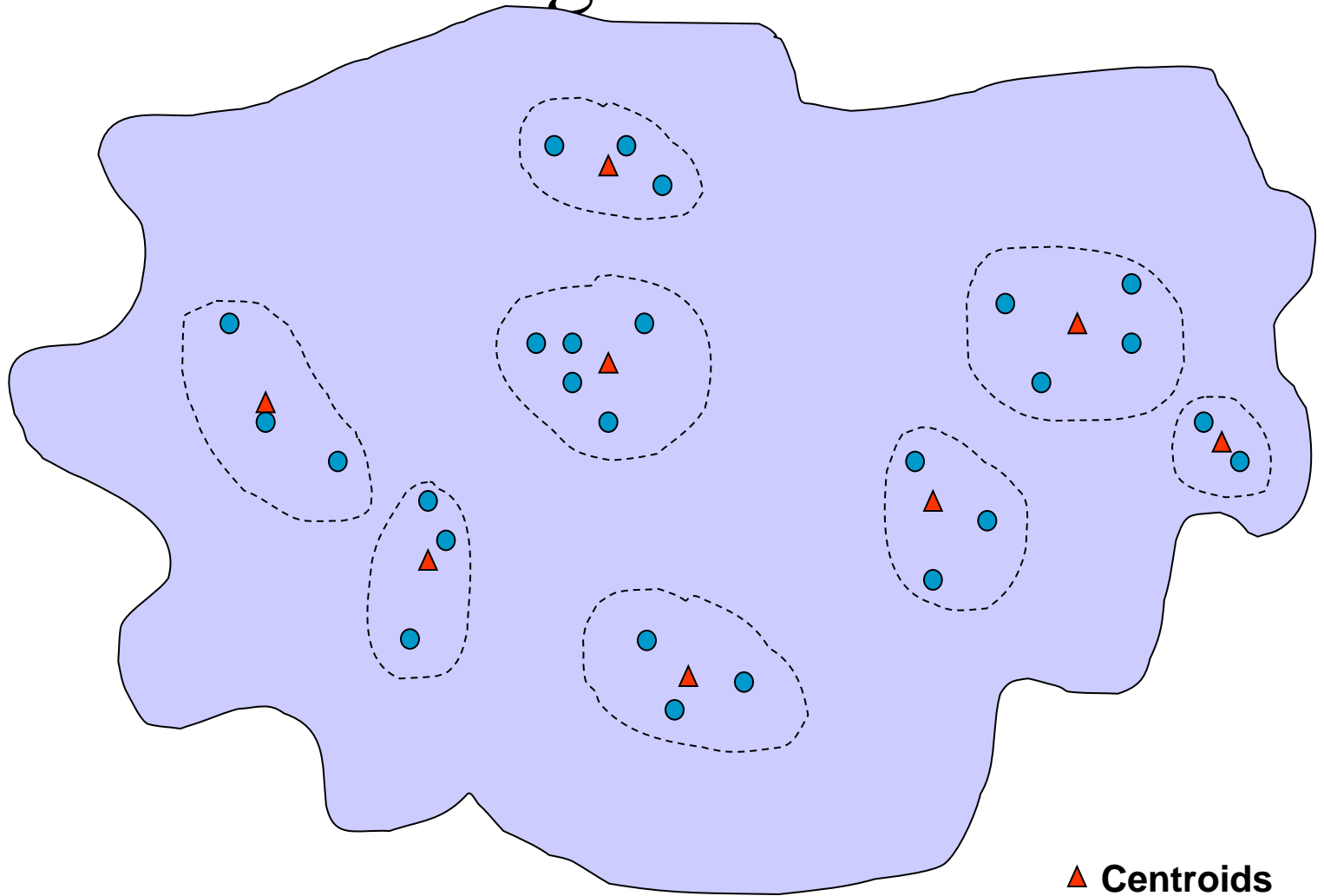
Click here to start search

1 540 000 Found results (1 -> 13)

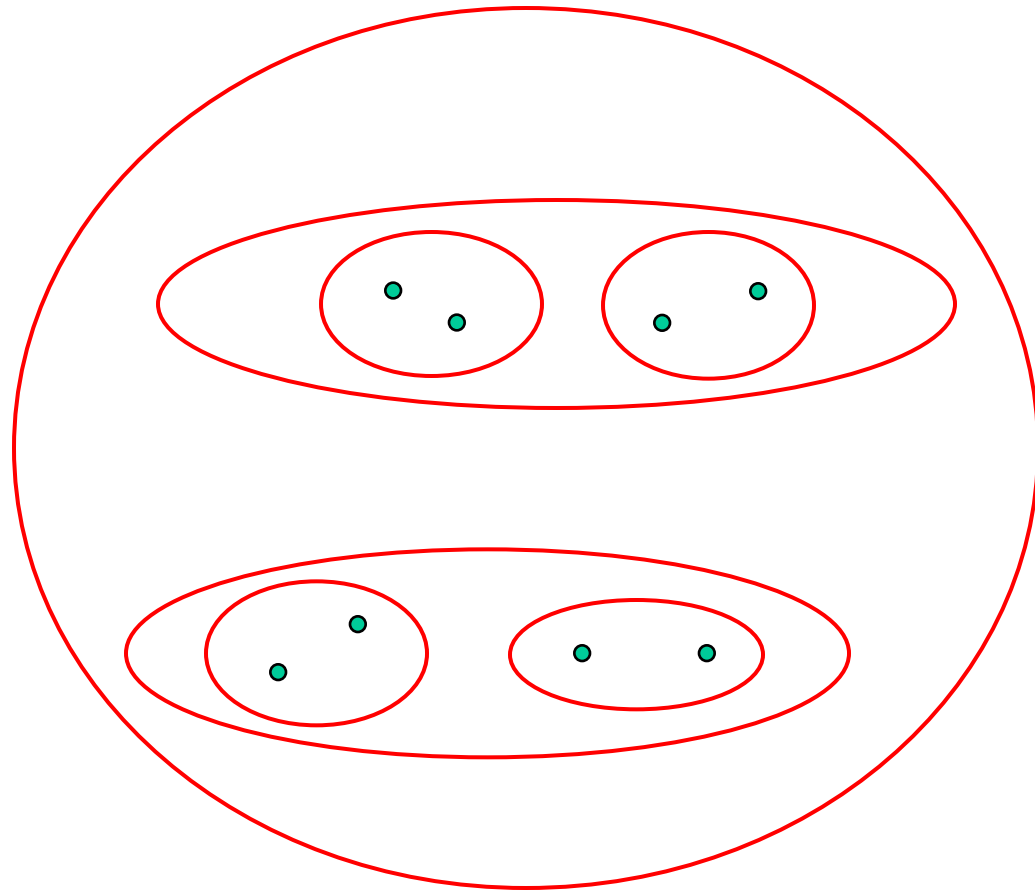
Clustering Result Sets

- 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

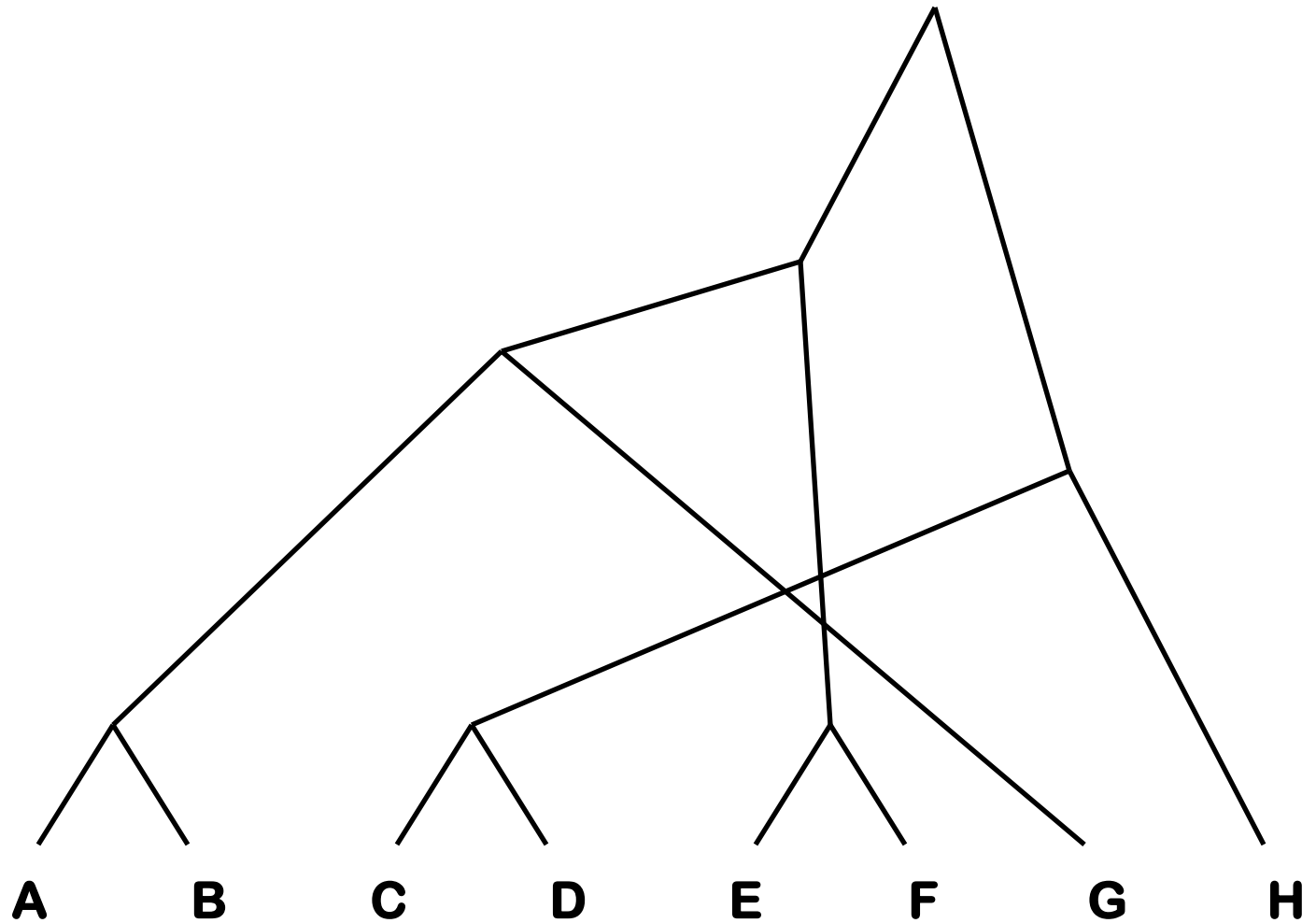
Visualizing Clusters



Hierarchical Agglomerative Clustering



Another Way to Look at H.A.C.



The H.A.C. Algorithm

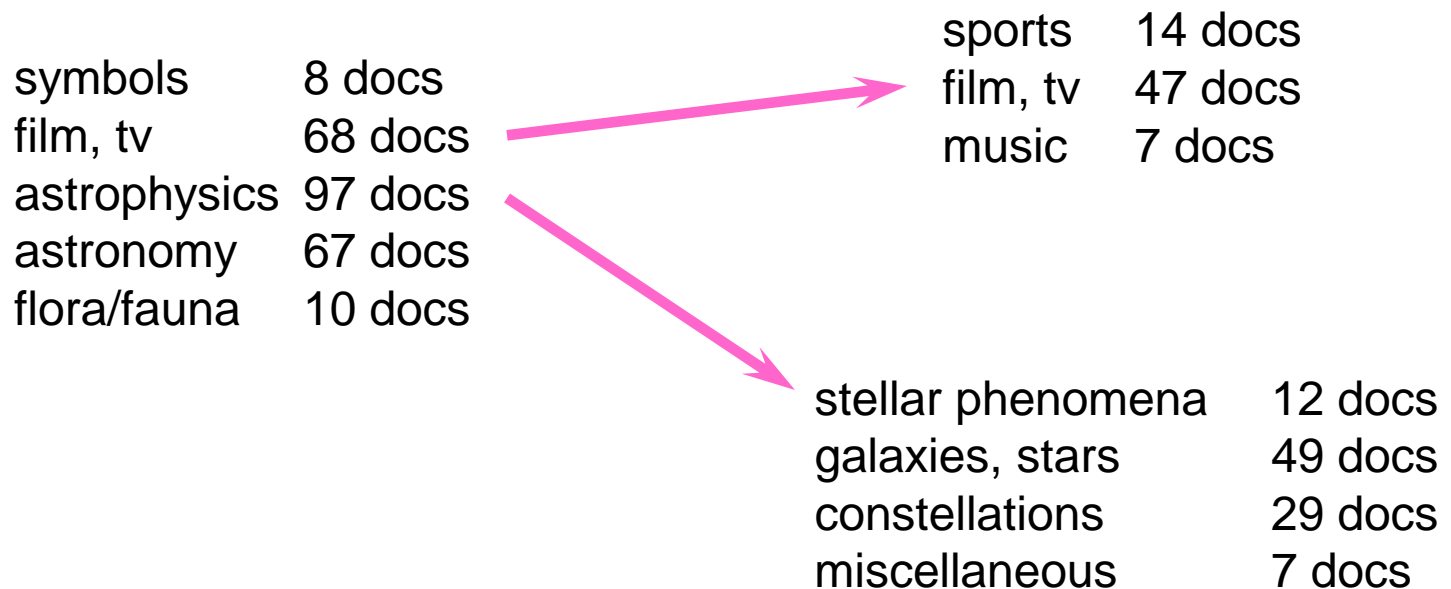
- Start with each document in its own cluster
- Until there is only one cluster:
 - Determine the two most similar clusters c_i and c_j
 - Replace c_i and c_j with a single cluster $c_i \cup c_j$
- The history of merging forms the hierarchy

Cluster Similarity

- Assume a similarity function that determines the similarity of two instances: $\text{sim}(x,y)$
 - 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

- System 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”

☐ 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
- ☐ Milledgeville, Georgia

☐ Cluster 2 Size: 68 film play career win television role record award york popular stage p

- ☐ Burstyn, Ellen
- ☐ Stanwyck, Barbara
- ☐ Berle, Milton
- ☐ Zukor, Adolph
- ☐ DeMille, Cecil

☐ Cluster 3 Size: 97 bright magnitude cluster constellation line type contain period spectr

- ☐ star
- ☐ Galaxy, The
- ☐ extragalactic systems
- ☐ interstellar matter
- ☐ cluster, star

☐ Cluster 4 Size: 67 astronomer observatory astronomy position measure celestial telescop

- ☐ astronomy and astrophysics
- ☐ astrometry
- ☐ Agena
- ☐ astronomical catalogs and atlases
- ☐ Herschel, Sir William

☐ 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 basebal

- ☐ 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

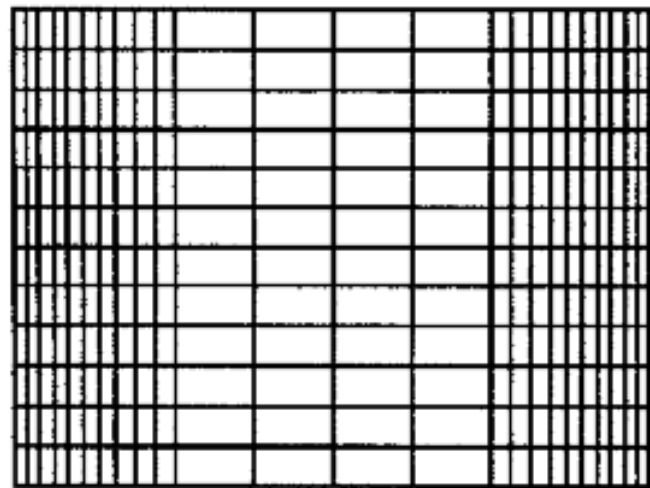
Recap

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

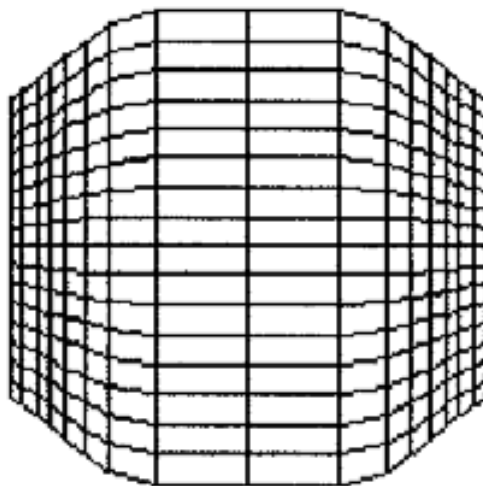
Agenda

- Query formulation
- Selection
- Examination
- Source selection

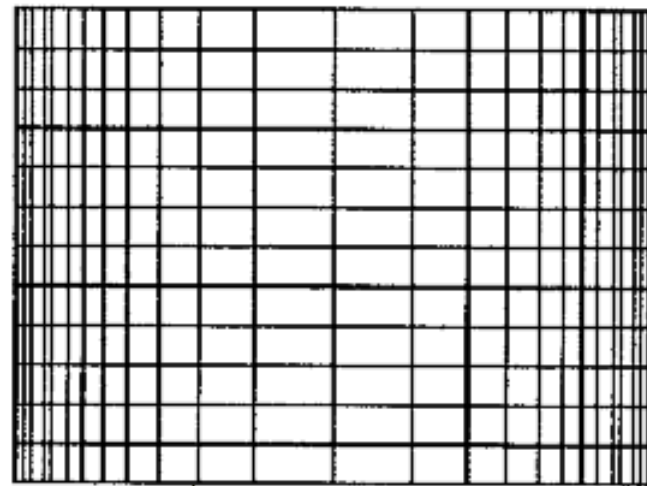
Distorting Reality



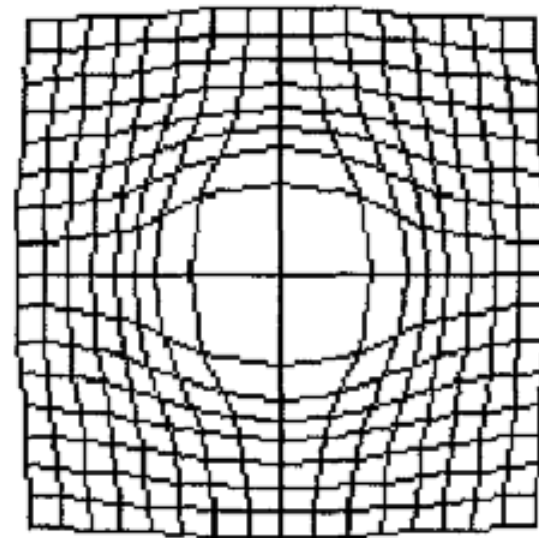
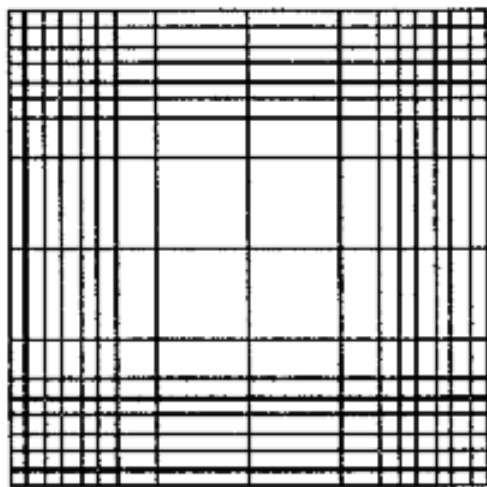
Bifocal



Perspective Wall

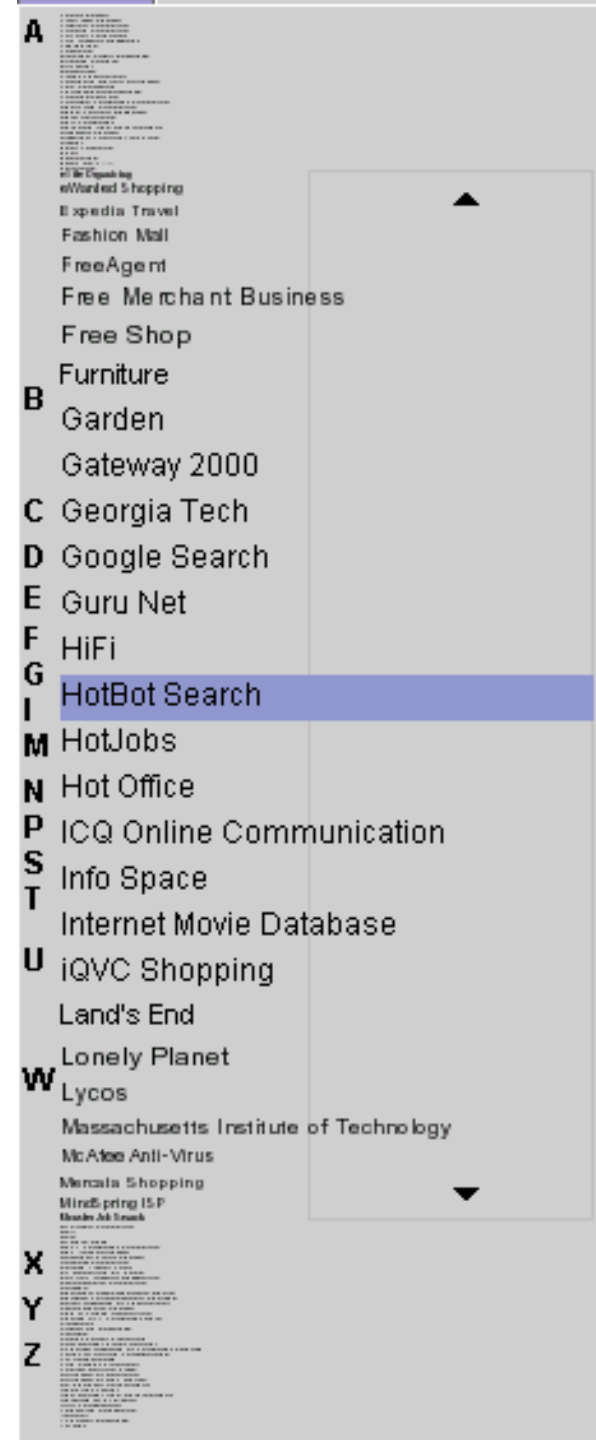


Fisheye

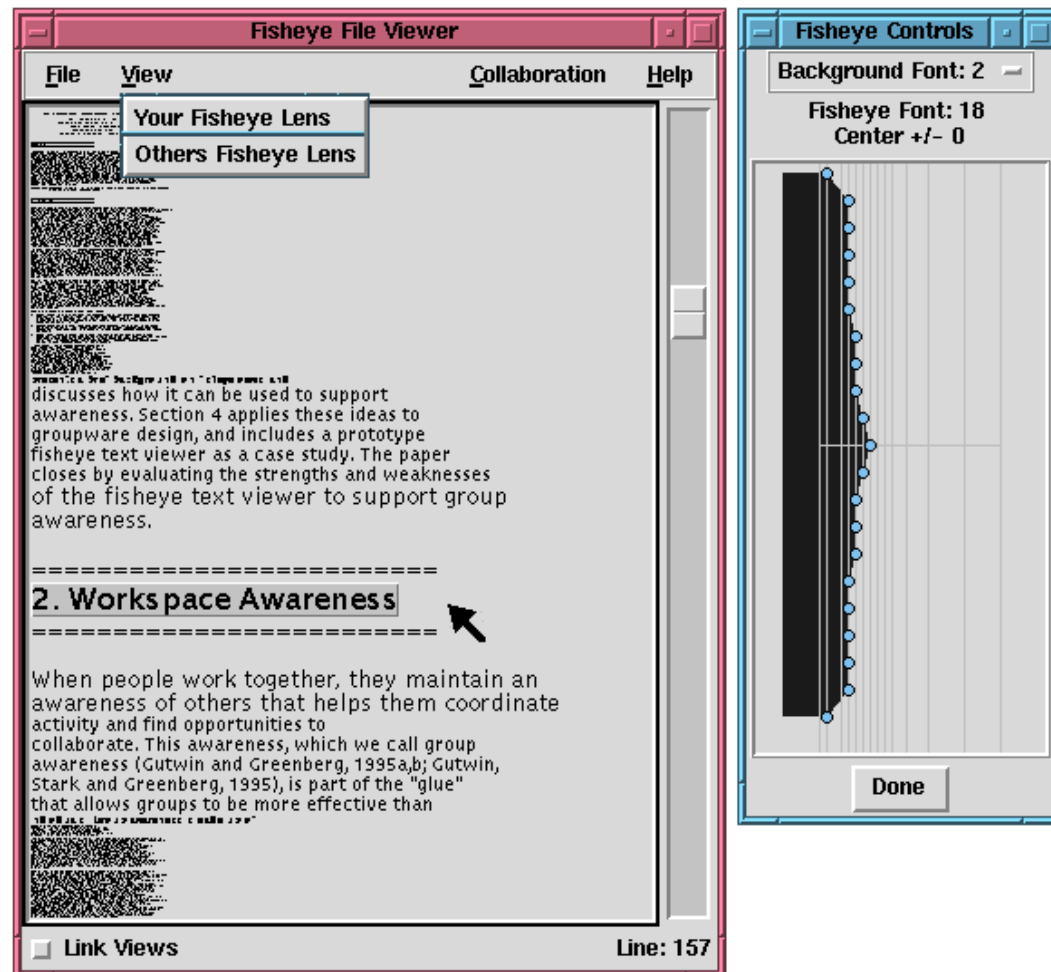


1-D Fisheye Menu

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



1-D Fisheye Document Viewer



TileBars

Topic: reliability of DBMS (database systems)

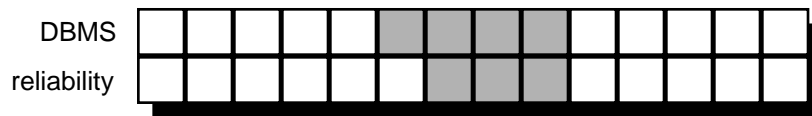
Query terms: DBMS, reliability



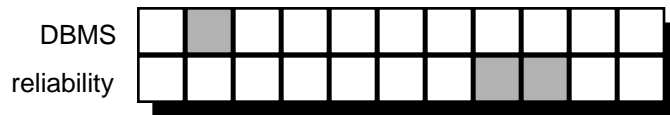
Mainly about both DBMS
and reliability



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

health and safety of staff and inmates, the Bureau of Prisons will restrict areas and circumstances in which **smoking** is permitted within its institutions and offices.

(a) All areas of Bureau of Prisons facilities and vehicles are no **smoking** areas unless specifically designated as **smoking** areas by the Chief Executive Officer consistent with the guidelines set forth in this rule.

(b) Chief Executive Officers shall limit **smoking** areas to a minimum number of locations, consistent with effective operations. Under no circumstances shall **smoking** be permitted in the following areas, except as noted in &Section; 551.162(a) :

- (1) Elevators,
- (2) Storage Rooms and Warehouses,
- (3) Libraries,
- (4) Corridors and Halls,
- (5) Dining Facilities,
- (6) Kitchen and Food Preparation Areas,
- (7) Medical/Dental Care Delivery Areas,
- (8) Institution/**Government** Vehicles,
- (9) Administrative Areas and Offices,
- (10) Auditoriums,
- (11) Class and Conference Rooms,
- (12) Gymnasiums and Exercise Rooms, and
- (13) Restrooms.

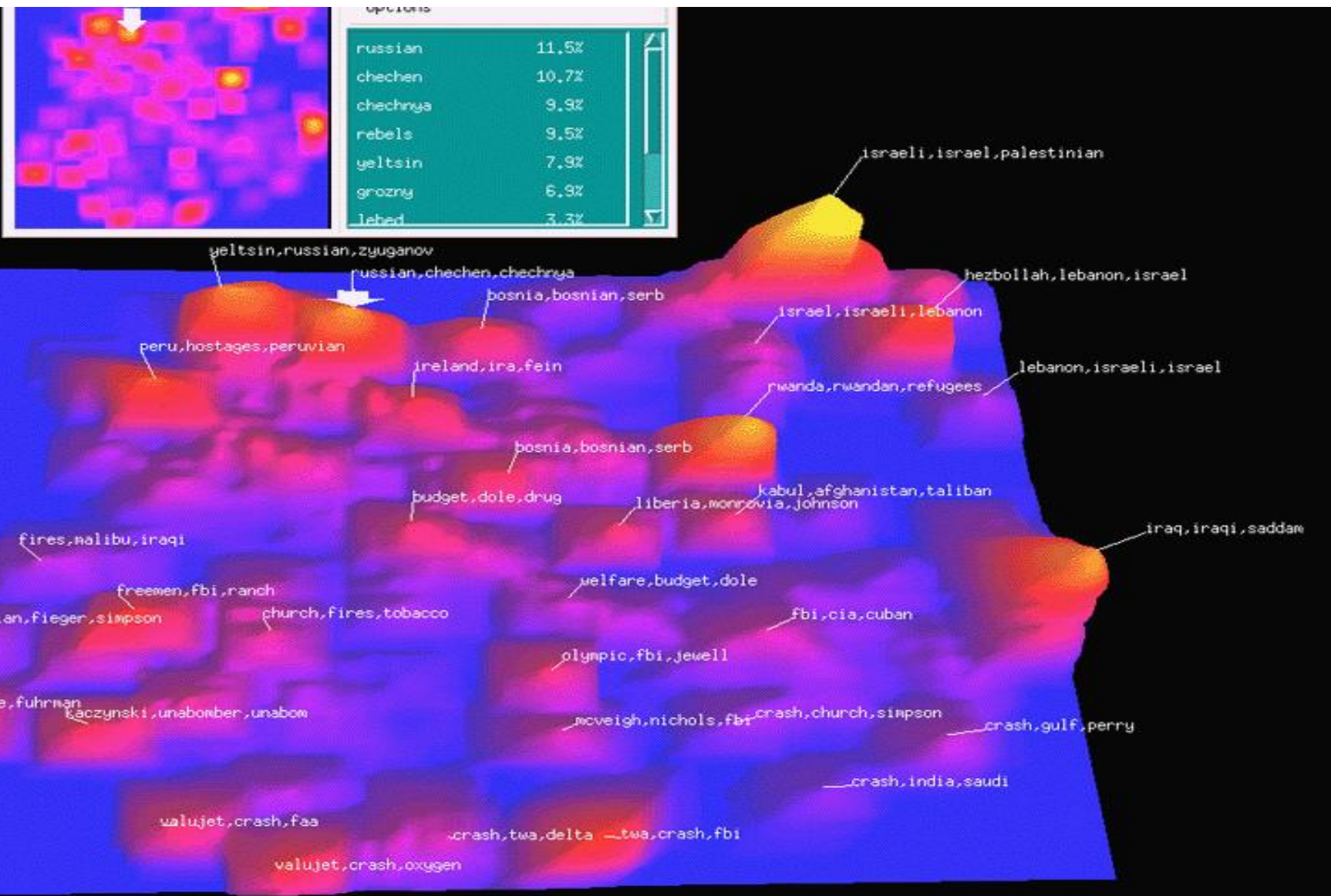
&Section; 551.161
Definition.

For purpose of this rule, **smoking** is defined as carrying or inhaling a lighted cigar, cigarette, pipe or other lighted tobacco products.

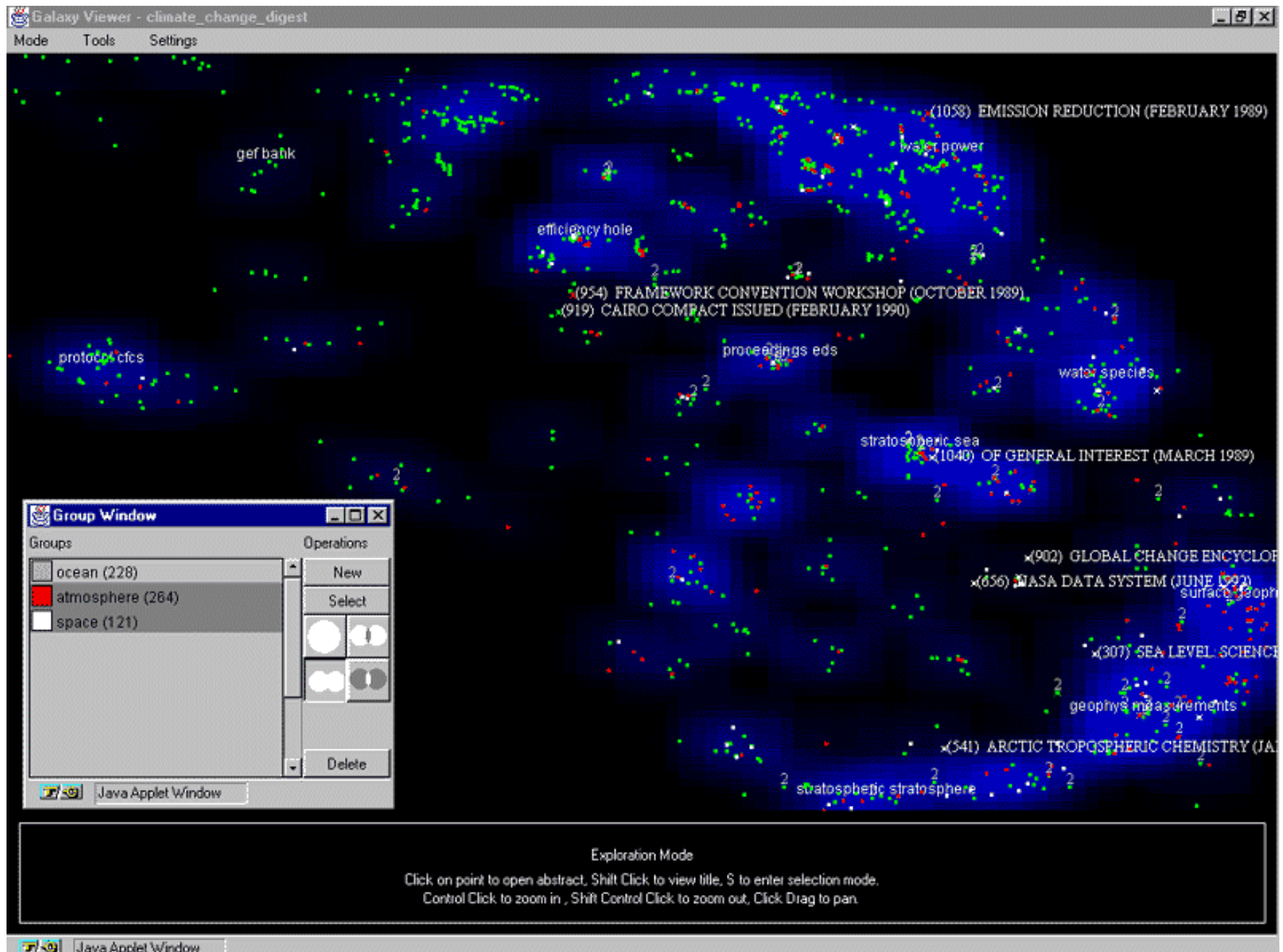
Agenda

- Query formulation
- Selection
- Examination
- Source selection

Theme View



WebTheme



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.